

09

Abstracts of Papers

TUTORIALS December 7, 2009 Hyatt Regency Vancouver, BC, Canada

CONFERENCE SESSIONS December 7-10, 2009 Hyatt Regency Vancouver, BC, Canada

MINI SYMPOSIA December 10, 2009 Hyatt Regency Vancouver, BC, Canada

WORKSHOP December 11-12, 2009 The Westin Resort & Spa The Hilton Whistler Resort & Spa Whistler, BC, Canada



Neural Information Processing Systems Foundation



Abstracts of Papers

9

TUTORIALS December 7, 2009 Hyatt Regency Vancouver, BC, Canada

CONFERENCE SESSIONS December 7-10, 2009 Hyatt Regency Vancouver, BC, Canada

MINI SYMPOSIA December 10, 2009 Hyatt Regency Vancouver, BC, Canada

WORKSHOP December 11-12, 2009 The Westin Resort & Spa The Hilton Whistler Resort & Spa Whistler, BC, Canada



Neural Information Processing Systems Foundation

Sponsored by the Neural Information Processing System Foundation, Inc

The technical program includes six invited talks and 263 accepted papers, selected from a total of 1,105 submissions considered by the program committee. Because the conference stresses interdisciplinary interactions, there are no parallel sessions. Papers presented at the conference will appear in "Advances in Neural Information Processing Systems 22," edited by Yoshua Bengio, Dale Schuurmans, John Lafferty Chris Williams and Aron Culotta.

Contents

Organizing Committee	. 7
Program Committee	. 7
NIPS Foundation Offices and Board Members	. 8
Sponsors	. 9
Core Logistics Team	. 9
Awards	. 10
Program Highlights	11
Schedule of Presentations	15
Schedule, Monday	. 15
Schedule, Tuesday	. 24
Schedule, Wednesday	. 35
Schedule, Thursday	. 46
Abstracts of Tutorials	47
Tutorials, Monday, Session 1	. 47
Tutorials, Monday, Session 1	. 48
Tutorials, Monday, Session 2	. 49
Tutorials, Monday, Session 2	. 50
Tutorials, Monday, Session 3	. 51
Tutorials, Monday, Session 3	. 52
Monday	55
Monday, Spotlights	. 55
Monday, Poster Session	. 57
Tuesday	99
Tuesday, Oral Session	. 99
Tuesday, Oral Session	. 100
Tuesday, Publication Models Debate	. 102
Tuesday, Oral Session	. 104
Tuesday, Spotlights	. 105
Tuesday, Oral Session	. 107
Tuesday, Spotlights	. 109
Tuesday, Poster Session	. 110
Tuesday, Demonstrations	. 149

Wednesday 15	3
Wednesday, Oral Session	63
Wednesday, Oral Session	55
Wednesday, Oral Session	57
Wednesday, Spotlights	68
Wednesday, Oral Session	50
Wednesday, Spotlights	51
Wednesday, Poster Session	i 3
Wednesday, Demonstrations	9
Thursday 20)3
Thursday, Oral Session)3
Thursday, Oral Session)4
Thursday, Mini Symposia)5
Poster and Demo Floor Plans 20	9
Reviewers 21	.7
Subject Index 22	:5
Author Index22	:9
Notes 23	57

Organizing Committee

General Chair

Program Chairs

Tutorials Chair Workshop Chairs Demonstration Chairs Publications Chair and Electronic Proceedings Chair Volunteers Chair Publicity Chair YOSHUA BENGIO, University of Montreal DALE SCHUURMANS, University of Alberta JOHN LAFFERTY, Carnegie Mellon University CHRIS WILLIAMS, University of Edinburgh YANN LECUN, New York University RICHARD ZEMEL, University of Toronto HANS PETER GRAF, NEC Laboratories America, Inc. ARON CULOTTA, Southeastern Louisiana University

NIPS Foundation Office NIPS Foundation Office

Program Committee

JEAN-YVES AUDIBERT. Ecole des Ponts ParisTech DAVID BLEI, Princeton University **KWABENA BOAHEN**, Stanford University MICHAEL BOWLING, University of Alberta NICOLO CESA-BIANCHI, University of Milan AARON COURVILLE, University of Montreal KOBY CRAMMER, University of Pennsylvania NATHANIEL DAW, New York University DAVID DUNSON, Duke University PAOLO FRASCONI, University of Florence NIR FRIEDMAN, Hebrew University of Jerusalem ARTHUR GRETTON, Carnegie Mellon University and Max Planck Institute MATTHIAS HEIN, Saarland University AAPO HYVARINEN, University of Helsinki THORSTEN JOACHIMS, Cornell University MARK JOHNSON, Brown University CHARLES KEMP, Carnegie Mellon University JOHN LAFFERTY, Carnegie Mellon University WEE SUN LEE, National University of Singapore TAI SING LEE, Carnegie Mellon University JON MCAULIFFE, University of Pennsylvania YAEL NIV, Princeton University ROBERT NOWAK, University of Wisconsin, Madison PASCAL POUPART, University of Waterloo CARL RASMUSSEN, University of Cambridge ERIK SUDDERTH, Brown University BEN TASKAR, University of Pennsylvania ANTONIO TORRALBA, Massachusetts Institute of Technology BILL TRIGGS, Laboratoire Jean Kuntzmann, CNRS SETHU VIJAYAKUMAR, University of Edinburgh CHRIS WILLIAMS, University of Edinburgh ANDREW ZISSERMAN, University of Oxford

NIPS Foundation Officers and Board Members

President	TERRENCE SEJNOWSKI, The Salk Institute
Vice President for Development	SEBASTIAN THRUN, Stanford University
Treasurer	MARIAN STEWART BARTLETT, University of California, San Diego
Secretary	MICHAEL MOZER, University of Colorado, Boulder
Legal Advisor	Phil Sotel, Pasadena, CA
Executive Board	SUE BECKER, McMaster University, Ontario, Canada
	JOHN C. PLATT, Microsoft Research
	LAWRENCE SAUL, University of Pennsylvania
	BERNHARD SCHÖLKOPF, Max Planck Institute
	YAIR WEISS, Hebrew University of Jerusalem
	DAPHNE KOLLER, Stanford University
Advisory Board	GARY BLASDEL, Harvard Medical School
	JACK COWAN, University of Chicago
	STEPHEN HANSON, Rutgers University
	MICHAEL I. JORDAN, University of California, Berkeley
	MICHAEL KEARNS, University of Pennsylvania
	SCOTT KIRKPATRICK, Hebrew University, Jerusalem
	RICHARD LIPPMANN, Massachusetts Institute of Technology
	TODD K. LEEN, Oregon Graduate Institute
	BARTLETT MEL, University of Southern California
	JOHN MOODY, International Computer Science Institute, Berkeley
	and Portland
	GERALD TESAURO, IBM Watson Labs
	DAVE TOURETZKY, Carnegie Mellon University
	THOMAS G. DIETTERICH, Oregon State University
	SARA A. SOLLA, Northwestern University Medical School
Emeritus Members	TERRENCE L. FINE, Cornell University
	EVE MARDER, Brandeis University

Sponsors

NIPS gratefully acknowledges the generosity of those individuals and organizations who have provided financial support for the NIPS 2009 conference. The financial support enabled us to sponsor student travel and participation, the outstanding student paper awards, the demonstration track and the opening buffet.

Air Force Office of Scientific Research Artificial Intelligence Google Microsoft Research IBM Research D.E. Shaw & Co. Yahoo Labs Alberta Ingenuity Center for Machine Learning MITACS Intel Pascal Toyota Springer

Core Logistics Team

The running of NIPS would not be possible without the help of many volunteers, students, researchers and administrators who donate their valuable time and energy to assist the conference in various ways. However, there is a core team at the Salk Institute whose tireless efforts make the conference run smoothly and efficiently every year. This year, NIPS would particularly like to acknowlege the exceptional work of:

YUCHENG LOW, Workflow Master Lee Campbell Chris Hiestand Sheri Leone Mary Ellen Perry

Awards

OUTSTANDING STUDENT PAPER AWARDS

An LP View of the M-best MAP problem MENACHEM FROMER^{*} and AMIR GLOBERSON

Fast subtree kernels on graphs NINO SHERVASHIDZE^{*} and KARSTEN BORGWARDT

STUDENT PAPER HONORABLE MENTIONS

Reading Tea Leaves: How Humans Interpret Topic Models JONATHAN CHANG*, JORDAN BOYD-GRABER*, SEAN GERRISH, CHONG WANG and DAVID BLEI

Multi-Label Prediction via Compressed Sensing DANIEL HSU^{*}, SHAM KAKADE, JOHN LANGFORD and TONG ZHANG

Kernel Choice and Classifiability for RKHS Embeddings of Probability Distributions

BHARATH SRIPERUMBUDUR*, KENJI FUKUMIZU, ARTHUR GRETTON, GERT LANCKRIET, and BERNHARD SCHÖELKOPF

Program Highlights

Monday, December 7th

7:30am-9:00am	Breakfast: Regency A — Conference Level	
$8:00 \mathrm{am}-6:00 \mathrm{pm}$	Registration Desk Open	
$8:00 \mathrm{am}-6:00 \mathrm{pm}$	Internet Access Room Open	
9:30am-5:30pm	Tutorials	15
1:00pm-6:30pm	Poster Setup and Preview	
6:15 pm-6:30 pm	Opening Remarks and Awards	
$6:30 \mathrm{pm}-6:50 \mathrm{pm}$	Spotlights	15
$7:00 \text{pm}{-}11:59 \text{pm}$	Poster Session	16

Tuesday, December 8th

$7:30 \mathrm{am}-9:00 \mathrm{am}$	Breakfast: Perspectives Level			
$8:00 \mathrm{am}-9:00 \mathrm{am}$	Breakfast: also in Regency A — Conference Level			
$8:00 \mathrm{am}-6:00 \mathrm{pm}$	Registration Desk Open			
$8:00 \mathrm{am}-6:00 \mathrm{pm}$	Internet Access Room Open			
8:30am–10:10am	Oral Session: Information Theory and Estimation			
	Invited Talk: <i>Relative Entropy</i> Sergio Verdu			
10:10am–10:40am	Break			
$10:40 \text{am}{-}12:00 \text{pm}$	Oral Session: Images and Codes			
12:00 pm - 2:00 pm	Break for Lunch			
12:30pm-1:45pm	Publication Models Debate 24			
2:00 pm-3:20 pm	Oral Session: Deep Learning and Network Models			
	Invited Talk: Deep Learning with Multiplicative Interactions Geoffrey Hinton			
3:20pm-3:40pm	Spotlights			
3:40pm-4:10pm	Break			
4:10 pm-5:10 pm	Oral Session: Cognitive Science and EEG Analysis			
$5:10 \mathrm{pm}{-}5:30 \mathrm{pm}$	Spotlights			
1:00pm-6:00pm	Poster/Demo Setup and Preview			

PROGRAM HIGHLIGHTS

7:30pm-11:59pm	Poster Session	27
7:30 pm - 11:59 pm	Demonstrations	34

Wednesday, December 9th

7:30am–9:00am	Breakfast: Perspectives Level
8:00am-9:00am	Breakfast: also in Regency A — Conference Level
$8:00 \mathrm{am}-6:00 \mathrm{pm}$	Registration Desk Open
$8:00 \mathrm{am}-6:00 \mathrm{pm}$	Internet Access Room Open
$8:30 \text{am}{-10:10} \text{am}$	Oral Session: Neuroscience
	Invited Talk: The Rat Vibrissal Array as a Model Sensorimotor System MITRA HARTMANN
10:10am–10:40am	Break
10:40 am - 12:00 pm	Oral Session: Theory, Optimization and Games
12:00pm-2:00pm	Break for lunch
2:00 pm-3:20 pm	Oral Session: Vision and Inference
	Invited Talk: Learning and Inference in Low-Level Vision YAIR WEISS
1:00pm-6:00pm	Poster/Demo Setup and Preview
3:20pm-3:40pm	Spotlights
3:40pm-4:10pm	Break
4:10pm-5:10pm	Oral Session: Probabilistic Models and Kernels
$5:10 \mathrm{pm}{-}5:30 \mathrm{pm}$	Spotlights
$7:30 \mathrm{pm}{-}11:59 \mathrm{pm}$	Poster Session
7:30pm-11:59pm	Demonstrations

Thursday, December 10th

7:30am-9:00am	Breakfast: Perspectives Level		
$8:00 \mathrm{am}-9:00 \mathrm{am}$	Breakfast: also in Regency A — Conference Level		
$8:00 \text{am}{-}11:00 \text{am}$	Registration Desk Open		
8:00am-12:00pm	Internet Access Room Open		
8:30am-9:50am	Oral Session: Bayesian Analysis		
	Invited Talk: Bayesian Analysis of Markov Chains Persi Diaconis		

PROGRAM HIGHLIGHTS

$9:50 \mathrm{am}{-10:30 \mathrm{am}}$	Break
10:30 am - 11:50 am	Oral Session: Neural Modeling and Imaging
	Invited Talk: Optogenetics: Development and Application KARL DEISSEROTH
12:00pm	Close of Main Conference
1:30 pm-2:50 pm	Mini Symposia
2:50 pm-3:10 pm	Break
3:10 pm-4:30 pm	Mini Symposia
2:00pm-6:00pm	Buses depart for Workshops

PROGRAM HIGHLIGHTS

Schedule of Presentations

Monday, December 7th

9:30am-11:30am Tutorial Session 1

Making Very Large-Scale Linear Algebraic Computations Possible Via Randomization GUNNAR MARTINSSON Location: Regency E/F

Understanding Visual Scenes ANTONIO TORRALBA Location: Regency D

1:00pm-3:00pm Tutorial Session 2

Model-Based Reinforcement Learning MICHAEL LITTMAN Location: Regency E/F

Sparse Methods for Machine Learning: Theory and Algorithms FRANCIS BACH Location: Regency D

3:30pm-5:30pm Tutorial Session 3

Deep Learning in Natural Language Processing RONAN COLLOBERT and JASON WESTON Location: Regency E/F

Sequential Monte-Carlo Methods ARNAUD DOUCET and NANDO DE FREITAS Location: Regency D

6:30pm-7:45pm Buffet, Opening Remarks and Awards

Dinner buffet open until 9pm; dessert buffet until 10pm.

6:30-6:50pm Spotlights

Learning	to	Rank	by	Optimizing NDCG Measure, HAMED	
VALIZADEC	GAN,	Rong	Jin,	Michigan State University, RUOFEI ZHANG and	
JIANCHANG	G MA	40, Yah	00!		88

On Learning Rotations, RAMAN ARORA, Univ. of Wisconsin-Madison...... 88

SCHEDULE, MONDAY

	Replacing supervised classification learning by Slow Feature Analysis in spiking neural networks, STEFAN KLAMPFL and WOLFGANG MAASS, Graz University of Technology	66
	Heavy-Tailed Symmetric Stochastic Neighbor Embedding, Zhirong Yang, Cuhk, TKK, Irwin King, Zenglin Xu, Cuhk, and Erkki Oja, TKK.	73
	Spatial Normalized Gamma Processes, VINAYAK RAO and YEE WHYE TEH, Gatsby Computational Neuroscience Unit, UCL	60
	Rethinking LDA: Why Priors Matter, HANNA WALLACH, DAVID MIMNO and ANDREW MCCALLUM, University of Massachusetts Amherst	75
	Nonparametric Latent Feature Models for Link Prediction , KURT MILLER, UC Berkeley, THOMAS GRIFFITHS, University of California, Berkeley, and MICHAEL JORDAN, University of California	63
	A Fast, Consistent Kernel Two-Sample Test, ARTHUR GRETTON, Carnegie Mellon University and Max Planck Institute, KENJI FUKUMIZU, The Institute of Statistical Mathematics, ZAID HARCHAOUI, Carnegie Mellon University, and BHARATH SRIPERUMBUDUR, UC San Diego	94
	Entropic Graph Regularization in Non-Parametric Semi-Supervised Classification, AMARNAG SUBRAMANYA and JEFF BILMES, University of Washington	57
	Asymptotic Analysis of MAP Estimation via the Replica Method and Compressed Sensing, SUNDEEP RANGAN, Qualcomm, Inc., ALYSON FLETCHER, University of California, Berkeley, and VIVEK GOYAL, Massachusetts Institute of Technology	63
	Fast Learning from Non-i.i.d. Observations, INGO STEINWART, Los Alamos National Laboratory, and ANDREAS CHRISTMANN, University of Bayreuth	84
	Convergent Temporal-Difference Learning with Arbitrary Smooth Function Approximation, HAMID MAEI, CSABA SZEPESVARI, University of Alberta, SHALABH BHATNAGAR, Indian Institute of Science, Bangalore, India, DOINA PRECUP, McGill University, DAVID SILVER and RICH SUTTON, University of Alberta	96
7:00–11:59pm	Poster Session	
M1	Entropic Graph Regularization in Non-Parametric Semi-Supervised Classification, AMARNAG SUBRAMANYA and JEFF BILMES, University of Washington	57
M2	A Sparse Non-Parametric Approach for Single Channel Separation of Known Sounds, PARIS SMARAGDIS, Adobe Systems Inc., MADHUSUDANA SHASHANKA, Mars Inc., and BHIKSHA RAJ, Carnegie Mellon University	57

M3	Sparse and Locally Constant Gaussian Graphical Models, JEAN HONORIO, LUIS ORTIZ, DIMITRIS SAMARAS, Stony Brook University, NIKOS PARAGIOS, Ecole Centrale Paris, and RITA GOLDSTEIN, Brookhaven National	
	Laboratory	7
M4	A Stochastic approximation method for inference in probabilistic graphical models, PETER CARBONETTO, University of Chicago, MATTHEW KING, University of British Columbia, and FIRAS HAMZE, D-Wave Systems 5	8
M5	Variational Inference for the Nested Chinese Restaurant Process, CHONG WANG and DAVID BLEI, Princeton University	,9
M6	Gaussian process regression with Student-t likelihood, JARNO VANHATALO, PASI JYLÄNKI and AKI VEHTARI, Helsinki University of Technology	9
M7	A Bayesian Analysis of Dynamics in Free Recall, RICHARD SOCHER, Stanford University, SAMUEL GERSHMAN, ADLER PEROTTE, PER SEDERBERG, DAVID BLEI and KENNETH NORMAN, Princeton University	9
M8	Streaming Pointwise Mutual Information, BENJAMIN VAN DURME, University of Rochester, and ASHWIN LALL, Georgia Institute of Technology 6	0
M9	Non-stationary continuous dynamic Bayesian networks, MARCO GRZEGORCZYK, Department of Statistics, TU Dortmund University, Germany, and DIRK HUSMEIER, Biomathematics and Statistics Scotland (BioSS), Edinburgh, UK	0
M10	Spatial Normalized Gamma Processes, VINAYAK RAO and YEE WHYE TEH, Gatsby Computational Neuroscience Unit, UCL	0
M11	Learning a Small Mixture of Trees, M. PAWAN KUMAR and DAPHNE KOLLER, Stanford University	1
M12	Which graphical models are difficult to learn?, ANDREA MONTANARI and JOSE AYRES PEREIRA, Stanford University	1
M13	Large Scale Nonparametric Bayesian Inference: Data Parallelisation in the Indian Buffet Process, SHAKIR MOHAMED, DAVID KNOWLES, ZOUBIN GHAHRAMANI, University of Cambridge, and FINALE DOSHI-VELEZ, Cambridge University, MIT	52
M14	Modelling Relational Data using Bayesian Clustered Tensor Factorization, ILYA SUTSKEVER, University of Toronto, RUSLAN SALAKHUTDINOV, Department of Brain and Cognitive Sciences and CSAIL MIT, and JOSHUA TENENBAUM, Massachusetts Institute of Technology	52
M15	Nonparametric Latent Feature Models for Link Prediction, KURT MILLER, UC Berkeley, THOMAS GRIFFITHS, University of California, Berkeley, and MICHAEL JORDAN, University of California	3

M16	Asymptotic	Analysis	of 1	MAP Estir	nation via	a the	Replica	Method	
	and Compr	essed Sens	ing	, SUNDEEP	RANGAN,	Qualco	mm, Inc.,	Alyson	
	Fletcher,	University	of	California,	Berkeley,	and	Vivek	Goyal,	
	Massachusetts	Institute of	Tech	nology	••••••	•••••			. 63

- M17 Sequential effects reflect parallel learning of multiple environmental regularities, MATTHEW WILDER, University of Colorado Boulder, MATT JONES and MICHAEL MOZER, University of Colorado at Boulder . 64
- M18 Noise Characterization, Modeling, and Reduction for In Vivo Neural Recording, ZHI YANG, QI ZHAO, UC Santa Cruz, EDWARD KEEFER, UT Southwestern Medical Center, and WENTAI LIU, UC Santa Cruz..... 64

- M22 fMRI-Based Inter-Subject Cortical Alignment Using Functional Connectivity, BRYAN CONROY, BEN SINGER, Princeton University, JAMES HAXBY, Dartmouth College, and PETER RAMADGE, Princeton University 67

M27	Unsupervised Feature Selection for the k-means Clustering Problem, CHRISTOS BOUTSIDIS, Rensselaer Polytechnic Instit., MICHAEL MAHONEY, Stanford University, and PETROS DRINEAS, Rensselaer Polytechnic Institute
M28	Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization, JOHN WRIGHT, Microsoft Research, ARVIND GANESH, University of Illinois, SHANKAR RAO, HRL Laboratories, YIGANG PENG, Microsoft Research Asia, and YI MA, University of Illinois at Urbana-Champaign
M29	Breaking Boundaries Between Induction Time and Diagnosis Time Active Information Acquisition, ASHISH KAPOOR and ERIC HORVITZ, Microsoft Research
M30	Manifold Embeddings for Model-Based Reinforcement Learning under Partial Observability, KEITH BUSH and JOELLE PINEAU, McGill University
M31	Modeling Social Annotation Data with Content Relevance using a Topic Model, TOMOHARU IWATA, NTT Communication Science Laboratories, TAKESHI YAMADA, NTT, and NAONORI UEDA, NTT Communication Science Laboratories
M32	Rank-Approximate Nearest Neighbor Search: Retaining Meaning and Speed in High Dimensions, PARIKSHIT RAM, Georgia Institute of Technolog, DONGRYEOL LEE, HUA OUYANG and ALEXANDER GRAY, Georgia Institute of Technology
M33	Distribution Matching for Transduction, NOVI QUADRIANTO, SML-NICTA and RSISE-ANU, JAMES PETTERSON, NICTA, and ALEX SMOLA, Yahoo!
M34	Heavy-Tailed Symmetric Stochastic Neighbor Embedding, Zhirong Yang, Cuhk, TKK, Irwin King, Zenglin Xu, Cuhk, and Erkki Oja, TKK. 73
M35	A Rate Distortion Approach for Semi-Supervised Conditional Random Fields, YANG WANG, GHOLAMREZA HAFFARI, Simon Fraser University, SHAOJUN WANG, Wright State University, and GREG MORI, Simon Fraser U
M36	Optimal Scoring for Unsupervised Learning, ZHIHUA ZHANG and GUANG DAI, Zhejiang University
M37	An Online Algorithm for Large Scale Image Similarity Learning, GAL CHECHIK, Stanford university, URI SHALIT, The Hebrew University, VARUN SHARMA and SAMY BENGIO, Google
M38	Rethinking LDA: Why Priors Matter, HANNA WALLACH, DAVID

M39	Semi-supervised Learning using Sparse Eigenfunction Bases, KAUSHIK SINHA and MIKHAIL BELKIN, Ohio State University	,)
M40	Streaming k-means approximation, NIR AILON, Google, RAGESH JAISWAL and CLAIRE MONTELEONI, Columbia University	5
M41	Polynomial Semantic Indexing, BING BAI, NEC Labs America, INC, JASON WESTON, NEC Labs America, DAVID GRANGIER, RONAN COLLOBERT, KUNIHIKO SADAMASA, YANJUN QI, NEC Labs America, INC, CORINNA CORTES, Google Research, and MEHRYAR MOHRI, Courant Institute of Mathematical Sciences and Google Research	Ĵ
M42	Matrix Completion from Power-Law Distributed Samples, RAGHU MEKA, PRATEEK JAIN, University of Texas at Austin, and INDERJIT DHILLON, University of Texas	,
M43	Learning from Neighboring Strokes: Combining Appearance and Context for Multi-Domain Sketch Recognition, TOM OUYANG and RANDALL DAVIS, MIT CSAIL	,
M44	Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise, JACOB WHITEHILL, PAUL RUVOLO, TING-FAN WU, JACOB BERGSMA, UCSD, and JAVIER MOVELLAN, University of California	8
M45	Individuation, Identification and Object Discovery, CHARLES KEMP, ALAN JERN, Carnegie Mellon University, and FEI XU, University of California, Berkeley	3
M46	The Wisdom of Crowds in the Recollection of Order Information, MARK STEYVERS, MICHAEL LEE, University of California, BRENT MILLER and PERNILLE HEMMER, University of California, Irvine)
M47	Measuring model complexity with the prior predictive, WOLF VANPAEMEL, Kuleuven)
M48	Structured output regression for detection with partial truncation, ANDREA VEDALDI and ANDREW ZISSERMAN, University of Oxford)
M49	Dual Averaging Method for Regularized Stochastic Learning and Online Optimization, LIN XIAO, Microsoft Research)
M50	Generalization Errors and Learning Curves for Regression with Multi-task Gaussian Processes, KIAN MING CHAI, University of Edinburgh)
M51	DUOL: A Double Updating Approach for Online Learning, PEILIN ZHAO, STEVEN CHU-HONG HOI, Nanyang Technological University, and RONG JIN, Michigan State University	_

M52 Monte Carlo Sampling for Regret Minimization in Extensive Games, MARC LANCTOT, University of Alberta, KEVIN WAUGH, Carnegie

	Mellon University, MARTIN ZINKEVICH, Yahoo!, and MICHAEL BOWLING, University of Alberta	81
M53	Beyond Convexity: Online Submodular Minimization, ELAD HAZAN, IBM, and SATYEN KALE, Yahoo! Research	82
M54	Asymptotically Optimal Regularization in Smooth Parametric Models, PERCY LIANG, University of California, FRANCIS BACH, Ecole Normale Superieure, GUILLAUME BOUCHARD, Xerox Research Centre Europe, and MICHAEL JORDAN, University of California	82
M55	Compressed Least-Squares Regression, ODALRIC MAILLARD and REMI MUNOS, INRIA	82
M56	Robust Nonparametric Regression with Metric-Space Valued Output, MATTHIAS HEIN, Saarland University	83
M57	On Invariance in Hierarchical Models, JAKE BOUVRIE, MIT, LORENZO ROSASCO, Massachusetts Institute of Technology, and TOMASO POGGIO, MIT	83
M58	Fast Learning from Non-i.i.d. Observations, INGO STEINWART, Los Alamos National Laboratory, and ANDREAS CHRISTMANN, University of Bayreuth	84
M59	Periodic Step Size Adaptation for Single Pass On-line Learning, CHUN-NAN HSU, YU-MING CHANG, HANSHEN HUANG, IIS, Academia Sinica, and YUH-JYE LEE, Department of CSIE, National Taiwan University of Science and Technology	84
M60	Nonparametric Bayesian Texture Learning and Synthesis, LONG ZHU, Massachusetts Institute of Technology, YUANAHAO CHEN, UCLA, BILL FREEMAN and ANTONIO TORRALBA, Massachusetts Institute of Technology	84
M61	The "tree-dependent components" of natural scenes are edge filters, DANIEL ZORAN, Hebrew University of Jerusalem, and YAIR WEISS, Hebrew University	85
M62	A Biologically Plausible Model for Rapid Natural Scene Identification, SENNAY GHEBREAB, STEVEN SCHOLTE, VICTOR LAMME and ARNOLD SMEULDERS, University of Amsterdam	85
M63	Learning models of object structure, JOSEPH SCHLECHT and KOBUS BARNARD, University of Arizona	86
M64	Dirichlet-Bernoulli Alignment: A Generative Model for Multi-Class Multi-Label Multi-Instance Corpora, SHUANG-HONG YANG, HONGYUAN ZHA, Georgia Tech, and BAO-GANG HU, Chinese Academy of Sciences	86
M65	Toward Provably Correct Feature Selection in Arbitrary Domains,	

SCHEDULE, MONDAY

M66	AUC optimization and the two-sample problem, STÉPHAN CLÉMENÇON, Telecom ParisTech, NICOLAS VAYATIS, ENS Cachan, and MARINE DEPECKER, Telecom ParisTech	87
M67	Learning to Rank by Optimizing NDCG Measure, HAMED VALIZADEGAN, RONG JIN, Michigan State University, RUOFEI ZHANG and JIANCHANG MAO, Yahoo!	88
M68	On Learning Rotations, RAMAN ARORA, Univ. of Wisconsin-Madison	88
M69	Zero-shot Learning with Semantic Output Codes, MARK PALATUCCI, Carnegie Mellon University, DEAN POMERLEAU, Intel Research, GEOFFREY HINTON, University of Toronto, and TOM MITCHELL, Carnegie Mellon University.	88
M70	Ranking Measures and Loss Functions in Learning to Rank, WEI CHEN, Chinese Academy of Sciences, TIE-YAN LIU, Microsoft Research Asia, YANYAN LAN, ZHI-MING MA, Chinese Academy of Sciences, and HANG LI, Microsoft Research Asia	89
M71	Nash Equilibria of Static Prediction Games, MICHAEL BRÜCKNER and TOBIAS SCHEFFER, University of Potsdam	90
M72	Sparsistent Learning of Varying-coefficient Models with Structural Changes, MLADEN KOLAR, LE SONG and ERIC XING, Carnegie Mellon University	90
M73	Multiple Incremental Decremental Learning of Support Vector Machines, MASAYUKI KARASUYAMA and ICHIRO TAKEUCHI, Nagoya Institute of Technology	90
M74	On the Algorithmics and Applications of a Mixed-norm based Kernel Learning Formulation, SAKETHA NATH JAGARLAPUDI, IIT-Bombay, DINESH G, RAMAN S, CHIRANJIB BHATTACHARYYA, Indian Institute of Science, Bangalore, AHARON BEN-TAL, Technion, Israel, and RAMAKRISHNAN K.R., Indian Institute of Science, Bangalore	91
M75	Graph-based Consensus Maximization among Multiple Supervised and Unsupervised Models, JING GAO, FENG LIANG, UIUC, WEI FAN, IBM TJ Waston Research, YIZHOU SUN and JIAWEI HAN, UIUC	91
M76	Positive Semidefinite Metric Learning with Boosting, CHUNHUA SHEN, JUNAE KIM, NICTA, LEI WANG, The Australian National University, and ANTON VAN DEN HENGEL, University of Adelaide	92
M77	Free energy score space, ALESSANDRO PERINA, MARCO CRISTANI, UMBERTO CASTELLANI, VITTORIO MURINO, University of Verona, and NEBOJSA JOJIC, Microsoft Research	93
M78	Conditional Neural Fields, JIAN PENG, LIEFENG BO and JINBO XU, TTI-C	93

M79	Distribution-Calibrated Hierarchical Classification, OFER DEKEL, Microsoft	94
M80	Learning with Compressible Priors, VOLKAN CEVHER, Rice University	94
M81	A Fast, Consistent Kernel Two-Sample Test, ARTHUR GRETTON, Carnegie Mellon University and Max Planck Institute, KENJI FUKUMIZU, The Institute of Statistical Mathematics, ZAID HARCHAOUI, Carnegie Mellon University, and BHARATH SRIPERUMBUDUR, UC San Diego	94
M82	Improving Existing Fault Recovery Policies, GUY SHANI and CHRISTOPHER MEEK, Microsoft Research	95
M83	Code-specific policy gradient rules for spiking neurons, HENNING SPREKELER, GUILLAUME HENNEQUIN and WULFRAM GERSTNER, Ecole Polytechnique Federale de Lausanne	95
M84	A Generalized Natural Actor-Critic Algorithm, TETSURO MORIMURA, IBM Research - Tokyo, EIJI UCHIBE, JUNICHIRO YOSHIMOTO and KENJI DOYA, Okinawa Institute of Science and Technology	96
M85	Convergent Temporal-Difference Learning with Arbitrary Smooth Function Approximation, HAMID MAEI, CSABA SZEPESVARI, University of Alberta, SHALABH BHATNAGAR, Indian Institute of Science, Bangalore, India, DOINA PRECUP, McGill University, DAVID SILVER and RICH SUTTON, University of Alberta	96
M86	Solving Stochastic Games, LIAM MAC DERMED and CHARLES ISBELL, Georgia Inst. of Technology	97
M87	White Functionals for Anomaly Detection in Dynamical Systems, MARCO CUTURI, Princeton University, JEAN-PHILIPPE VERT, ParisTech, and ALEXANDRE D'ASPREMONT, Princeton University	97

Tuesday, December 8th

8:30–10:10am Oral Session:

Information Theory and Estimation

- 8:30am Invited Talk: Relative Entropy, SERGIO VERDU, Princeton University 99

10:40am-12:00pm Oral Session:

Images and Codes

- 10:40am An Additive Latent Feature Model for Transparent Object Recognition, MARIO FRITZ, UC Berkeley EECS and ICSI, MICHAEL BLACK, Brown University, GARY BRADSKI, Willow Garage, SERGEY KARAYEV, UC Berkeley, and TREVOR DARRELL, UC Berkeley EECS and ICSI.. 100

12:30pm-1:45pm Publication Models Debate

12:30pm Debate on Future Publication Models for the NIPS Community, YOSHUA BENGIO, University of Montreal

2:00–3:20pm Oral Session:

Deep Learning and Network Models

3:00pm	Discriminative Network Models of Schizophrenia , GUILLERMO CECCHI, IRINA RISH, IBM T.J. Watson Research Center, BENJAMIN THYREAU, Neurospin, CEA-INSERM, BERTRAND THIRION, INRIA, MARION PLAZE, CEA-INSERM, MARIE-LAURE PAILLERE-MARTINOT, AP-HP, CATHERINE MARTELLI, Centre Hospitalier Paul Brousse, JEAN-LUC MARTINOT, INSERM - CEA - Univ. Paris Sud, and JEAN-BAPTISTE POLINE, Neurospin, CEA-INSERM	104
3:20–3:40pm	Spotlights	
	Kernels and learning curves for Gaussian process regression on random graphs, PETER SOLLICH, Kings College London, MATTHEW URRY, King's College London, and CAMILLE COTI, INRIA Saclay-Ile de France	110
	Robust Value Function Approximation Using Bilinear Programming, MAREK PETRIK, University of Massachusetts Amherst, and SHLOMO ZILBERSTEIN, University of Massachusetts at Amherst	147
	Analysis of SVM with Indefinite Kernels, YIMING YING, COLIN CAMPBELL, University of Bristol, and MARK GIROLAMI, University of Glasgow. 1	141
	Boosting with Spatial Regularization, ZHEN XIANG, Princeton University, YONGXIN XI, URI HASSON and PETER RAMADGE, Princeton University	142
	Abstraction and Relational learning, CHARLES KEMP and ALAN JERN, Carnegie Mellon University	122
	Probabilistic Relational PCA , WU-JUN LI, DIT-YAN YEUNG, HKUST, and ZHIHUA ZHANG, Zhejiang University	111
	Decoupling Sparsity and Smoothness in the Discrete Hierarchical Dirichlet Process, CHONG WANG and DAVID BLEI, Princeton University 1	112
	A joint maximum-entropy model for binary neural population patterns and continuous signals, SEBASTIAN GERWINN, PHILIPP BERENS and MATTHIAS BETHGE, MPI for Biological Cybernetics & University of Tübingen	137
	Efficient Bregman Range Search, LAWRENCE CAYTON, Max Planck Institute for Biological Cybernetics	112
	Graph Zeta Function in the Bethe Free Energy and Loopy Belief Propagation, YUSUKE WATANABE and KENJI FUKUMIZU, The Institute of Statistical Mathematics	118
	Online Learning of Assignments, MATTHEW STREETER, Google, DANIEL GOLOVIN, Carnegie Mellon University, and ANDREAS KRAUSE, California Institute of Technology	127
	Linear-time Algorithms for Pairwise Statistical Problems,	

PARIKSHIT RAM, Georgia Institute of Technolog, DONGRYEOL LEE, Georgia

4:10–5:10pm Oral Session: Cognitive Science and EEG Analysis

4:10pm Differential Use of Implicit Negative Evidence in Generative and Discriminative Language Learning, ANNE HSU, University of California Berkeley, and THOMAS GRIFFITHS, University of California, Berkeley...... 107

- 4:30pm Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model, ED VUL, MICHAEL FRANK, Massachusetts Institute of Technology, GEORGE ALVAREZ, Dept. of Psychology, Harvard University, and JOSHUA TENENBAUM, Massachusetts Institute of Technology 107
- 4:50pm Optimizing Multi-Class Spatio-Spectral Filters via Bayes Error Estimation for EEG Classification, WENMING ZHENG, Southeast University, P.R. China, and ZHOUCHEN LIN, Microsoft Research Asia, P.R. China. 108

5:10-5:30pm Spotlights

Indian Buffet Processes with Power-law Behavior, YEE WHYE TEH and DILAN GORUR, Gatsby Computational Neuroscience Unit, UCL 119

STDP	enables	spiking	neurons	to	detect	hidden	caus	es c	of their	
inputs,	Bern	ihard N	ESSLER,	MIC	hael 1	Pfeiffer	and	Wo	LFGANG	
MAASS,	Graz Univ	versity of '	Technology							. 139

3D	Object	Recogn	ition	with	Deep	Belief	Nets,	Vinoi) Nai	R and	
Gec	FFREY	Hinton,	Univer	sity of	Toront	0					144

Time-Varying	Dynamic	Bayesian	Networks,	Le Song,	Mladen	
KOLAR and ERIC	XING, Car	negie Mellon	University			125

	Training Factor Graphs with Reinforcement Learning for Efficient MAP Inference, MICHAEL WICK, KHASHAYAR ROUANNAANESH SAMEED SINCH and ANDREW MCCALLUM University of
	Massachusetts, Amherst
	Riffled Independence for Ranked Data, JONATHAN HUANG and CARLOS GUESTRIN, Carnegie Mellon University
	A Smoothed Approximate Linear Program, VIJAY DESAI, Columbia University, VIVEK FARIAS, MIT Sloan, and CIAMAC MOALLEMI, Columbia University
	Randomized Pruning: Efficiently Calculating Expectations in Large Dynamic Programs, ALEXANDRE BOUCHARD-CÔTÉ, UC Berkeley, SLAV PETROV, Google Research, and DAN KLEIN, University of California Berkeley
	Subject independent EEG-based BCI decoding, SIAMAC FAZLI, Technical University Berlin, CRISTIAN GROZEA, Fraunhofer Institute FIRST, MARTON DANOCZY, TU Berlin, BENJAMIN BLANKERTZ, Berlin Institute of Technology, FLORIN POPESCU, Fraunhofer FIRST, and KLAUS-ROBERT MULLER, Berlin Institute of Technology
7:30–11:59pm	Poster Session
T1	Kernels and learning curves for Gaussian process regression on random graphs, PETER SOLLICH, Kings College London, MATTHEW URRY, King's College London, and CAMILLE COTI, INRIA Saclay-Ile de France 110
T2	Locality-sensitive binary codes from shift-invariant kernels, MAXIM RAGINSKY, Duke University, and SVETLANA LAZEBNIK, University of North Carolina
Τ3	Probabilistic Relational PCA , WU-JUN LI, DIT-YAN YEUNG, HKUST, and ZHIHUA ZHANG, Zhejiang University
T4	Decoupling Sparsity and Smoothness in the Discrete Hierarchical Dirichlet Process, CHONG WANG and DAVID BLEI, Princeton University 112
T5	<i>Efficient Bregman Range Search</i> , LAWRENCE CAYTON, Max Planck Institute for Biological Cybernetics
Τ6	Learning Bregman Distance Functions and Its Application for Semi-Supervised Clustering, LEI WU, Nanyang Technological University, RONG JIN, Michigan State University, STEVEN CHU-HONG HOI, Nanyang Technological University, JIANKE ZHU, ETH Zurich, and NENGHAI YU, University of Science and Technology of China
T7	Linearly constrained Bayesian matrix factorization for blind source separation, MIKKEL SCHMIDT, University of Cambridge

T 8	Matrix Completion from Noisy Entries, RAGHUNANDAN KESHAVAN, ANDREA MONTANARI and SEWOONG OH, Stanford University
Τ9	Nonlinear Learning using Local Coordinate Coding, KAI YU, NEC Laboratories, TONG ZHANG, Rutgers, and YIHONG GONG, NEC Laboratories America
T10	Semi-supervised Regression using Hessian energy with an application to semi-supervised dimensionality reduction, KWANG IN KIM, Saarland University, FLORIAN STEINKE, Siemens Corporate Technology, Munich, Germany, and MATTHIAS HEIN, Saarland University 114
T 11	Group Sparse Coding, SAMY BENGIO, FERNANDO PEREIRA, YORAM SINGER and DENNIS STRELOW, Google
T12	Label Selection on Graphs, ANDREW GUILLORY and JEFF BILMES, University of Washington
T13	Local Rules for Global MAP: When Do They Work ?, KYOMIN JUNG, KAIST, PUSHMEET KOHLI, Microsoft, and DEVAVRAT SHAH, MIT 116
T14	Non-Parametric Bayesian Dictionary Learning for Sparse Image Representations, MINGYUAN ZHOU, Duke ECE, HAOJUN CHEN, Duke University, JOHN PAISLEY, LU REN, Duke ECE, GUILLERMO SAPIRO, University of Minnesota, and LAWRENCE CARIN, Duke ECE
T15	Exponential Family Graph Matching and Ranking, JAMES PETTERSON, TIBERIO CAETANO, JULIAN MCAULEY, NICTA, and JIN YU, Nicta
T16	Sensitivity analysis in HMMs with application to likelihood maximization, PIERRE-ARNAUD COQUELIN, Vekia, ROMAIN DEGUEST, Columbia University, and REMI MUNOS, INRIA
T17	A Gaussian Tree Approximation for Integer Least-Squares, JACOB GOLDBERGER, Bar-Ilan U, and AMIR LESHEM, Bar-Ilan University 117
T18	Conditional Random Fields with High-Order Features for Sequence Labeling, NAN YE, National University of Singapo, WEE SUN LEE, National University of Singapore, HAI LEONG CHIEU, DSO National Laboratories, and DAN WU, Singapore MIT Alliance
T19	Graph Zeta Function in the Bethe Free Energy and Loopy Belief Propagation, YUSUKE WATANABE and KENJI FUKUMIZU, The Institute of Statistical Mathematics
T20	Indian Buffet Processes with Power-law Behavior, YEE WHYE TEH and DILAN GORUR, Gatsby Computational Neuroscience Unit, UCL 119
T2 1	Inter-domain Gaussian Processes for Sparse Inference using Inducing Features, MIGUEL LAZARO-GREDILLA and ANIBAL FIGUEIRAS-VIDAL, Universidad Carlos III Madrid

T22	Bayesian Sparse Factor Models and DAGs Inference and Comparison, RICARDO HENAO and OLE WINTHER, Technical University of Denmark	120
T23	Learning in Markov Random Fields using Tempered Transitions, RUSLAN SALAKHUTDINOV, Department of Brain and Cognitive Sciences and CSAIL MIT	120
T24	Maximum likelihood trajectories for continuous-time Markov chains, THEODORE PERKINS, Ottawa Hospital Research Inst.	120
T25	Riffled Independence for Ranked Data, JONATHAN HUANG and CARLOS GUESTRIN, Carnegie Mellon University	121
T26	Nonlinear directed acyclic structure learning with weakly additive noise models, ROBERT TILLMAN, Carnegie Mellon University, ARTHUR GRETTON, Carnegie Mellon and MPI for Biological Cybernetics, and PETER SPIRTES, Carnegie Mellon	121
T27	Abstraction and Relational learning, CHARLES KEMP and ALAN JERN, Carnegie Mellon University	122
T28	Human Rademacher Complexity, XIAOJIN ZHU, TIMOTHY ROGERS, University of Wisconsin-Madison, and BRYAN GIBSON, University of Wisconsin	122
T29	Parallel Inference for Latent Dirichlet Allocation on Graphics Processing Units, FENG YAN, Purdue University, NINGYI XU, Microsoft Research Asia, and YUAN QI, Purdue University	123
T30	Differential Use of Implicit Negative Evidence in Generative and Discriminative Language Learning, ANNE HSU, University of California Berkeley, and THOMAS GRIFFITHS, University of California, Berkeley	107
T31	Adaptive Design Optimization in Experiments with People, DANIEL CAVAGNARO, The Ohio State University, MARK PITT and JAY MYUNG, Ohio State University	123
T32	Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model, ED VUL, MICHAEL FRANK, Massachusetts Institute of Technology, GEORGE ALVAREZ, Dept. of Psychology, Harvard University, and JOSHUA TENENBAUM, Massachusetts Institute of Technology	107
T33	Factor Modeling for Advertisement Targeting, YE CHEN, EBay Research Labs, MICHAEL KAPRALOV, Stanford University, DMITRY PAVLOV, Yandex Labs, and JOHN CANNY, University of California, Berkeley	124
T3 4	Clustering sequence sets for motif discovery, Jong Kyoung Kim and Seungjin Choi, POSTECH	125
T35	<i>Time-Varying Dynamic Bayesian Networks,</i> LE SONG, MLADEN KOLAR and ERIC XING, Carnegie Mellon University	125

SCHEDULE, TUESDAY

T36	Randomized Pruning: Efficiently Calculating Expectations in Large Dynamic Programs, ALEXANDRE BOUCHARD-CÔTÉ, UC Berkeley, SLAV PETROV, Google Research, and DAN KLEIN, University of California Berkeley
T38	Multi-Label Prediction via Compressed Sensing, DANIEL HSU, University of California, SHAM KAKADE, Toyota Technological Institute, JOHN LANGFORD, Yahoo! Research, and TONG ZHANG, Rutgers University
T39	Data-driven calibration of linear estimators with minimal penalties, SYLVAIN ARLOT, CNRS - ENS, and FRANCIS BACH, Ecole Normale Superieure
T40	Regularized Distance Metric Learning: Theory and Algorithm, RONG JIN, Michigan State University, SHIJUN WANG, National Institutes of Health, and YANG ZHOU, Michigan State University
T41	Online Learning of Assignments, MATTHEW STREETER, Google, DANIEL GOLOVIN, Carnegie Mellon University, and ANDREAS KRAUSE, California Institute of Technology
T42	A General Projection Property for Distribution Families, YAO- LIANG YU, YUXI LI, DALE SCHUURMANS and CSABA SZEPESVARI, University of Alberta
T43	Submodularity Cuts and Applications, YOSHINOBU KAWAHARA, Osaka University, KIYOHITO NAGANO, Tokyo Institute of Technology, KOJI TSUDA, Advanced Industrial Science and Technology, and JEFF BILMES, University of Washington
T44	Linear-time Algorithms for Pairwise Statistical Problems, PARIKSHIT RAM, Georgia Institute of Technolog, DONGRYEOL LEE, Georgia Institute of Technology, WILLIAM MARCH, Georgia Tech, and ALEXANDER GRAY, Georgia Institute of Technology
T 45	On the Convergence of the Concave-Convex Procedure, BHARATH SRIPERUMBUDUR, UC San Diego, and GERT LANCKRIET, University of California
T46	Slow Learners are Fast, MARTIN ZINKEVICH, ALEX SMOLA and JOHN LANGFORD, Yahoo!
T47	Accelerated Gradient Methods for Stochastic Optimization and Online Learning, CHONGHAI HU, JAMES KWOK and WEIKE PAN, Hong Kong University of Science and Technology
T48	Noisy Generalized Binary Search, ROB NOWAK, University of Wisconsin, Madison
T49	Thresholding Procedures for High Dimensional Variable Selection and Statistical Estimation, SHUHENG ZHOU, ETH Zurich

T50	Construction of Nonparametric Bayesian Models from Parametric Bayes Equations, PETER ORBANZ, University of Cambridge
T51	Adapting to the Shifting Intent of Search Queries, UMAR SYED, Princeton University, ALEKSANDRS SLIVKINS and NINA MISHRA, Microsoft Research
T52	A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers, SAHAND NEGAHBAN, UC, Berkeley, PRADEEP RAVIKUMAR, MARTIN WAINWRIGHT and BIN YU, University of California
T53	Beyond Categories: The Visual Memex Model for Reasoning About Object Relationships, TOMASZ MALISIEWICZ and ALYOSHA EFROS, Carnegie Mellon University
T54	L ₁ -Penalized Robust Estimation for a Class of Inverse Problems Arising in Multiview Geometry, ARNAK DALALYAN and RENAUD KERIVEN, Université Paris-Est, ENPC
T55	An Additive Latent Feature Model for Transparent Object Recognition, MARIO FRITZ, UC Berkeley EECS and ICSI, MICHAEL BLACK, Brown University, GARY BRADSKI, Willow Garage, SERGEY KARAYEV, UC Berkeley, and TREVOR DARRELL, UC Berkeley EECS and ICSI 100
T56	Semi-Supervised Learning in Gigantic Image Collections, ROB FERGUS, New York University, YAIR WEISS, Hebrew University, and ANTONIO TORRALBA, Massachusetts Institute of Technology
T57	Speaker Comparison with Inner Product Discriminant Functions, WILLIAM CAMPBELL, MIT Lincoln Laboratory, ZAHI KARAM, MIT, and DOUGLAS STURIM, MIT Lincoln Laboratory
T58	Efficient Recovery of Jointly Sparse Vectors, LIANG SUN, JUN LIU, JIANHUI CHEN and JIEPING YE, Arizona State University
T59	Extending Phase Mechanism to Differential Motion Opponency for Motion Pop-out, YICONG MENG and BERTRAM SHI, Hong Kong University of Science and Technology
T60	An Integer Projected Fixed Point Method for Graph Matching and MAP Inference, MARIUS LEORDEANU, Carnegie Mellon University, MARTIAL HEBERT, CMU, and RAHUL SUKTHANKAR, Intel
T61	Filtering Abstract Senses From Image Search Results, KATE SAENKO, MIT, and TREVOR DARRELL, UC Berkeley EECS and ICSI 136
T62	Optimizing Multi-Class Spatio-Spectral Filters via Bayes Error Estimation for EEG Classification, WENMING ZHENG, Southeast University, P.R. China, and ZHOUCHEN LIN, Microsoft Research Asia, P.R. China. 108

SCHEDULE, TUESDAY

T63	No evidence for active sparsification in the visual cortex, PIETRO BERKES, BEN WHITE and JOZSEF FISER, Brandeis University
T64	A joint maximum-entropy model for binary neural population patterns and continuous signals, SEBASTIAN GERWINN, PHILIPP BERENS and MATTHIAS BETHGE, MPI for Biological Cybernetics & University of Tübingen
T65	Unsupervised feature learning for audio classification using convolutional deep belief networks, HONGLAK LEE, PETER PHAM, YAN LARGMAN and ANDREW NG, Stanford University
T66	Hierarchical Mixture of Classification Experts Uncovers Interactions between Brain Regions, BANGPENG YAO, Stanford University, DIRK WALTHER, DIANE BECK, University of Illinois, and LI FEI-FEI, Princeton University
T67	Statistical Models of Linear and Nonlinear Contextual Interactions in Early Visual Processing, RUBEN COEN-CAGLI, Albert Einstein College of Medicine, PETER DAYAN, University College London, and ODELIA SCHWARTZ, Albert Einstein College of Medicine
T68	Discriminative Network Models of Schizophrenia, GUILLERMO CECCHI, IRINA RISH, IBM T.J. Watson Research Center, BENJAMIN THYREAU, Neurospin, CEA-INSERM, BERTRAND THIRION, INRIA, MARION PLAZE, CEA-INSERM, MARIE-LAURE PAILLERE-MARTINOT, AP-HP, CATHERINE MARTELLI, Centre Hospitalier Paul Brousse, JEAN-LUC MARTINOT, INSERM - CEA - Univ. Paris Sud, and JEAN-BAPTISTE POLINE, Neurospin, CEA-INSERM
T69	STDP enables spiking neurons to detect hidden causes of their inputs, BERNHARD NESSLER, MICHAEL PFEIFFER and WOLFGANG MAASS, Graz University of Technology
T7 0	A Neural Implementation of the Kalman Filter, ROBERT WILSON, Princeton University, and LEIF FINKEL, University of Pennsylvania
T71	Speeding up Magnetic Resonance Image Acquisition by Bayesian Multi-Slice Adaptive Compressed Sensing, MATTHIAS SEEGER, Saarland University and Max Planck Institute for Informatics
T72	Subject independent EEG-based BCI decoding, SIAMAC FAZLI, Technical University Berlin, CRISTIAN GROZEA, Fraunhofer Institute FIRST, MARTON DANOCZY, TU Berlin, BENJAMIN BLANKERTZ, Berlin Institute of Technology, FLORIN POPESCU, Fraunhofer FIRST, and KLAUS-ROBERT MULLER, Berlin Institute of Technology
T73	Analysis of SVM with Indefinite Kernels, YIMING YING, COLIN CAMPBELL, University of Bristol, and MARK GIROLAMI, University of Glasgow. 141
T74	Boosting with Spatial Regularization, ZHEN XIANG, Princeton University, YONGXIN XI, URI HASSON and PETER RAMADGE, Princeton University

T75	Directed Regression, YI-HAO KAO, BENJAMIN VAN ROY and XIANG YAN, Stanford University	142
T76	Measuring Invariances in Deep Networks, IAN GOODFELLOW, QUOC LE, ANDREW SAXE and ANDREW NG, Stanford University	143
$\mathbf{T77}$	Nonparametric Greedy Algorithms for the Sparse Learning Problem, HAN LIU and XI CHEN, Carnegie Mellon University	143
T78	Learning Non-Linear Combinations of Kernels, CORINNA CORTES, Google Research, MEHRYAR MOHRI, Courant Institute of Mathematical Sciences and Google Research, and AFSHIN ROSTAMIZADEH, Courant Institute, New York University	144
T79	3D Object Recognition with Deep Belief Nets, VINOD NAIR and GEOFFREY HINTON, University of Toronto	144
T80	Grouped Orthogonal Matching Pursuit for Variable Selection and Prediction, AURELIE LOZANO, IBM T.J. Watson Research, GRZEGORZ SWIRSZCZ, IBM Research, and NAOKI ABE, IBM T. J. Watson Research Center	145
T81	Efficient and Accurate Lp-Norm Multiple Kernel Learning, MARIUS KLOFT, Technical University of Berlin, ULF BREFELD, TU Berlin, SOEREN SONNENBURG, Friedrich Miescher Laboratory, Max Planck Society, PAVEL LASKOV, University of Tuebingen, KLAUS-ROBERT MÜLLER, Technical University of Berlin, and ALEXANDER ZIEN, LIFE Biosystems	145
T82	Training Factor Graphs with Reinforcement Learning for Efficient MAP Inference, MICHAEL WICK, KHASHAYAR ROHANIMANESH, SAMEER SINGH and ANDREW MCCALLUM, University of Massachusetts, Amherst	146
T83	Fast, smooth and adaptive regression in metric spaces, SAMORY KPOTUFE, UCSD CSE	146
T84	RobustValueFunctionApproximationUsingBilinearProgramming,MAREK PETRIK, University of Massachusetts Amherst, andSHLOMOZILBERSTEIN, University of Massachusetts at Amherst	147
T85	The Infinite Partially Observable Markov Decision Process, FINALE DOSHI-VELEZ, Cambridge University, MIT	147
T86	A Smoothed Approximate Linear Program , VIJAY DESAI, Columbia University, VIVEK FARIAS, MIT Sloan, and CIAMAC MOALLEMI, Columbia University	147
T 87	Learning to Explore and Exploit in POMDPs, CHENGHUI CAI, XUEJUN LIAO, Duke University, and LAWRENCE CARIN, Duke ECE	148
T88	Compositionality of optimal control laws, EMANUEL TODOROV, UW	148

SCHEDULE, TUESDAY

T89	Constructing Topological Maps using Markov Random Fields and Loop-Closure Detection, ROY ANATI and KOSTAS DANIILIDIS, University of Pennsylvania
7:30–11:59pm	Demonstrations
D1	Activity Recognition and Human Gait Recognition, JORDAN FRANK, McGill University
D2	Automatically Adaptable Spike Sorting Software, JORGE GAMEZ DE LEON, U.N.A.M
D3	Car Detector trained with the Quantum Adiabatic Algorithm, HARTMUT NEVEN, Google
D4	Open-access datasets for time series causality discovery validation, ISABELLE GUYON, Clopinet
D5	Demonstration of the TreeRank Software , MARINE DEPECKER, Renault SA-Telecom ParisTech
D6	The IID: A Natively Probabilistic Reconfigurable Computer, VIKASH MANSINGHKA, Massachusetts Institute of Technology
D7	Vision Toolkit for Image Recognition, JAMIE NIEMASIK, Numenta 151

Wednesday, December 9th

8:30–10:10am	Oral	Session:
	Neuro	oscience

8:30am	Invited Talk: The Rat Vibrissal Array as a Model Sensorimotor System, MITRA HARTMANN, Northwestern University
9:30am	Functional network reorganization in motor cortex can be explained by reward-modulated Hebbian learning, ROBERT LEGENSTEIN, Graz University of Technology, STEVEN CHASE, Carnegie Mellon University, ANDREW SCHWARTZ, University of Pittsburgh, and WOLFGANG MAASS, Graz University of Technology
9:50am	Know Thy Neighbour: A Normative Theory of Synaptic Depression, JEAN-PASCAL PFISTER, University of Cambridge, PETER DAYAN, University College London, and MATE LENGYEL, University of Cambridge
10:40am–12:00pm	Oral Session: Theory, Optimization and Games
10:40am	Efficient Learning using Forward-Backward Splitting, JOHN DUCHI, UC Berkeley, and YORAM SINGER, Google
11:00am	On Stochastic and Worst-case Models for Investing, ELAD HAZAN, IBM, and SATYEN KALE, Yahoo! Research
11:20am	Bootstrapping from Game Tree Search, JOEL VENESS, UNSW / NICTA, DAVID SILVER, University of Alberta, WILLIAM UTHER, NICTA / University of New South Wales, and ALAN BLAIR, NICTA / UNSW
11:40am	Kernel Choice and Classifiability for RKHS Embeddings of Probability Distributions, BHARATH SRIPERUMBUDUR, UC San Diego, KENJI FUKUMIZU, The Institute of Statistical Mathematics, ARTHUR GRETTON, Carnegie Mellon University and Max Planck Institute, GERT LANCKRIET, University of California, and BERNHARD SCHOELKOPF, MPI for Biological Cybernetics
$2:00-3:20 \mathrm{pm}$	Oral Session: Vision and Inference
2:00pm	Invited Talk: Learning and Inference in Low-Level Vision, YAIR WEISS, Hebrew University
3:00pm	An LP View of the M-best MAP problem, MENACHEM FROMER and AMIR GLOBERSON, Hebrew University

3:20–3:40pm Spotlights

Quantification and the language of thought, CHARLES KEMP, Carnegie Mellon University
Fast Image Deconvolution using Hyper-Laplacian Priors, DILIP KRISHNAN and ROB FERGUS, New York University
Efficient Large-Scale Distributed Training of Conditional Maximum Entropy Models, Gideon Mann, Ryan McDonald, Google, Mehryar Mohri, Google Research, Nathan Silberman, Google Inc., and Dan Walker, New York University
A Data-Driven Approach to Modeling Choice, VIVEK FARIAS, MIT Sloan, SRIKANTH JAGABATHULA and DEVAVRAT SHAH, MIT
Statistical Analysis of Semi-Supervised Learning: The Limit of Infinite Unlabelled Data, BOAZ NADLER, Weizmann Institute of Science, NATHAN SREBRO, Toyota Technological Institute at Chicago, and XUEYUAN ZHOU, University of Chicago
Region-based Segmentation and Object Detection, STEPHEN GOULD, TIANSHI GAO and DAPHNE KOLLER, Stanford University
Anomaly Detection with Score functions based on Nearest Neighbor Graphs, MANQI ZHAO and VENKATESH SALIGRAMA, Boston University
Adaptive Regularization of Weight Vectors, KOBY CRAMMER, The Technion, ALEX KULESZA, University of Pennsylvania, and MARK DREDZE, Johns Hopkins University
Adaptive Regularization for Transductive Support Vector Machine, ZENGLIN XU, CUHK, RONG JIN, Michigan State University, JIANKE ZHU, ETH Zurich, IRWIN KING, CUHK, MICHAEL LYU, The Chinese University of HK, and ZHIRONG YANG, CUHK, TKK
Variational Gaussian-process factor analysis for modeling spatio-temporal data, JAAKKO LUTTINEN and ALEXANDER ILIN, Helsinki University of Technology
Predicting the Optimal Spacing of Study: A Multiscale Context Model of Memory, MICHAEL MOZER, University of Colorado at Boulder, HAROLD PASHLER, UC San Diego, NICHOLAS CEPEDA, York University, ROBERT LINDSEY, University of Colorado, and ED VUL, Massachusetts Institute of Technology
Fast Graph Laplacian Regularized Kernel Learning via Semidefinite-Quadratic-Linear Programming, XIAO-MING WU, ANTHONY MAN-CHO SO, ZHENGUO LI and SHUO-YEN ROBERT LI, The Chinese University of Hong Kong

4:10–5:10pm
4:10pm
4:30pm
4:50pm
5:10–5:30pm

	Hierarchical Learning of Dimensional Biases in Human Categorization, KATHERINE HELLER, University of Cambridge, ADAM SANBORN and NICK CHATER, University College London			
	Lower bounds on minimax rates for nonparametric regression with additive sparsity and smoothness, GARVESH RASKUTTI, UC Berkeley, MARTIN WAINWRIGHT and BIN YU, University of California			
	Neural Implementation of Hierarchical Bayesian Inference by Importance Sampling, LEI SHI, UC Berkeley, and THOMAS GRIFFITHS, University of California, Berkeley			
	Posterior vs Parameter Sparsity in Latent Variable Models, JOAO GRACA, University of Lisbon, KUZMAN GANCHEV, University of Pennsylvania, BEN TASKAR, University of Pennsylvania, and FERNANDO PEREIRA, Google 189			
	Structural inference affects depth perception in the context of potential occlusion, IAN STEVENSON and KONRAD KOERDING, Northwestern University			
	Bilinear classifiers for visual recognition , HAMED PIRSIAVASH, DEVA RAMANAN and CHARLESS FOWLKES, UC Irvine			
7:30–11:59pm	30–11:59pm Poster Session			
W1	Functional network reorganization in motor cortex can be explained by reward-modulated Hebbian learning, ROBERT LEGENSTEIN, Graz University of Technology, STEVEN CHASE, Carnegie Mellon University, ANDREW SCHWARTZ, University of Pittsburgh, and WOLFGANG MAASS, Graz University of Technology			
W2	Efficient Large-Scale Distributed Training of Conditional Maximum Entropy Models, GIDEON MANN, RYAN MCDONALD, Google, MEHRYAR MOHRI, Google Research, NATHAN SILBERMAN, Google Inc., and DAN WALKER, New York University			
W3	Ensemble Nystrom Method, SANJIV KUMAR, Google Research NY, MEHRYAR MOHRI, Courant Institute of Mathematical Sciences and Google Research, and AMEET TALWALKAR, Courant Institute (NYU)			
W4	Learning Label Embeddings for Nearest-Neighbor Multi-class Classification with an Application to Speech Recognition, NATASHA SINGH-MILLER, MIT, and MICHAEL COLLINS, Massachusetts Institute of Technology			
W5	From PAC-Bayes Bounds to KL Regularization, PASCAL GERMAIN, ALEXANDRE LACASSE, Laval University, FRANCOIS LAVIOLETTE, MARIO MARCHAND, Université Laval, and SARA SHANIAN, Laval University 164			
W6	Fast subtree kernels on graphs, NINO SHERVASHIDZE and KARSTEN BORGWARDT, MPIs Tuebingen			

W7	Adaptive Regularization of Weight Vectors, KOBY CRAMMER, The Technion, ALEX KULESZA, University of Pennsylvania, and MARK DREDZE, Johns Hopkins University	165
W8	Kernel Methods for Deep Learning, YOUNGMIN CHO, University of California, San Diego, and LAWRENCE SAUL, University of California	165
W9	Lattice Regression, ERIC GARCIA and MAYA GUPTA, University of Washington	166
W10	Slow, Decorrelated Features for Pretraining Complex Cell-like Networks, JAMES BERGSTRA and YOSHUA BENGIO, University of Montreal	166
W11	Orthogonal Matching Pursuit From Noisy Random Measurements: A New Analysis, ALYSON FLETCHER, University of California, Berkeley, and SUNDEEP RANGAN, Qualcomm, Inc.	166
W12	Sparse Estimation Using General Likelihoods and Non-Factorial Priors, DAVID WIPF and SRIKANTAN NAGARAJAN, UC San Francisco	167
W13	Efficient Learning using Forward-Backward Splitting, JOHN DUCHI, UC Berkeley, and YORAM SINGER, Google	155
W14	A Data-Driven Approach to Modeling Choice, VIVEK FARIAS, MIT Sloan, SRIKANTH JAGABATHULA and DEVAVRAT SHAH, MIT	168
W15	Unsupervised Detection of Regions of Interest Using Iterative Link Analysis, GUNHEE KIM, Carnegie Mellon University, and ANTONIO TORRALBA, Massachusetts Institute of Technology	168
W16	Maximin affinity learning of image segmentation, SRINIVAS TURAGA, MIT, KEVIN BRIGGMAN, MORITZ HELMSTAEDTER, WINFRIED DENK, Max-Planck Institute for Medical Research, and SEBASTIAN SEUNG, MIT.	168
W17	Evaluating multi-class learning strategies in a generative hierarchical framework for object detection, SANJA FIDLER, MARKO BOBEN and ALES LEONARDIS, University of Ljubljana	169
W18	Hierarchical Modeling of Local Image Features through L_p -Nested Symmetric Distributions, FABIAN SINZ, MPI for Biological Cybernetics & University of Tübingen, EERO SIMONCELLI, New York University, and MATTHIAS BETHGE, MPI for Biological Cybernetics & University of Tübingen	169
W19	Fast Image Deconvolution using Hyper-Laplacian Priors, DILIP KRISHNAN and ROB FERGUS, New York University	170
W20	The Ordered Residual Kernel for Robust Motion Subspace Clustering, TAT-JUN CHIN, HANZI WANG and DAVID SUTER, The University of Adelaide	170
W21	Region-based Segmentation and Object Detection, STEPHEN GOULD, TIANSHI GAO and DAPHNE KOLLER, Stanford University	171

SCHEDULE, WEDNESDAY

W22	Efficient Match Kernel between Sets of Features for Visual Recognition, LIEFENG BO, Toyota Technological Institute at Chicago (TTI-C), and CRISTIAN SMINCHISESCU, U Bonn
W23	Segmenting Scenes by Matching Image Composites, BRYAN RUSSELL, INRIA - ENS Paris, ALYOSHA EFROS, Carnegie Mellon University, JOSEF SIVIC, INRIA ENS, BILL FREEMAN, Massachusetts Institute of Technology, and ANDREW ZISSERMAN, University of Oxford
W24	Structural inference affects depth perception in the context of potential occlusion, IAN STEVENSON and KONRAD KOERDING, Northwestern University
W25	Who's Doing What: Joint Modeling of Names and Verbs for Simultaneous Face and Pose Annotation, JIE LUO, Idiap / EPF Lausanne, BARBARA CAPUTO, Idiap Research Institute, and VITTORIO FERRARI, ETH Zurich
W26	Bilinear classifiers for visual recognition , HAMED PIRSIAVASH, DEVA RAMANAN and CHARLESS FOWLKES, UC Irvine
W27	Neurometric function analysis of population codes, PHILIPP BERENS, SEBASTIAN GERWINN, ALEXANDER ECKER and MATTHIAS BETHGE, MPI for Biological Cybernetics & University of Tübingen
W28	Time-rescaling methods for the estimation and assessment of non- Poisson neural encoding models, JONATHAN PILLOW, University of Texas at Austin
W29	Know Thy Neighbour: A Normative Theory of Synaptic Depression, JEAN-PASCAL PFISTER, University of Cambridge, PETER DAYAN, University College London, and MATE LENGYEL, University of Cambridge
W30	Learning Brain Connectivity of Alzheimer's Disease from Neuroimaging Data, SHUAI HUANG, Arizona State University, JING LI, jing.li.8@asu.edu, LIANG SUN, JUN LIU, TERESA WU, Arizona State University, KEWEI CHEN, ADAM FLEISHER, ERIC REIMAN, Banner Alzheimer?s Institute and Banner PET Center, and JIEPING YE, Arizona State University
W31	Neural Implementation of Hierarchical Bayesian Inference by Importance Sampling, LEI SHI, UC Berkeley, and THOMAS GRIFFITHS, University of California, Berkeley
W32	Exploring Functional Connectivities of the Human Brain using Multivariate Information Analysis, BARRY CHAI, Stanford, DIRK WALTHER, DIANE BECK, University of Illinois, and LI FEI-FEI, Princeton University
W33	Quantification and the language of thought, CHARLES KEMP, Carnegie Mellon University

W34	Bayesian Belief Polarization, ALAN JERN, KAI-MIN CHANG and CHARLES KEMP, Carnegie Mellon University
W35	Predicting the Optimal Spacing of Study: A Multiscale Context Model of Memory, MICHAEL MOZER, University of Colorado at Boulder, HAROLD PASHLER, UC San Diego, NICHOLAS CEPEDA, York University, ROBERT LINDSEY, University of Colorado, and ED VUL, Massachusetts Institute of Technology
W36	Perceptual Multistability as Markov Chain Monte Carlo Inference, SAMUEL GERSHMAN, Princeton University, ED VUL and JOSHUA TENENBAUM, Massachusetts Institute of Technology
W37	Multi-Step Dyna Planning for Policy Evaluation and Control, HENGSHUAI YAO, University of alberta, RICH SUTTON, University of Alberta, SHALABH BHATNAGAR, Indian Institute of Science, DONGCUI DIAO, South China Normal University, and CSABA SZEPESVARI, University of Alberta
W38	Bootstrapping from Game Tree Search, JOEL VENESS, UNSW / NICTA, DAVID SILVER, University of Alberta, WILLIAM UTHER, NICTA / University of New South Wales, and ALAN BLAIR, NICTA / UNSW
W39	Skill Discovery in Continuous Reinforcement Learning Domains using Skill Chaining, GEORGE KONIDARIS, University of Massachusetts, and ANDREW BARTO, University of Massachusetts Amherst
W40	Strategy Grafting in Extensive Games, KEVIN WAUGH, NOLAN BARD and MICHAEL BOWLING, University of Alberta
W41	Complexity of Decentralized Control: Special Cases, MARTIN ALLEN, Connecticut College, and SHLOMO ZILBERSTEIN, University of Massachusetts at Amherst
W42	Discrete MDL Predicts in Total Variation, MARCUS HUTTER, Australian National University
W43	Potential-Based Agnostic Boosting, ADAM KALAI, Microsoft, and VARUN KANADE, Harvard University
W44	On Stochastic and Worst-case Models for Investing, ELAD HAZAN, IBM, and SATYEN KALE, Yahoo! Research
W45	Statistical Consistency of Top-k Ranking, FEN XIA, Institute of Automation, Chinese Academy of Sciences, TIE-YAN LIU and HANG LI, Microsoft Research Asia
W46	Learning from Multiple Partially Observed Views - an Application to Multilingual Text Categorization, MASSIH AMINI, National Research Councl Canad, NICOLAS USUNIER, Université Pierre et Marie Cur, and CYRIL GOUTTE, National Research Council Canada

W47	Tracking Dynamic Sources of Malicious Activity at Internet Scale, SHOBHA VENKATARAMAN, AT&T Labs – Research, AVRIM BLUM, CMU, DAWN SONG, UC Berkeley, SUBHABRATA SEN and OLIVER SPATSCHECK, AT&T Labs – Research
W48	Kernel Choice and Classifiability for RKHS Embeddings of Probability Distributions, BHARATH SRIPERUMBUDUR, UC San Diego, KENJI FUKUMIZU, The Institute of Statistical Mathematics, ARTHUR GRETTON, Carnegie Mellon University and Max Planck Institute, GERT LANCKRIET, University of California, and BERNHARD SCHOELKOPF, MPI for Biological Cybernetics
W49	Information-theoretic lower bounds on the oracle complexity of convex optimization, Alekh Agarwal, Peter Bartlett, UC Berkeley, PRADEEP RAVIKUMAR, University of California, and MARTIN WAINWRIGHT, UC Berkeley
W50	Lower bounds on minimax rates for nonparametric regression with additive sparsity and smoothness, GARVESH RASKUTTI, UC Berkeley, MARTIN WAINWRIGHT and BIN YU, University of California
W51	A Parameter-free Hedging Algorithm, KAMALIKA CHAUDHURI, YOAV FREUND, UC San Diego, and DANIEL HSU, University of California
W52	Reading Tea Leaves: How Humans Interpret Topic Models, JONATHAN CHANG, JORDAN BOYD-GRABER, SEAN GERRISH, CHONG WANG and DAVID BLEI, Princeton University
W53	Sufficient Conditions for Agnostic Active Learnable, LIWEI WANG, Peking University
W54	Sparse Metric Learning via Smooth Optimization, YIMING YING, University of Bristol, KAIZHU HUANG, NLPR, Institute of Automation, Chinese Academy of Sciences, and COLIN CAMPBELL, University of Bristol
W55	Statistical Analysis of Semi-Supervised Learning: The Limit of Infinite Unlabelled Data, BOAZ NADLER, Weizmann Institute of Science, NATHAN SREBRO, Toyota Technological Institute at Chicago, and XUEYUAN ZHOU, University of Chicago
W56	Manifold Regularization for SIR with Rate Root-n Convergence, WEI BIAN, NTU, and DACHENG TAO, Nanyang Technological University 186
W57	Anomaly Detection with Score functions based on Nearest Neighbor Graphs, MANQI ZHAO and VENKATESH SALIGRAMA, Boston University
W58	Adaptive Regularization for Transductive Support Vector Machine, ZENGLIN XU, CUHK, RONG JIN, Michigan State University, JIANKE ZHU, ETH Zurich, IRWIN KING, CUHK, MICHAEL LYU, The Chinese University of HK, and ZHIRONG YANG, CUHK, TKK

W59	Canonical Time Warping for Alignment of Human Behavior, FENG ZHOU and FERNANDO DE LA TORRE, Carnegie Mellon University	187
W60	Fast Graph Laplacian Regularized Kernel Learning via Semidefinite-Quadratic-Linear Programming, XIAO-MING WU, ANTHONY MAN-CHO SO, ZHENGUO LI and SHUO-YEN ROBERT LI, The Chinese University of Hong Kong	187
W61	Replicated Softmax: an Undirected Topic Model, RUSLAN SALAKHUTDINOV, Department of Brain and Cognitive Sciences and CSAIL MIT, and GEOFFREY HINTON, University of Toronto	188
W62	A Game-Theoretic Approach to Hypergraph Clustering, SAMUEL ROTA BULÒ, Università Ca' Foscari di Venezia, and MARCELLO PELILLO, Università Ca Foscari di Venezia	188
W63	Learning to Hash with Binary Reconstructive Embeddings, BRIAN KULIS, UC Berkeley EECS & ICSI, and TREVOR DARRELL, UC Berkeley EECS and ICSI	189
W64	Posterior vs Parameter Sparsity in Latent Variable Models, JOAO GRACA, University of Lisbon, KUZMAN GANCHEV, University of Pennsylvania, BEN TASKAR, University of Pennsylvania, and FERNANDO PEREIRA, Google	189
W65	Occlusive Components Analysis, JÖRG LÜCKE, FIAS, Goethe-University Frankfurt, RICHARD TURNER, Gatsby Computational Neuroscience Unit, UCL, MANEESH SAHANI, University College London, and MARC HENNIGES, FIAS, Goethe-University Frankurt	190
W66	Bayesian Source Localization with the Multivariate Laplace Prior, MARCEL VAN GERVEN, BOTOND CSEKE, ICIS, ROBERT OOSTENVELD, DCCN, and TOM HESKES, Radboud University Nijmegen	191
W67	A Bayesian Model for Simultaneous Image Clustering, Annotation and Object Segmentation, LAN DU, Duke University, ECE Department, LU REN, Duke ECE, DAVID DUNSON, Duke, and LAWRENCE CARIN, Duke ECE	191
W68	Bayesian Nonparametric Models on Decomposable Graphs, FRANCOIS CARON, INRIA Bordeaux, and ARNAUD DOUCET, Institute of Statistical Mathematics	191
W69	<i>Efficient Moments-based Permutation Tests,</i> CHUNXIAO ZHOU, UIUC, HUIXIA JUDY WANG, North Carolina State University, and YONGMEI MICHELLE WANG, University of Illinois at Urbana-Champaign	192
W70	Localizing Bugs in Program Executions with Graphical Models, LAURA DIETZ, Max Planck Institute for Infor, VALENTIN DALLMEIER, ANDREAS ZELLER, Saarland University, and TOBIAS SCHEFFER, University of Potsdam	192

SCHEDULE, WEDNESDAY

W71	Learning transport operators for image manifolds, BENJAMIN CULPEPPER, UC Berkeley, and BRUNO OLSHAUSEN, University of California, Berkeley
W72	Variational Gaussian-process factor analysis for modeling spatio-temporal data, JAAKKO LUTTINEN and ALEXANDER ILIN, Helsinki University of Technology
W73	An LP View of the M-best MAP problem, MENACHEM FROMER and AMIR GLOBERSON, Hebrew University
W74	Nonparametric Bayesian Models for Unsupervised Event Coreference Resolution, COSMIN BEJAN, University of Texas at Dallas, MATTHEW TITSWORTH, ANDREW HICKL, Language Computer Corporation, and SANDA HARABAGIU, University of Texas at Dallas
W75	Sharing Features among Dynamical Systems with Beta Processes, EMILY FOX, Massachusetts Institute of Technology, ERIK SUDDERTH, Brown University, MICHAEL JORDAN, UC Berkeley, and ALAN WILLSKY, MIT 160
W76	Convex Relaxation of Mixture Regression with Efficient Algorithms, NOVI QUADRIANTO, SML-NICTA and RSISE-ANU, TIBERIO CAETANO, NICTA, JOHN LIM, NICTA and ANU, and DALE SCHUURMANS, University of Alberta
W77	FACTORIE: Probabilistic Programming via Imperatively Defined Factor Graphs, ANDREW MCCALLUM, KARL SCHULTZ and SAMEER SINGH, University of Massachusetts, Amherst
W78	Particle-based Variational Inference for Continuous Systems, ALEXANDER IHLER, ANDREW FRANK and PADHRAIC SMYTH, UC Irvine 195
W79	Approximating MAP by Compensating for Structural Relaxations, ARTHUR CHOI, UCLA, and ADNAN DARWICHE, University of California 195
W80	AcceleratingBayesianStructuralInferenceforNon-DecomposableGaussianGraphicalModels,BABACKMOGHADDAM, Caltech, BENJAMINMARLIN, EMTIYAZKHAN and KEVINMURPHY, University of British Columbia203
W81	An Infinite Factor Model Hierarchy Via a Noisy-Or Mechanism, AARON COURVILLE, DOUGLAS ECK and YOSHUA BENGIO, University of Montreal
W82	Multi-Label Prediction via Sparse Infinite CCA, PIYUSH RAI and HAL DAUME, University of Utah
W83	Submanifold density estimation, ARKADAS OZAKIN, Georgia Tech Research Institute, and ALEXANDER GRAY, Georgia Institute of Technology 197
W84	Heterogeneous multitask learning with joint sparsity constraints, XIAOLIN YANG, SEYOUNG KIM and ERIC XING, Carnegie Mellon University 197

W85	Modeling the spacing effect in sequential category learning, HONGJING LU, Dept. of Psychology & Statistics, UCLA, MATTHEW WEIDEN, Dept. of Psychology, UCLA, and ALAN YUILLE, UCLA		
W86	Hierarchical Learning of Dimensional Biases in Human Categorization, KATHERINE HELLER, University of Cambridge, ADAM SANBORN and NICK CHATER, University College London		
W87	Help or Hinder: Bayesian Models of Social Goal Inference, TOMER ULLMAN, CHRIS BAKER, OWEN MACINDOE, OWAIN EVANS, NOAH GOODMAN and JOSHUA TENENBAUM, Massachusetts Institute of Technology		
7:30–11:59pm	Demonstrations		
D8	An FPGA-based Multi-Algorithm Learning and Classification Engine, HARI CADAMBI, NEC Laboratories America		
D9	Curious George: An Attentive Semantic Robotic Vision System, ANKUR GUPTA, University of British Columbia		
D10	Deformed Lattice Detection in Real-World Images Using a Novel Mean-Shift Belief, MINWOO PARK, Penn State Univ		
D11	Monte: An Interactive Ssytem for Massively Parallel Probabilistic Programming, VIKASH MANSINGHKA, Massachusetts Institute of Technology. 200		
D12	Nengo: Large-Scale Integrated Neural Modeling, CHRIS ELIASMITH, U of Waterloo		
D13	Real-Time Adaptive Face Recognition, BRANISLAV KVETON, Intel Corporation		
D14	Robust Speaker Recognition Using Approximate Bayesian Inference, CIIRA MAINA, Drexel University		

Thursday, December 10th

8:30–9:50am	Oral Session: Bayesian Analysis		
8:30am	Invited Talk: Bayesian Analysis of Markov Chains, PERSI DIACONIS, Stanford		
9:30am	AcceleratingBayesianStructuralInferenceforNon-DecomposableGaussianGraphicalModels,BABACKMOGHADDAM, Caltech, BENJAMINMARLIN, EMTIYAZKHAN and KEVINMURPHY, University of British Columbia203		
10:30–11:50am	Oral Session: Neural Modeling and Imaging		
10:30am	Invited Talk: Optogenetics: Development and Application, KARL DEISSEROTH, Stanford University		
11:30am	Time-rescaling methods for the estimation and assessment of non- Poisson neural encoding models, JONATHAN PILLOW, University of Texas at Austin		
1:30–4:30pm	Mini Symposia Parallel sessions with one break		
MS1	Assistive Machine Learning for People with Disabilities, FERNANDO PEREZ-CRUZ, University Carlos III in Madrid, EMILIO PARRADO-HERNANDEZ, University Carlos III de Madrid, DAVID HARDOON, Institute for Infocomm Research, and JAISIEL MADRID-SANCHEZ, Fundosa Technosite Location: Regency B		
MS2	Causality and Time Series Analysis, FLORIN POPESCU, Fraunhofer FIRST, ISABELLE GUYON, Clopinet, and GUIDO NOLTE, Location: Regency A		
MS3	Machine Learning for Sustainability, J. ZICO KOLTER, Stanford University, THOMAS DIETTERICH, Oregon State University, and ANDREW NG, Stanford University Location: Regency D		
MS4	Machine Learning in Computational Biology, YANJUN QI, NEC Labs America, GAL CHECHIK, Google, ALEXANDER ZIEN, LIFE Biosystems GmbH, TOMER HERTZ, Fred Hutchnison Cancer Research Center, and WILLIAM NOBLE, University of Washington Location: Regency C		
MS5	Partially Observable Reinforcement Learning, MARCUS HUTTER, Australian National University, WILL UTHER, NICTA Neville Road Research Lab, and PASCAL POUPART, University of Waterloo Location: Regency E/F		

Abstracts of Tutorials

Tutorial Session 1, 9:30am–11:30am

Tutorial: Making Very Large-Scale Linear Algebraic Computations Possible Via Randomization

GUNNAR MARTINSSON Per-gunnar.Martinsson@Colorado.EDU University of Colorado at Boulder

The demands on software for analyzing data are rapidly increasing as ever larger data sets are generated in medical imaging, in analyzing large networks such as the World Wide Web, in image and video processing, and in a range of other applications. To handle this avalanche of data, any software used must be able to fully exploit modern hardware characterized by multiple processors and capacious but slow memory. The evolution of computer architecture is currently forcing a shift in algorithm design away from the classical algorithms that were designed for single-processor computers with all the data available in Random Access Memory (RAM), towards algorithms that aggressively minimize communication costs. This tutorial will describe a set of recently developed techniques for standard linear algebraic computations (such as computing a partial singular value decomposition of a matrix) that are very well suited for implementation on multi-core or other parallel architectures, and for processing data stored on disk, or streamed. These techniques are based on the use of randomized sampling to reduce the effective dimensionality of the data. Remarkably, randomized sampling does not only loosen communication constraints, but does so while maintaining, or even improving, the accuracy and robustness of existing deterministic techniques.

Gunnar Martinsson is an assistant professor of applied mathematics at the University of Colorado at Boulder. He graduated from Chalmers University of Technology in Gothenburg, Sweden, in 1998, and earned his PhD in 2002 from the Computational and Applied Mathematics program at the University of Texas at Austin. His research interests include numerical linear algebra, fast solvers for linear PDEs, and applied harmonic analysis. A particular focus of his research is the development of algorithms specifically designed for very large scale problems. Such problems require algorithms whose complexity scales linearly with problem size, that can efficiently exploit parallel processors in a variety of configurations, that can run with the data stored out-of-core or streamed, and that maintain optimal accuracy even for very large problems.

Tutorial Session 1, 9:30am–11:30am

Tutorial: Understanding Visual Scenes

ANTONIO TORRALBA torralba@csail.mit.edu Massachusetts Institute of Technology

One remarkable aspect of human vision is the ability to understand the meaning of a novel image or event quickly and effortlessly. Within a single glance, we can comprehend the semantic category of a place, its spatial layout as well as identify many of the objects that compose the scene. Approaching human abilities at scene understanding is a current challenge for computer vision systems. The field of scene understanding is multidisciplinary, involving a growing collection of works in computer vision, machine learning, cognitive psychology, and neuroscience. In this tutorial, we will focus on recent work on computer vision attempting to solve the tasks of scene recognition and classification, visual context representation, object recognition in context, drawing parallelism with work in psychology and neuroscience. Devising systems that solve scene and object recognition in an integrated fashion will lead towards more efficient and robust artificial visual understanding systems.

Antonio Torralba is associate Professor of Electrical Engineering and Computer Science at the Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT. Following his degree in telecommunications engineering, obtained at the Universidad Politecnica de Catalunya, Spain, he was awarded a Ph.D. in Signal, Image, and Speech processing from the Institut National Polytechnique de Grenoble, France. Thereafter, he spent post-doctoral training at the Brain and Cognitive Science Department and the Computer Science and Artificial Intelligence Laboratory at MIT.

Tutorial Session 2, 1:00pm-3:00pm

Tutorial: Model-Based Reinforcement Learning

MICHAEL LITTMAN Rutgers University mlittman@cs.rutgers.edu

In model-based reinforcement learning, an agent uses its experience to construct a representation of the control dynamics of its environment. It can then predict the outcome of its actions and make decisions that maximize its learning and task performance. This tutorial will survey work in this area with an emphasis on recent results. Topics will include: Efficient learning in the PAC-MDP formalism, Bayesian reinforcement learning, models and linear function approximation, recent advances in planning.

Michael L. Littman is professor and chair of the Department of Computer Science at Rutgers University and directs the Rutgers Laboratory for Real-Life Reinforcement Learning (RL3). His research in machine learning examines algorithms for decision making under uncertainty. Littman has earned multiple awards for teaching and his research has been recognized with three best-paper awards on the topics of meta-learning for computer crossword solving, complexity analysis of planning under uncertainty, and algorithms for efficient reinforcement learning. He has served on the editorial boards for several machine-learning journals and was Programme Co-chair of ICML 2009.

Tutorial Session 2, 1:00pm-3:00pm

Tutorial: Sparse Methods for Machine Learning: Theory and Algorithms

FRANCIS BACH francis.bach@mines.org INRIA - Ecole Normale Superieure

Regularization by the L1-norm has attracted a lot of interest in recent years in statistics, machine learning and signal processing. In the context of least-square linear regression, the problem is usually referred to as the Lasso or basis pursuit. Much of the early effort has been dedicated to algorithms to solve the optimization problem efficiently, either through first-order methods, or through homotopy methods that leads to the entire regularization path (i.e., the set of solutions for all values of the regularization parameters) at the cost of a single matrix inversion. A well-known property of the regularization by the L1-norm is the sparsity of the solutions, i.e., it leads to loading vectors with many zeros, and thus performs model selection on top of regularization. Recent works have looked precisely at the model consistency of the Lasso, i.e., if we know that the data were generated from a sparse loading vector, does the Lasso actually recover the sparsity pattern when the number of observations grows? Moreover, how many irrelevant variables could we consider while still being able to infer correctly the relevant ones? The objective of the tutorial is to give a unified overview of the recent contributions of sparse convex methods to machine learning, both in terms of theory and algorithms. The course will be divided in three parts: in the first part, the focus will be on the regular L1-norm and variable selection, introducing key algorithms and key theoretical results. Then, several more structured machine learning problems will be discussed, on vectors (second part) and matrices (third part), such as multi-task learning, sparse principal component analysis, multiple kernel learning and sparse coding. Web page: http://www.di.ens.fr/fbach/nips2009tutorial/

Francis Bach is a researcher in the Willow INRIA project-team, in the Computer Science Department of Ecole Normale Supérieure, Paris, France. He graduated from Ecole Polytechnique, Palaiseau, France, in 1997, and earned his PhD in 2005 from the Computer Science division at the University of California, Berkeley. His research interests include machine learning, statistics, convex and combinatorial optimization, graphical models, kernel methods, sparse methods, signal processing and computer vision.

Tutorial Session 3, 3:30pm-5:30pm

Tutorial: Deep Learning in Natural Language Processing

RONAN COLLOBERT NEC Laboratories of America JASON WESTON Google ronan@collobert.com

This tutorial will describe recent advances in deep learning techniques for Natural Language Processing (NLP). Traditional NLP approaches favour shallow systems, possibly cascaded, with adequate hand-crafted features. In constrast, we are interested in end-to-end architectures: these systems include several feature layers, with increasing abstraction at each layer. Compared to shallow systems, these feature layers are learnt for the task of interest, and do not require any engineering. We will show how neural networks are naturally well suited for end-to-end learning in NLP tasks. We will study multi-tasking different tasks, new semi-supervised learning techniques adapted to these deep architectures, and review end-to-end structured output learning. Finally, we will highlight how some of these advances can be applied to other fields of research, like computer vision, as well.

Ronan Collobert received his master degree in pure mathematics from University of Rennes (France) in 2000. He then performed graduate studies in University of Montreal and IDIAP (Switzerland) under the Bengio brothers, and received his PhD in 2004 from University of Paris VI. He joined NEC Labs (USA) in January 2005 as a postdoc, and became a research staff member after about one year. His research interests always focused on large-scale machine-learning algorithms, with a particular interest in semi-supervised learning and deep learning architectures. Two years ago, his research shifted in the natural language processing area, slowly going towards automatic text understanding.

Jason Weston received a PhD. (2000) from Royal Holloway, University of London under the supervision of Vladimir Vapnik. From 2000 to 2002, he was a researcher at Biowulf technologies, New York, applying machine learning to bioinformatics. From 2002 to 2003 he was a research scientist at the Max Planck Institute for Biological Cybernetics, Tuebingen, Germany. From 2004 to June 2009 he was a research staff member at NEC Labs America, Princeton. From July 2009 onwards he has been a research scientist at Google, New York. Jason Weston's current research focuses on various aspects of statistical machine learning and its applications, particularly in text and images.

Tutorial Session 3, 3:30pm–5:30pm

Tutorial: Sequential Monte-Carlo Methods

ARNAUD DOUCET University of British Columbia NANDO DE FREITAS UBC

nando@cs.ubc.ca

arnaud@cs.ubc.ca

Over the last fifteen years, sequential Monte Carlo (SMC) methods gained popularity as powerful tools for solving intractable inference problems arising in the modelling of sequential data. Much effort was devoted to the development of SMC methods, known as particle filters (PFs), for estimating the filtering distribution of the latent variables in This line of research produced many algorithms, including dynamic models. auxiliary-variable PFs. marginal PFs. the resample-move algorithm and Rao-Blackwellised PFs. It also led to many applications in tracking, computer vision, robotics and econometrics. The theoretical properties of these methods were also studied extensively in this period. Although PFs occupied the center-stage, significant progress was also attained in the development of SMC methods for parameter estimation, online EM, particle smoothing and SMC techniques for control and planning. Various SMC algorithms were also designed to approximate sequences of unnormalized functions, thus allowing for the computation of eigen-pairs of large matrices and kernel operators. Recently, frameworks for building efficient high-dimensional proposal distributions for MCMC using SMC methods were proposed. These allow us to design effective MCMC algorithms in complex scenarios where standard strategies failed. Such methods have been demonstrated on a number of domains, including simulated tempering, Dirichlet process mixtures, nonlinear non-Gaussian state-space models, protein folding and stochastic differential equations. Finally, SMC methods were also generalized to carry out approximate inference in static models. This is typically done by constructing a sequence of probability distributions, which starts with an easy-to-sample-from distribution and which converges to the desired target distribution. These SMC methods have been successfully applied to notoriously hard problems, such as inference in Boltzmann machines, marginal parameter estimation and nonlinear Bayesian experimental design. In this tutorial, we will introduce the classical SMC methods and expose the audience to the new developments in the field.

Arnaud Doucet obtained a PhD in Information Engineering from University Paris XI in 1997. He has held previously academic appointments at Melbourne University and Cambridge University. He is currently Professor at the Institute of Statistical Mathematics in Tokyo and Associate Professor and Canada Research Chair in the Departments of Computer Science and Statistics of the University of British Columbia. He is especially interested in Bayesian statistics, Monte Carlo methods and their applications.

Nando de Freitas is an associate professor in Computer Science and Cognitive Systems at the University of British Columbia (UBC). He is also an associate member of the Statistics Department. He obtained his PhD on "Bayesian methods for neural networks" at Cambridge University in 2000. During the next two years he was a postdoctoral scholar at UC Berkeley with Prof Stuart Russell. Nando has contributed to many areas of machine learning including fast n-body methods, hierarchical Bayesian models for model selection, Markov decision processes, Bayesian optimization, robotics, image tracking, Bayesian experimental design, multimedia retrieval and machine translation approaches for object recognition. Monte Carlo methods have been central to Nando's research. In 2000, he edited the first book on sequential Monte Carlo methods with Arnaud Doucet and Neil Gordon. This booked played an important role in unifying the field and it is now the 5th most cited computer science publication in 2000 according to the Citeseer index. With his students and collaborators, Nando has contributed to the development of many well known Monte Carlo methods, including Rao-Blackwellized particle filters for dynamic Bayesian networks, unscented and boosted particle filters for tracking, efficient marginal particle filters and particle smoothers, sequential Monte Carlo methods for Boltzmann machines (hot coupling), tree sampling methods for random fields and reversible jump MCMC algorithms for control and planning.

Abstracts of Presentations

Monday, December 7th

Spotlights (6:30–6:50pm)

- Learning to Rank by Optimizing NDCG Measure HAMED VALIZADEGAN, RONG JIN, Michigan State University, RUOFEI ZHANG and JIANCHANG MAO, Yahoo!. Subject Area: Supervised Learning See abstract, page 88.
- On Learning Rotations RAMAN ARORA, Univ. of Wisconsin-Madison. Subject Area: Supervised Learning See abstract, page 88.
- Sparsistent Learning of Varying-coefficient Models with Structural Changes MLADEN KOLAR, LE SONG and ERIC XING, Carnegie Mellon University. Subject Area: Supervised Learning See abstract, page 90.
- Replacing supervised classification learning by Slow Feature Analysis in spiking neural networks
 STEFAN KLAMPFL and WOLFGANG MAASS, Graz University of Technology.
 Subject Area: Neuroscience See abstract, page 66.
- Heavy-Tailed Symmetric Stochastic Neighbor Embedding ZHIRONG YANG, CUHK, TKK, IRWIN KING, ZENGLIN XU, CUHK, and ERKKI OJA, TKK. Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 73.
- Spatial Normalized Gamma Processes VINAYAK RAO and YEE WHYE TEH, Gatsby Computational Neuroscience Unit, UCL. Subject Area: Probabilistic Models and Methods See abstract, page 60.
- Rethinking LDA: Why Priors Matter HANNA WALLACH, DAVID MIMNO and ANDREW MCCALLUM, University of Massachusetts Amherst.
 Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 75.

MONDAY, SPOTLIGHTS

Nonparametric Latent Feature Models for Link Prediction
 KURT MILLER, UC Berkeley, THOMAS GRIFFITHS, University of California, Berkeley, and
 MICHAEL JORDAN, University of California.

 Subject Area: Probabilistic Models and Methods
 See abstract, page 63.

• A Fast, Consistent Kernel Two-Sample Test ARTHUR GRETTON, Carnegie Mellon University and Max Planck Institute, KENJI FUKUMIZU, The Institute of Statistical Mathematics, ZAID HARCHAOUI, Carnegie Mellon University, and BHARATH SRIPERUMBUDUR, UC San Diego. Subject Area: Supervised Learning See abstract, page 94.

- Entropic Graph Regularization in Non-Parametric Semi-Supervised Classification AMARNAG SUBRAMANYA and JEFF BILMES, University of Washington. Subject Area: Speech and Signal Processing See abstract, page 57.
- Asymptotic Analysis of MAP Estimation via the Replica Method and Compressed Sensing
 SUNDEEP RANGAN, Qualcomm, Inc., ALYSON FLETCHER, University of California, Berkeley, and VIVEK GOYAL, Massachusetts Institute of Technology.
 Subject Area: Probabilistic Models and Methods See abstract, page 63.

Fast Learning from Non-i.i.d. Observations
 INGO STEINWART, Los Alamos National Laboratory, and ANDREAS CHRISTMANN,
 University of Bayreuth.
 Subject Area: Theory
 See abstract, page 84.

 Convergent Temporal-Difference Learning with Arbitrary Smooth Function Approximation
 HAMID MAEI, CSABA SZEPESVARI, University of Alberta, SHALABH BHATNAGAR, Indian Institute of Science, Bangalore, India, DOINA PRECUP, McGill University, DAVID SILVER and RICH SUTTON, University of Alberta.
 Subject Area: Control and Reinforcement Learning See abstract, page 96.

Poster Session (7:00–11:59pm)

M1 Entropic Graph Regularization in Non-Parametric Semi-Supervised Classification

Amarnag Subramanya	asubram@ee.washington.edu
Jeff Bilmes	bilmes@ee.washington.edu
University of Washington	

We prove certain theoretical properties of a graph-regularized transductive learning objective that is based on minimizing a Kullback-Leibler divergence based loss. These include showing that the iterative alternating minimization procedure used to minimize the objective converges to the correct solution and deriving a test for convergence. We also propose a graph node ordering algorithm that is cache cognizant and leads to a linear speedup in parallel computations. This ensures that the algorithm scales to large data sets. By making use of empirical evaluation on the TIMIT and Switchboard I corpora, we show this approach is able to out-perform other state-of-the-art SSL approaches. In one instance, we solve a problem on a 120 million node graph. **Subject Area:** Speech and Signal Processing

Spotlight presentation, Monday, 6:30.

M2 A Sparse Non-Parametric Approach for Single Channel Separation of Known Sounds

Paris Smaragdis	paris@media.mit.edu
Adobe Systems Inc.	
Madhusudana Shashanka	shashanka@alum.bu.edu
Mars Inc.	
Bhiksha Raj	bhiksha@cs.cmu.edu
Carnegie Mellon University	

In this paper we present an algorithm for separating mixed sounds from a monophonic recording. Our approach makes use of training data which allows us to learn representations of the types of sounds that compose the mixture. In contrast to popular methods that attempt to extract com- pact generalizable models for each sound from training data, we employ the training data itself as a representation of the sources in the mixture. We show that mixtures of known sounds can be described as sparse combinations of the training data itself, and in doing so produce significantly better separation results as compared to similar systems based on compact statistical models. **Subject Area:** Speech and Signal Processing

M3 Sparse and Locally Constant Gaussian Graphical Models

Jean Honorio	jhonorio@cs.sunysb.edu
Luis Ortiz	leortiz@cs.sunysb.edu
Dimitris Samaras	samaras@cs.sunysb.edu
Stony Brook University	
Nikos Paragios	nikos.paragios@ecp.fr
Ecole Centrale Paris	
RITA GOLDSTEIN	rgoldstein@bnl.gov
Brookhaven National Laboratory	

Locality information is crucial in datasets where each variable corresponds to a measurement in a manifold (silhouettes, motion trajectories, 2D and 3D images). Although these datasets are typically under-sampled and high-dimensional, they often need to be represented with low-complexity statistical models, which are comprised of only the important probabilistic dependencies in the datasets. Most methods attempt to reduce model complexity by enforcing structure sparseness. However, sparseness cannot describe inherent regularities in the structure. Hence, in this paper we first propose a new class of Gaussian graphical models which, together with sparseness, imposes local constancy through ℓ_1 -norm penalization. Second, we propose an efficient algorithm which decomposes the strictly convex maximum likelihood estimation into a sequence of problems with closed form solutions. Through synthetic experiments, we evaluate the closeness of the recovered models to the ground truth. We also test the generalization performance of our method in a wide range of complex real-world datasets and demonstrate that it can capture useful structures such as the rotation and shrinking of a beating heart, motion correlations between body parts during walking and functional interactions of brain regions. Our method outperforms the state-of-the-art structure learning techniques for Gaussian graphical models both for small and large datasets. Subject Area: Probabilistic Models and Methods

M4 A Stochastic approximation method for inference in probabilistic graphical models

Peter Carbonetto	pcarbo@cs.ubc.ca
University of Chicago	
MATTHEW KING	kingdom@interchange.ubc.ca
University of British Columbia	
Firas Hamze	orbis.tertius@gmail.com
D-Wave Systems	

We describe a new algorithmic framework for inference in probabilistic models, and apply it to inference for latent Dirichlet allocation. Our framework adopts the methodology of variational inference, but unlike existing variational methods such as mean field and expectation propagation it is not restricted to tractable classes of approximating distributions. Our approach can also be viewed as a sequential Monte Carlo (SMC) method, but unlike existing SMC methods there is no need to design the artificial sequence of distributions. Notably, our framework offers a principled means to exchange the variance of an importance sampling estimate for the bias incurred through variational approximation. Experiments on a challenging inference problem in population genetics demonstrate improvements in stability and accuracy over existing methods, and at a comparable cost.

Subject Area: Probabilistic Models and Methods

M5 Variational Inference for the Nested Chinese Restaurant Process

CHONG WANG DAVID BLEI Princeton University chongw@cs.princeton.edu blei@cs.princeton.edu

The nested Chinese restaurant process (nCRP) is a powerful nonparametric Bayesian model for learning tree-based hierarchies from data. Since its posterior distribution is intractable, current inference methods have all relied on MCMC sampling. In this paper, we develop an alternative inference technique based on variational methods. To employ variational methods, we derive a tree-based stick-breaking construction of the nCRP mixture model, and a novel variational algorithm that efficiently explores a posterior over a large set of combinatorial structures. We demonstrate the use of this approach for text and hand written digits modeling, where we show we can adapt the nCRP to continuous data as well.

Subject Area: Probabilistic Models and Methods

M6 Gaussian process regression with Student-t likelihood

Jarno Vanhatalo	jarno.vanhatalo@tkk.fi
Pasi Jylänki	pasi.jylanki@tkk.fi
Aki Vehtari	aki.vehtari@hut.fi
Helsinki University of Technology	

In the Gaussian process regression the observation model is commonly assumed to be Gaussian, which is convenient in computational perspective. However, the drawback is that the predictive accuracy of the model can be significantly compromised if the observations are contaminated by outliers. A robust observation model, such as the Student-t distribution, reduces the influence of outlying observations and improves the predictions. The problem, however, is the analytically intractable inference. In this work, we discuss the properties of a Gaussian process regression model with the Student-t likelihood and utilize the Laplace approximation for approximate inference. We compare our approach to a variational approximation and a Markov chain Monte Carlo scheme, which utilize the commonly used scale mixture representation of the Student-t distribution.

Subject Area: Probabilistic Models and Methods

M7 A Bayesian Analysis of Dynamics in Free Recall

RICHARD SOCHERrichard@socher.orgStanford Universitysjgershm@princeton.eduSAMUEL GERSHMANsjgershm@princeton.eduADLER PEROTTEaperotte@Princeton.EDUPER SEDERBERGpersed@princeton.eduDAVID BLEIblei@cs.princeton.eduKENNETH NORMANknorman@Princeton.EDUPrinceton UniversityStateman

We develop a probabilistic model of human memory performance in free recall experiments. In these experiments, a subject first studies a list of words and then tries to recall them. To model these data, we draw on both previous psychological research and statistical topic models of text documents. We assume that memories are formed by assimilating the semantic meaning of studied words (represented as a distribution over topics) into a slowly changing latent context (represented in the same space). During recall, this context is reinstated and used as a cue for retrieving studied words. By conceptualizing memory retrieval as a dynamic latent variable model, we are able to use Bayesian inference to represent uncertainty and reason about the cognitive processes underlying memory. We present a particle filter algorithm for performing approximate posterior inference, and evaluate our model on the prediction of recalled words in experimental data. By specifying the model hierarchically, we are also able to capture inter-subject variability. **Subject Area:** Probabilistic Models and Methods

M8 Streaming Pointwise Mutual Information

Benjamin Van Durme	vandurme@cs.rochester.edu
University of Rochester	
Ashwin Lall	alall@cs.rochester.edu
Georgia Institute of Technology	

Recent work has led to the ability to perform space efficient, approximate counting over large vocabularies in a streaming context. Motivated by the existence of data structures of this type, we explore the computation of associativity scores, other- wise known as pointwise mutual information (PMI), in a streaming context. We give theoretical bounds showing the impracticality of perfect online PMI compu- tation, and detail an algorithm with high expected accuracy. Experiments on news articles show our approach gives high accuracy on real world data.

Subject Area: Probabilistic Models and Methods

M9 Non-stationary continuous dynamic Bayesian networks

MARCO GRZEGORCZYK grzegorczyk@statistik.tu-dortmund.de Department of Statistics, TU Dortmund University, Germany DIRK HUSMEIER dirk@bioss.ac.uk Biomathematics and Statistics Scotland (BioSS), Edinburgh, UK

Dynamic Bayesian networks have been applied widely to reconstruct the structure of regulatory processes from time series data. The standard approach is based on the assumption of a homogeneous Markov chain, which is not valid in many real-world scenarios. Recent research efforts addressing this shortcoming have considered undirected graphs, directed graphs for discretized data, or over-flexible models that lack any information sharing between time series segments. In the present article, we propose a non-stationary dynamic Bayesian network for continuous data, in which parameters are allowed to vary between segments, and in which a common network structure provides essential information sharing across segments. Our model is based on a Bayesian change-point process, and we apply a variant of the allocation sampler of Nobile and Fearnside to infer the number and location of the change-points. **Subject Area:** Probabilistic Models and Methods

M10 Spatial Normalized Gamma Processes

Vinayak Rao	vrao@gatsby.ucl.ac.uk
Yee Whye Teh	ywteh@gatsby.ucl.ac.uk
Gatsby Computational	Neuroscience Unit, UCL

Dependent Dirichlet processes (DPs) are dependent sets of random measures, each being marginally Dirichlet process distributed. They are used in Bayesian nonparametric models when the usual exchangebility assumption does not hold. We propose a simple and general framework to construct dependent DPs by marginalizing and normalizing a single gamma process over an extended space. The result is a set of DPs, each located at a point in a space such that neighboring DPs are more dependent. We describe Markov chain Monte Carlo inference, involving the typical Gibbs sampling and three different Metropolis-Hastings proposals to speed up convergence. We report an empirical study of convergence speeds on a synthetic dataset and demonstrate an application of the model to topic modeling through time.

Subject Area: Probabilistic Models and Methods

Spotlight presentation, Monday, 6:30.

M11 Learning a Small Mixture of Trees

M. PAWAN KUMAR DAPHNE KOLLER Stanford University

pawan@cs.stanford.edu
koller@cs.stanford.edu

The problem of approximating a given probability distribution using a simpler distribution plays an important role in several areas of machine learning, e.g. variational inference and classification. Within this context, we consider the task of learning a mixture of tree distributions. Although mixtures of trees can be learned by minimizing the KL-divergence using an EM algorithm, its success depends heavily on the initialization. We propose an efficient strategy for obtaining a good initial set of trees that attempts to cover the entire observed distribution by minimizing the α -divergence with $\alpha = \infty$. We formulate the problem using the fractional covering framework and present a convergent sequential algorithm that only relies on solving a convex program at each iteration. Compared to previous methods, our approach results in a significantly smaller mixture of trees that provides similar or better accuracies. We demonstrate the usefulness of our approach by learning pictorial structures for face recognition.

Subject Area: Probabilistic Models and Methods

M12 Which graphical models are difficult to learn?

ANDREA MONTANARI JOSE AYRES PEREIRA Stanford University montanari@stanford.edu jbento@stanford.edu We consider the problem of learning the structure of Ising models (pairwise binary Markov random fields) from i.i.d. samples. While several methods have been proposed to accomplish this task, their relative merits and limitations remain somewhat obscure. By analyzing a number of concrete examples, we show that low-complexity algorithms systematically fail when the Markov random field develops long-range correlations. More precisely, this phenomenon appears to be related to the Ising model phase transition (although it does not coincide with it).

Subject Area: Probabilistic Models and Methods

M13 Large Scale Nonparametric Bayesian Inference: Data Parallelisation in the Indian Buffet Process

Shakir Mohamed	sm694@eng.cam.ac.uk
David Knowles	dak33@cam.ac.uk
Zoubin Ghahramani	zoubin@eng.cam.ac.uk
University of Cambridge	
Finale Doshi-Velez	finale@mit.edu
Cambridge University, MIT	

Nonparametric Bayesian models provide a framework for flexible probabilistic modelling of complex datasets. Unfortunately, Bayesian inference methods often require high-dimensional averages and can be slow to compute, especially with the potentially unbounded representations associated with nonparametric models. We address the challenge of scaling nonparametric Bayesian inference to the increasingly large datasets found in real-world applications, focusing on the case of parallelising inference in the Indian Buffet Process (IBP). Our approach divides a large data set between multiple processors. The processors use message passing to compute likelihoods in an asynchronous, distributed fashion and to propagate statistics about the global Bayesian posterior. This novel MCMC sampler is the first parallel inference scheme for IBP-based models, scaling to datasets orders of magnitude larger than had previously been possible. **Subject Area:** Probabilistic Models and Methods

M14 Modelling Relational Data using Bayesian Clustered Tensor Factorization

Ilya Sutskever	ilya@cs.utoronto.ca
University of Toronto	
Ruslan Salakhutdinov	rsalakhu@cs.toronto.edu
Department of Brain and Cogr	itive Sciences and CSAIL MIT
Joshua Tenenbaum	jbt@mit.edu
Massachusetts Institute of Tech	hnology

We consider the problem of learning probabilistic models for complex relational structures between various types of objects. A model can help us "understand" a dataset of relational facts in at least two ways, by finding interpretable structure in the data, and by supporting predictions, or inferences about whether particular unobserved relations are likely to be true. Often there is a tradeoff between these two aims: cluster-based models yield more easily interpretable representations, while factorization-based approaches have better predictive performance on large data sets. We introduce the Bayesian Clustered Tensor Factorization (BCTF) model, which embeds a factorized representation of relations in a nonparametric Bayesian clustering framework. Inference is fully Bayesian but scales well to large data sets. The model simultaneously discovers interpretable clusters and yields predictive performance that matches or beats previous probabilistic models for relational data.

Subject Area: Probabilistic Models and Methods

M15 Nonparametric Latent Feature Models for Link Prediction

Kurt Miller	tadayuki@cs.berkeley.edu
UC Berkeley	
THOMAS GRIFFITHS	tom_griffiths@berkeley.edu
University of California, Berkeley	
Michael Jordan	jordan@cs.berkeley.edu
University of California	

As the availability and importance of relational data – such as the friendships summarized on a social networking website – increases, it becomes increasingly important to have good models for such data. The kinds of latent structure that have been considered for use in predicting links in such networks have been relatively limited. In particular, the machine learning community has focused on latent class models, adapting nonparametric Bayesian methods to jointly infer how many latent classes there are while learning which entities belong to each class. We pursue a similar approach with a richer kind of latent variable – latent features – using a nonparametric Bayesian technique to simultaneously infer the number of features at the same time we learn which entities have each feature. The greater expressiveness of this approach allows us to improve link prediction on three datasets. **Subject Area:** Probabilistic Models and Methods

Spotlight presentation, Monday, 6:30.

M16 Asymptotic Analysis of MAP Estimation via the Replica Method and Compressed Sensing

 SUNDEEP RANGAN
 sdrangan@yahoo.com

 Qualcomm, Inc.
 alyson@eecs.berkeley.edu

 ALYSON FLETCHER
 alyson@eecs.berkeley.edu

 University of California, Berkeley
 vgoyal@mit.edu

 VIVEK GOYAL
 vgoyal@mit.edu

 Massachusetts Institute of Technology
 vgoyal@mit.edu

The replica method is a non-rigorous but widely-used technique from statistical physics used in the asymptotic analysis of many large random nonlinear problems. This paper applies the replica method to non-Gaussian MAP estimation. It is shown that with large random linear measurements and Gaussian noise, the asymptotic behavior of the MAP estimate of an n-dimensional vector "decouples" as n scalar MAP estimators. The result is a counterpart to Guo and Verdu's replica analysis on MMSE estimation. The replica MAP analysis can be readily applied to many estimators used in compressed sensing, including basis pursuit, lasso, linear estimation with thresholding and zero-norm estimation. In the case of lasso estimation, the scalar estimator reduces to a soft-thresholding operator and for zero-norm estimation it reduces to a hard-threshold. Among other benefits, the replica method provides a computationally tractable method for exactly computing various performance metrics including MSE and sparsity recovery.

Subject Area: Probabilistic Models and Methods

Spotlight presentation, Monday, 6:30.

M17 Sequential effects reflect parallel learning of multiple environmental regularities

MATTHEW WILDER mattwilder.cu@gmail.com University of Colorado Boulder MATT JONES mcj@colorado.edu MICHAEL MOZER mozer@colorado.edu University of Colorado at Boulder

Across a wide range of cognitive tasks, recent experience influences behavior. For example, when individuals repeatedly perform a simple two-alternative forced-choice task (2AFC), response latencies vary dramatically based on the immediately preceding trial sequence. These sequential effects have been interpreted as adaptation to the statistical structure of an uncertain, changing environment (e.g. Jones & Sieck, 2003; Mozer, Kinoshita, & Shettel, 2007; Yu & Cohen, 2008). The Dynamic Belief Model (DBM) (Yu & Cohen, 2008) explains sequential effects in 2AFC tasks as a rational consequence of a dynamic internal representation that tracks second-order statistics of the trial sequence (repetition rates) and predicts whether the upcoming trial will be a repetition or an alternation of the previous trial. Experimental results suggest that first-order statistics (base rates) also influence sequential effects. We propose a model that learns both firstand second-order sequence properties, each according to the basic principles of the DBM but under a unified inferential framework. This model, the Dynamic Belief Mixture Model (DBM2), obtains precise, parsimonious fits to data. Furthermore, the model predicts dissociations in behavioral (Maloney, Dal Martello, Sahm, & Spillmann, 2005) and electrophysiological studies (Jentzsch & Sommer, 2002), supporting the psychological and neurobiological reality of its two components. Subject Area: Probabilistic Models and Methods

M18 Noise Characterization, Modeling, and Reduction for In Vivo Neural Recording

Zhi Yang	yangzhi@soe.ucsc.edu
Qi Zhao	zhaoqi_zju@hotmail.com
UC Santa Cruz	
Edward Keefer	Edward.Keefer@utsouthwestern.edu
UT Southwestern Medical Center	
Wentai Liu	wentai@soe.ucsc.edu
UC Santa Cruz	

Studying signal and noise properties of recorded neural data is critical in developing more efficient algorithms to recover the encoded information. Important issues exist in this research including the variant spectrum spans of neural spikes that make it difficult to choose a global optimal bandpass filter. Also, multiple sources produce aggregated noise that deviates from the conventional white Gaussian noise. In this work, the spectrum variability of spikes is addressed, based on which the concept of adaptive bandpass filter that fits the spectrum of individual spikes is proposed. Multiple noise sources have been studied through analytical models as well as empirical measurements. The dominant noise source is identified as neuron noise followed by interface noise of the electrode. This suggests that major efforts to reduce noise from electronics are not well spent. The measured noise from in vivo experiments shows a family of $1/f^x(x = 1.5 \pm 0.5)$ spectrum that can be reduced using noise shaping techniques. In summary, the methods of adaptive bandpass filtering and noise shaping together result in several dB signal-to-noise ratio (SNR) enhancement.

Subject Area: Neuroscience

M19 Optimal context separation of spiking haptic signals by second-order somatosensory neurons

ROMAIN BRASSELETromain.brasselet@upmc.frCNRS - University Pierre & Marie Curie Paris 6ROLAND JOHANSSONRoland.S.Johansson@physiol.umu.seUMEA UniversityANGELO ARLEOangelo.arleo@upmc.frCNRS - University Pierre & Marie Curie Paris 6

We study an encoding/decoding mechanism accounting for the relative spike timing of the signals propagating from peripheral nerve fibers to second-order somatosensory neurons in the cuneate nucleus (CN). The CN is modeled as a population of spiking neurons receiving as inputs the spatiotemporal responses of real mechanoreceptors obtained via microneurography recordings in humans. The efficiency of the haptic discrimination process is quantified by a novel definition of entropy that takes into full account the metrical properties of the spike train space. This measure proves to be a suitable decoding scheme for generalizing the classical Shannon entropy to spike-based neural codes. It permits an assessment of neurotransmission in the presence of a large output space (i.e. hundreds of spike trains) with 1 ms temporal precision. It is shown that the CN population code performs a complete discrimination of 81 distinct stimuli already within 35 ms of the first afferent spike, whereas a partial discrimination (80% of the maximum information transmission) is possible as rapidly as 15 ms. This study suggests that the CN may not constitute a mere synaptic relay along the somatosensory pathway but, rather, it may convey optimal contextual accounts (in terms of fast and reliable information transfer) of peripheral tactile inputs to downstream structures of the central nervous system.

Subject Area: Neuroscience

M20 Bayesian estimation of orientation preference maps

Jakob Macke	Jakob.Macke@gmx.de
Sebastian Gerwinn	sgerwinn@tuebingen.mpg.de
MPI for Biological Cybernetics &	University of Tübingen
LEONARD WHITE	white033@mc.duke.edu
Duke University Medical Center	
Matthias Kaschube	kaschube@princeton.edu
Princeton University	
Matthias Bethge	mbethge@tuebingen.mpg.de
MPI for Biological Cybernetics &	University of Tübingen

Imaging techniques such as optical imaging of intrinsic signals, 2-photon calcium imaging and voltage sensitive dye imaging can be used to measure the functional organization of visual cortex across different spatial scales. Here, we present Bayesian methods based on Gaussian processes for extracting topographic maps from functional imaging data. In particular, we focus on the estimation of orientation preference maps (OPMs) from intrinsic signal imaging data. We model the underlying map as a bivariate Gaussian process, with a prior covariance function that reflects known properties of OPMs, and a noise covariance adjusted to the data. The posterior mean can be interpreted as an optimally smoothed estimate of the map, and can be used for model based interpolations of the map from sparse measurements. By sampling from the posterior distribution, we can get error bars on statistical properties such as preferred orientations, pinwheel locations or -counts. Finally, the use of an explicit probabilistic model facilitates interpretation of parameters and provides the basis for decoding studies. We demonstrate our model both on simulated data and on intrinsic signaling data from ferret visual cortex. **Subject Area:** Neuroscience

M21 Replacing supervised classification learning by Slow Feature Analysis in spiking neural networks

Stefan Klampfl	klampfl@igi.tugraz.at
Wolfgang Maass	maass@igi.tugraz.at
Graz University of Technology	

Many models for computations in recurrent networks of neurons assume that the network state moves from some initial state to some fixed point attractor or limit cycle that represents the output of the computation. However experimental data show that in response to a sensory stimulus the network state moves from its initial state through a trajectory of network states and eventually returns to the initial state, without reaching an attractor or limit cycle in between. This type of network response, where salient information about external stimuli is encoded in characteristic trajectories of continuously varying network states, raises the question how a neural system could compute with such code, and arrive for example at a temporally stable classification of the external stimulus. We show that a known unsupervised learning algorithm, Slow Feature Analysis (SFA), could be an important ingredient for extracting stable information from these network trajectories. In fact, if sensory stimuli are more often followed by another stimulus from the same class than by a stimulus from another class, SFA approaches the classification capability of Fisher's Linear Discriminant (FLD), a powerful algorithm for supervised learning. We apply this principle to simulated cortical microcircuits, and show that it enables readout neurons to learn discrimination of spoken digits and detection of repeating firing patterns within a stream of spike trains with the same firing statistics, without requiring any supervision for learning. Subject Area: Neuroscience

Spotlight presentation, Monday, 6:30.

M22 fMRI-Based Inter-Subject Cortical Alignment Using Functional Connectivity

Bryan Conroy	bconroy@princeton.edu
Ben Singer	bdsinger@princeton.edu
Princeton University	
JAMES HAXBY	james.v.haxby@dartmouth.edu
Dartmouth College	
Peter Ramadge	ramadge@princeton.edu
Princeton University	

The inter-subject alignment of functional MRI (fMRI) data is important for improving the statistical power of fMRI group analyses. In contrast to existing anatomically-based methods, we propose a novel multi-subject algorithm that derives a functional correspondence by aligning spatial patterns of functional connectivity across a set of We test our method on fMRI data collected during a movie viewing subjects. By cross-validating the results of our algorithm, we show that the experiment. correspondence successfully generalizes to a secondary movie dataset not used to derive the alignment.

Subject Area: Neuroscience

M23 Correlation Coefficients are Insufficient for Analyzing Spike Count Dependencies

Arno Onken	aonken@cs.tu-berlin.de
Berlin Institute of Technology	
Steffen Grünewälder	S.Grunewalder@cs.ucl.ac.uk
University College London	
Klaus Obermayer	oby@cs.tu-berlin.de
Berlin Institute of Technology	

The linear correlation coefficient is typically used to characterize and analyze dependencies of neural spike counts. Here, we show that the correlation coefficient is in general insufficient to characterize these dependencies. We construct two neuron spike count models with Poisson-like marginals and vary their dependence structure using copulas. To this end, we construct a copula that allows to keep the spike counts uncorrelated while varying their dependence strength. Moreover, we employ a network of leaky integrate-and-fire neurons to investigate whether weakly correlated spike counts with strong dependencies are likely to occur in real networks. We find that the entropy of uncorrelated but dependent spike count distributions can deviate from the corresponding distribution with independent components by more than 25% and that weakly correlated but strongly dependent spike counts are very likely to occur in biological networks. Finally, we introduce a test for deciding whether the dependence structure of distributions with Poisson-like marginals is well characterized by the linear correlation coefficient and verify it for different copula-based models. Subject Area: Neuroscience

M24 Estimating image bases for visual image reconstruction from human brain activity

Yusuke Fujiwara Yoichi Miyawaki Yukiyasu Kamitani ATR, JAPAN yureisoul@gmail.com yoichi_m@atr.jp kmtn@atr.jp

Image representation based on image bases provides a framework for understanding neural representation of visual perception. A recent fMRI study has shown that arbitrary contrast-defined visual images can be reconstructed from fMRI activity patterns using a combination of multi-scale local image bases. In the reconstruction model, the mapping from an fMRI activity pattern to the contrasts of the image bases was learned from measured fMRI responses to visual images. But the shapes of the images bases were fixed, and thus may not be optimal for reconstruction. Here, we propose a method to build a reconstruction model in which image bases are automatically extracted from the measured data. We constructed a probabilistic model that relates the fMRI activity space to the visual image space via a set of latent variables. The mapping from the latent variables to the visual image space can be regarded as a set of image bases. We found that spatially localized, multi-scale image bases were estimated near the fovea, and that the model using the estimated image bases was able to accurately reconstruct novel visual images. The proposed method provides a means to discover a novel functional mapping between stimuli and brain activity patterns. Subject Area: Neuroscience

M25 Reconstruction of Sparse Circuits Using Multi-neuronal Excitation (RESCUME)

Tao Hu	hut@janelia.hhmi.org
Anthony Leonardo	leonardoa@janelia.hhmi.org
Howard Hughes Medical Institute,	Janelia Farm Research Campus
Dmitri Chklovskii	chklovskiid@janelia.hhmi.org
Janelia Farm Research Campus, H	oward Hughes Medical Institute

One of the central problems in neuroscience is reconstructing synaptic connectivity in neural circuits. Synapses onto a neuron can be probed by sequentially stimulating potentially pre-synaptic neurons while monitoring the membrane voltage of the post-synaptic neuron. Reconstructing a large neural circuit using such a "brute force" approach is rather time-consuming and inefficient because the connectivity in neural circuits is sparse. Instead, we propose to measure a post-synaptic neuron's voltage while stimulating simultaneously multiple randomly chosen potentially pre-synaptic neurons. To extract the weights of individual synaptic connections we apply a decoding algorithm recently developed for compressive sensing. Compared to the brute force approach, our method promises significant time savings that grow with the size of the circuit. We use computer simulations to find optimal stimulation parameters and explore the feasibility of our reconstruction method under realistic experimental conditions including noise and non-linear synaptic integration. Multiple-neuron stimulation allows reconstructing synaptic connectivity just from the spiking activity of post-synaptic neurons, even when sub-threshold voltage is unavailable. By using calcium indicators, voltage-sensitive dyes, or multi-electrode arrays one could monitor activity of multiple post-synaptic neurons simultaneously, thus mapping their synaptic inputs in parallel, potentially reconstructing a complete neural circuit.

Subject Area: Neuroscience

M26 Augmenting Feature-driven fMRI Analyses: Semi-supervised learning and resting state activity

Mathew Blaschko	blaschko@robots.ox.ac.uk
University of Oxford	
Jacquelyn Shelton	jshelton@tuebingen.mpg.de
Universitaet Tuebingen	
Andreas Bartels	andreas.bartels@tuebingen.mpg.de
MPI for Biological Cybernetics	

Resting state activity is brain activation that arises in the absence of any task, and is usually measured in awake subjects during prolonged fMRI scanning sessions where the only instruction given is to close the eyes and do nothing. It has been recognized in recent years that resting state activity is implicated in a wide variety of brain function. While certain networks of brain areas have different levels of activation at rest and during a task, there is nevertheless significant similarity between activations in the two cases. This suggests that recordings of resting state activity can be used as a source of unlabeled data to augment discriminative regression techniques in a semi-supervised setting. We evaluate this setting empirically yielding three main results: (i) regression tends to be improved by the use of Laplacian regularization even when no additional unlabeled data are available, (ii) resting state data may have a similar marginal distribution to that recorded during the execution of a visual processing task reinforcing the hypothesis that these conditions have similar types of activation, and (iii) this source of information can be broadly exploited to improve the robustness of empirical inference in fMRI studies, an inherently data poor domain.

Subject Area: Neuroscience

M27 Unsupervised Feature Selection for the k-means Clustering Problem

Christos Boutsidis	boutsc@cs.rpi.edu
Rensselaer Polytechnic Instit.	
Michael Mahoney	mmahoney@cs.stanford.edu
Stanford University	
Petros Drineas	drinep@cs.rpi.edu
Rensselaer Polytechnic Institute	

We present a novel feature selection algorithm for the k-means clustering problem. Our algorithm is randomized and, assuming an accuracy parameter $\epsilon \in (0, 1)$, selects and appropriately rescales in an unsupervised manner $\Theta(k \log(k/\epsilon)/\epsilon^2)$ features from a dataset of arbitrary dimensions. We prove that, if we run any γ -approximate k-means algorithm $(\gamma \geq 1)$ on the features selected using our method, we can find a $(1+(1+\epsilon)\gamma)$ -approximate partition with high probability.

Subject Area: Unsupervised and Semi-supervised Learning

M28 Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization

JOHN WRIGHT	jowrig@microsoft.com	
Microsoft Research		
Arvind Ganesh	abalasu2@uiuc.edu	
University of Illinois		
Shankar Rao	srrao@uiuc.edu	
HRL Laboratories		
YIGANG PENG	pengyigang@gmail.com	
Microsoft Research Asia		
YI MA	yima@uiuc.edu	
University of Illinois at Urbana-Champaign		

Principal component analysis is a fundamental operation in computational data analysis, with myriad applications ranging from web search to bioinformatics to computer vision and image analysis. However, its performance and applicability in real scenarios are limited by a lack of robustness to outlying or corrupted observations. This paper considers the idealized "robust principal component analysis" problem of recovering a low rank matrix A from corrupted observations D = A + E. Here, the error entries E can be arbitrarily large (modeling grossly corrupted observations common in visual and bioinformatic data), but are assumed to be sparse. We prove that most matrices A can be efficiently and exactly recovered from most error sign-and-support patterns, by solving a simple convex program. Our result holds even when the rank of A grows nearly proportionally (up to a logarithmic factor) to the dimensionality of the observation space and the number of errors E grows in proportion to the total number of entries in the matrix. A by-product of our analysis is the first proportional growth results for the related problem of completing a low-rank matrix from a small fraction of its entries. Simulations and real-data examples corroborate the theoretical results, and suggest potential applications in computer vision. Subject Area: Unsupervised and Semi-supervised Learning

M29 Breaking Boundaries Between Induction Time and Diagnosis Time Active Information Acquisition

ASHISH KAPOOR akapoor@microsoft.com ERIC HORVITZ horvitz@microsoft.com Microsoft Research

There has been a clear distinction between induction or training time and diagnosis time active information acquisition. While active learning during induction focuses on acquiring data that promises to provide the best classification model, the goal at diagnosis time focuses completely on next features to observe about the test case at hand in order to make better predictions about the case. We introduce a model and inferential methods that breaks this distinction. The methods can be used to extend case libraries under a budget but, more fundamentally, provide a framework for guiding agents to collect data under scarce resources, focused by diagnostic challenges. This extension to active learning leads to a new class of policies for real-time diagnosis, where recommended informationgathering sequences include actions that simultaneously seek new data for the case at hand and for cases in the training set.

Subject Area: Unsupervised and Semi-supervised Learning

M30 Manifold Embeddings for Model-Based Reinforcement Learning under Partial Observability

KEITH BUSH JOELLE PINEAU McGill University kbush@cs.mcgill.ca jpineau@cs.mcgill.ca

Interesting real-world datasets often exhibit nonlinear, noisy, continuous-valued states that are unexplorable, are poorly described by first principles, and are only partially observable. If partial observability can be overcome, these constraints suggest the use of model-based reinforcement learning. We experiment with manifold embeddings as the reconstructed observable state-space of an off-line, model-based reinforcement learning approach to control. We demonstrate the embedding of a system changes as a result of learning and that the best performing embeddings well-represent the dynamics of both the uncontrolled and adaptively controlled system. We apply this approach in simulation to learn a neurostimulation policy that is more efficient in treating epilepsy than conventional policies. We then demonstrate the learned policy completely suppressing seizures in realworld neurostimulation experiments on actual animal brain slices. **Subject Area:** Unsupervised and Semi-supervised Learning

M31 Modeling Social Annotation Data with Content Relevance using a Topic Model

TOMOHARU IWATAiwata@cslab.kecl.ntt.co.jpNTT Communication Science LaboratoriesTAKESHI YAMADAYAMADAYAMADAVAONORI UEDAUeda@cslab.kecl.ntt.co.jpNTT Communication Science Laboratories

We propose a probabilistic topic model for analyzing and extracting content-related annotations from noisy annotated discrete data such as web pages stored in social bookmarking services. In these services, since users can attach annotations freely, some annotations do not describe the semantics of the content, thus they are noisy, i.e. not content-related. The extraction of content-related annotations can be used as a preprocessing step in machine learning tasks such as text classification and image recognition, or can improve information retrieval performance. The proposed model is a generative model for content and annotations, in which the annotations are assumed to originate either from topics that generated the content or from a general distribution unrelated to the content. We demonstrate the effectiveness of the proposed method by using synthetic data and real social annotation data for text and images.

Subject Area: Unsupervised and Semi-supervised Learning

M32 Rank-Approximate Nearest Neighbor Search: Retaining Meaning and Speed in High Dimensions

Parikshit Ram	p.ram@gatech.edu
Georgia Institute of Technolog	
Dongryeol Lee	dongryel@cc.gatech.edu
Hua Ouyang	houyang@gatech.edu
Alexander Gray	agray@cc.gatech.edu
Georgia Institute of Technology	

The long-standing problem of efficient nearest-neighbor (NN) search has ubiquitous applications ranging from astrophysics to MP3 fingerprinting to bioinformatics to movie recommendations. As the dimensionality of the dataset increases, exact NN search becomes computationally prohibitive; (1+eps)-distance-approximate NN search can provide large speedups but risks losing the meaning of NN search present in the ranks (ordering) of the distances. This paper presents a simple, practical algorithm allowing the user to, for the first time, directly control the true accuracy of NN search (in terms of ranks) while still achieving the large speedups over exact NN. Experiments with high-dimensional datasets show that it often achieves faster and more accurate results than the best-known distance-approximate method, with much more stable behavior. Subject Area: Unsupervised and Semi-supervised Learning

M33 Distribution Matching for Transduction

Novi Quadrianto novi.quad@gmail.com SML-NICTA and RSISE-ANU JAMES PETTERSON NICTA ALEX SMOLA Yahoo!

james.petterson@nicta.com.au

alex@smola.org
Many transductive inference algorithms assume that distributions over training and test estimates should be related, e.g. by providing a large margin of separation on both sets. We use this idea to design a transduction algorithm which can be used without modification for classification, regression, and structured estimation. At its heart we exploit the fact that for a good learner the distributions over the outputs on training and test sets should match. This is a classical two-sample problem which can be solved efficiently in its most general form by using distance measures in Hilbert Space. It turns out that a number of existing heuristics can be viewed as special cases of our approach.

Subject Area: Unsupervised and Semi-supervised Learning

M34 Heavy-Tailed Symmetric Stochastic Neighbor Embedding

Zhirong Yang	zhirong.yang@tkk.fi
CUHK,TKK	
Irwin King	king@cse.cuhk.edu.hk
Zenglin Xu	zlxu@cse.cuhk.edu.hk
CUHK	
Erkki Oja	erkki.oja@tkk.fi
ТКК	

Stochastic Neighbor Embedding (SNE) has shown to be quite promising for data visualization. Currently, the most popular implementation, t-SNE, is restricted to a particular Student t-distribution as its embedding distribution. Moreover, it uses a gradient descent algorithm that may require users to tune parameters such as the learning step size, momentum, etc., in finding its optimum. In this paper, we propose the Heavy-tailed Symmetric Stochastic Neighbor Embedding (HSSNE) method, which is a generalization of the t-SNE to accommodate various heavy-tailed embedding similarity functions. With this generalization, we are presented with two difficulties. The first is how to select the best embedding similarity among all heavy-tailed functions and the second is how to optimize the objective function once the heave-tailed function has been selected. Our contributions then are: (1) we point out that various heavy-tailed embedding similarities can be characterized by their negative score functions. Based on this finding, we present a parameterized subset of similarity functions for choosing the best tail-heaviness for HSSNE; (2) we present a fixed-point optimization algorithm that can be applied to all heavy-tailed functions and does not require the user to set any parameters; and (3) we present two empirical studies, one for unsupervised visualization showing that our optimization algorithm runs as fast and as good as the best known t-SNE implementation and the other for semi-supervised visualization showing quantitative superiority using the homogeneity measure as well as qualitative advantage in cluster separation over t-SNE.

Subject Area: Unsupervised and Semi-supervised Learning

Spotlight presentation, Monday, 6:30.

M35 A Rate Distortion Approach for Semi-Supervised Conditional Random Fields

YANG WANG	ywang12@cs.sfu.ca
Gholamreza Haffari	ghaffar1@cs.sfu.ca
Simon Fraser University	
Shaojun Wang	shaojun.wang@wright.edu
Wright State University	
Greg Mori	mori@cs.sfu.ca
Simon Fraser U	

We propose a novel information theoretic approach for semi-supervised learning of conditional random fields. Our approach defines a training objective that combines the conditional likelihood on labeled data and the mutual information on unlabeled data. Different from previous minimum conditional entropy semi-supervised discriminative learning methods, our approach can be naturally cast into the rate distortion theory framework in information theory. We analyze the tractability of the framework for structured prediction and present a convergent variational training algorithm to defy the combinatorial explosion of terms in the sum over label configurations. Our experimental results show that the rate distortion approach outperforms standard l_2 regularization and minimum conditional entropy regularization on both multi-class classification and sequence labeling problems.

Subject Area: Unsupervised and Semi-supervised Learning

M36 Optimal Scoring for Unsupervised Learning

Zhihua Zhang	zhzhang@gmail.com
Guang Dai	daiguang1160gmail.com
Zhejiang University	

We are often interested in casting classification and clustering problems in a regression framework, because it is feasible to achieve some statistical properties in this framework by imposing some penalty criteria. In this paper we illustrate optimal scoring, which was originally proposed for performing Fisher linear discriminant analysis by regression, in the application of unsupervised learning. In particular, we devise a novel clustering algorithm that we call optimal discriminant clustering (ODC). We associate our algorithm with the existing unsupervised learning algorithms such as spectral clustering, discriminative clustering and sparse principal component analysis. Thus, our work shows that optimal scoring provides a new approach to the implementation of unsupervised learning. This approach facilitates the development of new unsupervised learning algorithms. **Subject Area:** Unsupervised and Semi-supervised Learning

M37 An Online Algorithm for Large Scale Image Similarity Learning

gal@robotics.stanford.edu
0
urish22@gmail.com
vasharma@google.com
bengio@google.com

Learning a measure of similarity between pairs of objects is a fundamental problem in machine learning. It stands in the core of classification methods like kernel machines, and is particularly useful for applications like searching for images that are similar to a given image or finding videos that are relevant to a given video. In these tasks, users look for objects that are not only visually similar but also semantically related to a given object. Unfortunately, current approaches for learning similarity may not scale to large datasets with high dimensionality, especially when imposing metric constraints on the learned similarity. We describe OASIS, a method for learning pairwise similarity that is fast and scales linearly with the number of objects and the number of non-zero features. Scalability is achieved through online learning of a bilinear model over sparse representations using a large margin criterion and an efficient hinge loss cost. OASIS is accurate at a wide range of scales: on a standard benchmark with thousands of images, it is more precise than state-of-the-art methods, and faster by orders of magnitude. On 2 million images collected from the web. OASIS can be trained within 3 days on a single CPU. The non-metric similarities learned by OASIS can be transformed into metric similarities, achieving higher precisions than similarities that are learned as metrics in the first place. This suggests an approach for learning a metric from data that is larger by an order of magnitude than was handled before.

Subject Area: Unsupervised and Semi-supervised Learning

M38 Rethinking LDA: Why Priors Matter

Hanna Wallach	wallach@cs.umass.edu
David Mimno	mimno@cs.umass.edu
Andrew McCallum	mccallum@cs.umass.edu
University of Massachusetts A	mherst

Implementations of topic models typically use symmetric Dirichlet priors with fixed concentration parameters, with the implicit assumption that such "smoothing parameters" have little practical effect. In this paper, we explore several classes of structured priors for topic models. We find that an asymmetric Dirichlet prior over the document-topic distributions has substantial advantages over a symmetric prior, while an asymmetric prior over the topic-word distributions provides no real benefit. Approximation of this prior structure through simple, efficient hyperparameter optimization steps is sufficient to achieve these performance gains. The prior structure we advocate substantially increases the robustness of topic models to variations in the number of topics and to the highly skewed word frequency distributions common in natural language. Since this prior structure can be implemented using efficient algorithms that add negligible cost beyond standard inference techniques, we recommend it as a new standard for topic modeling.

Subject Area: Unsupervised and Semi-supervised Learning

Spotlight presentation, Monday, 6:30.

M39 Semi-supervised Learning using Sparse Eigenfunction Bases

KAUSHIK SINHA MIKHAIL BELKIN Ohio State University

sinhak@cse.ohio-state.edu
mbelkin@cse.ohio-state.edu

We present a new framework for semi-supervised learning with sparse eigenfunction bases of kernel matrices. It turns out that when the *cluster assumption* holds, that is, when the high density regions are sufficiently separated by low density valleys, each high density area corresponds to a unique representative eigenvector. Linear combination of such eigenvectors (or, more precisely, of their Nystrom extensions) provide good candidates for good classification functions. By first choosing an appropriate basis of these eigenvectors from unlabeled data and then using labeled data with Lasso to select a classifier in the span of these eigenvectors, we obtain a classifier, which has a very sparse representation in this basis. Importantly, the sparsity appears naturally from the cluster assumption. Experimental results on a number of real-world data-sets show that our method is competitive with the state of the art semi-supervised learning algorithms and outperforms the natural base-line algorithm (Lasso in the Kernel PCA basis). **Subject Area:** Unsupervised and Semi-supervised Learning

M40 Streaming k-means approximation

NIR AILON	nailon@gmail.com
Google	
RAGESH JAISWAL	rjaiswal@gmail.com
Claire Monteleoni	cmontel@ccls.columbia.edu
Columbia University	

We provide a clustering algorithm that approximately optimizes the k-means objective, in the one-pass streaming setting. We make no assumptions about the data, and our algorithm is very light-weight in terms of memory, and computation. This setting is applicable to unsupervised learning on massive data sets, or resource-constrained devices. The two main ingredients of our theoretical work are: a derivation of an extremely simple pseudo-approximation batch algorithm for k-means, in which the algorithm is allowed to output more than k centers (based on the recent "k-means++"), and a streaming clustering algorithm in which batch clustering algorithms are performed on small inputs (fitting in memory) and combined in a hierarchical manner. Empirical evaluations on real and simulated data reveal the practical utility of our method.

Subject Area: Unsupervised and Semi-supervised Learning

M41 Polynomial Semantic Indexing

BING BAI	bbai@nec-labs.com
NEC Labs America, INC	
JASON WESTON	jasonw@nec-labs.com
NEC Labs America	
David Grangier	dgrangier@nec-labs.com
Ronan Collobert	collober@nec-labs.com
Kunihiko Sadamasa	kunihiko@nec-labs.com
Yanjun Qi	yanjun@nec-labs.com
NEC Labs America, INC	
Corinna Cortes	corinna@google.com
Google Research	
Mehryar Mohri	mohri@cs.nyu.edu
Courant Institute of Mathematical	Sciences and Google Research

We present a class of nonlinear (polynomial) models that are discriminatively trained to directly map from the word content in a query-document or document-document pair to a ranking score. Dealing with polynomial models on word features is computationally challenging. We propose a low rank (but diagonal preserving) representation of our polynomial models to induce feasible memory and computation requirements. We provide an empirical study on retrieval tasks based on Wikipedia documents, where we obtain state-of-the-art performance while providing realistically scalable methods. **Subject Area:** Applications

M42 Matrix Completion from Power-Law Distributed Samples

Raghu Meka	raghu@cs.utexas.edu
Prateek Jain	pjain@cs.utexas.edu
University of Texas at Austin	
Inderjit Dhillon	inderjit@cs.utexas.edu
University of Texas	

The low-rank matrix completion problem is a fundamental problem with many important applications. Recently, Candes & Recht, Keshavan et al. and Candes & Tao obtained the first non-trivial theoretical results for the problem assuming that the observed entries are sampled uniformly at random. Unfortunately, most real-world datasets do not satisfy this assumption, but instead exhibit power-law distributed samples. In this paper, we propose a graph theoretic approach to matrix completion that solves the problem for more realistic sampling models. Our method is easier to analyze than previous methods with the analysis reducing to computing the threshold for complete cascades in random graphs, a problem of independent interest. By analyzing the graph theoretic problem, we show that our method achieves exact recovery when the observed entries are sampled from the Chung-Lu-Vu model, which can generate power-law distributed graphs. We also hypothesize that our algorithm solves the matrix completion problem from an optimal number of entries for the popular preferential attachment model and provide strong empirical evidence for the claim. Furthermore, our method is easier to implement and is substantially faster than existing methods. We demonstrate the effectiveness of our method on examples when the low-rank matrix is sampled according to the prevalent random graph models for complex networks and also on the Netflix challenge dataset. Subject Area: Applications

M43 Learning from Neighboring Strokes: Combining Appearance and Context for Multi-Domain Sketch Recognition

Tom Ouyang Randall Davis MIT CSAIL ouyang@csail.mit.edu davis@csail.mit.edu We propose a new sketch recognition framework that combines a rich representation of low level visual appearance with a graphical model for capturing high level relationships between symbols. This joint model of appearance and context allows our framework to be less sensitive to noise and drawing variations, improving accuracy and robustness. The result is a recognizer that is better able to handle the wide range of drawing styles found in messy freehand sketches. We evaluate our work on two real-world domains, molecular diagrams and electrical circuit diagrams, and show that our combined approach significantly improves recognition performance. Subject Area: Applications

M44 Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise

Jacob Whitehill	jake@mplab.ucsd.edu
Paul Ruvolo	pruvolo@cs.ucsd.edu
TING-FAN WU	t3wu@ucsd.edu
Jacob Bergsma	jbergsma@mplab.ucsd.edu
UCSD	
JAVIER MOVELLAN	movellan@mplab.ucsd.edu
University of California	-

Modern machine learning-based approaches to computer vision require very large databases of labeled images. Some contemporary vision systems already require on the order of millions of images for training (e.g., Omron face detector). While the collection of these large databases is becoming a bottleneck, new Internet-based services that allow labelers from around the world to be easily hired and managed provide a promising solution. However, using these services to label large databases brings with it new theoretical and practical challenges: (1) The labelers may have wide ranging levels of expertise which are unknown a priori, and in some cases may be adversarial; (2) images may vary in their level of difficulty; and (3) multiple labels for the same image must be combined to provide an estimate of the actual label of the image. Probabilistic approaches provide a principled way to approach these problems. In this paper we present a probabilistic model and use it to simultaneously infer the label of each image, the expertise of each labeler, and the difficulty of each image. On both simulated and real data, we demonstrate that the model outperforms the commonly used "Majority" Vote" heuristic for inferring image labels, and is robust to both adversarial and noisy labelers.

Subject Area: Cognitive Science

M45 Individuation, Identification and Object Discovery

CHARLES KEMP ckemp@cmu.edu Alan Jern ajern@cmu.edu Carnegie Mellon University Fei Xu University of California, Berkeley

fei_xu@berkeley.edu

Humans are typically able to infer how many objects their environment contains and to recognize when the same object is encountered twice. We present a simple statistical model that helps to explain these abilities and evaluate it in three behavioral experiments. Our first experiment suggests that humans rely on prior knowledge when deciding whether an object token has been previously encountered. Our second and third experiments suggest that humans can infer how many objects they have seen and can learn about categories and their properties even when they are uncertain about which tokens are instances of the same object.

Subject Area: Cognitive Science

M46 The Wisdom of Crowds in the Recollection of Order Information

MARK STEYVERS	mark.steyvers@uci.edu
Michael Lee	mdlee@uci.edu
University of California	
Brent Miller	brentm@uci.edu
Pernille Hemmer	phemmer@uci.edu
University of California, Irvine	

When individuals independently recollect events or retrieve facts from memory, how can we aggregate these retrieved memories to reconstruct the actual set of events or facts? In this research, we report the performance of individuals in a series of general knowledge tasks, where the goal is to reconstruct from memory the order of historic events, or the order of items along some physical dimension. We introduce two Bayesian models for aggregating order information based on a Thurstonian approach and Mallows model. Both models assume that each individual's reconstruction is based on either a random permutation of the unobserved ground truth, or by a pure guessing strategy. We apply MCMC to make inferences about the underlying truth and the strategies employed by individuals. The models demonstrate a "wisdom of crowds" effect, where the aggregated orderings are closer to the true ordering than the orderings of the best individual. **Subject Area:** Cognitive Science

M47 Measuring model complexity with the prior predictive

WOLF VANPAEMEL wolf.vanpaemel@psy.kuleuven.be Kuleuven

In the last few decades, model complexity has received a lot of press. While many methods have been proposed that jointly measure a model's descriptive adequacy and its complexity, few measures exist that measure complexity in itself. Moreover, existing measures ignore the parameter prior, which is an inherent part of the model and affects the complexity. This paper presents a stand alone measure for model complexity, that takes the number of parameters, the functional form, the range of the parameters and the parameter prior into account. This Prior Predictive Complexity (PPC) is an intuitive and easy to compute measure. It starts from the observation that model complexity is the property of the model that enables it to fit a wide range of outcomes. The PPC then measures how wide this range exactly is.

Subject Area: Cognitive Science

M48 Structured output regression for detection with partial truncation

ANDREA VEDALDI ANDREW ZISSERMAN University of Oxford

vedaldi@robots.ox.ac.uk az@robots.ox.ac.uk

We develop a structured output model for object category detection that explicitly accounts for alignment, multiple aspects and partial truncation in both training and inference. The model is formulated as large margin learning with latent variables and slack rescaling, and both training and inference are computationally efficient. We make the following contributions: (i) we note that extending the Structured Output Regression formulation of Blaschko and Lampert (ECCV 2008) to include a bias term significantly improves performance; (ii) that alignment (to account for small rotations and anisotropic scalings) can be included as a latent variable and efficiently determined and implemented; (iii) that the latent variable extends to multiple aspects (e.g. left facing, right facing, front) with the same formulation; and (iv), most significantly for performance, that truncated and truncated instances can be included in both training and inference with an explicit truncation mask. We demonstrate the method by training and testing on the PASCAL VOC 2007 data set - training includes the truncated examples, and in testing object instances are detected at multiple scales, alignments, and with significant truncations.

Subject Area: Vision

M49 Dual Averaging Method for Regularized Stochastic Learning and Online Optimization

LIN XIAO

lin.xiao@microsoft.com

Microsoft Research

We consider regularized stochastic learning and online optimization problems, where the objective function is the sum of two convex terms: one is the loss function of the learning task, and the other is a simple regularization term such as L1-norm for sparsity. We develop a new online algorithm, the regularized dual averaging method, that can explicitly exploit the regularization structure in an online setting. In particular, at each iteration, the learning variables are adjusted by solving a simple optimization problem that involves the running average of all past subgradients of the loss functions and the whole regularization term, not just its subgradient. This method achieves the optimal convergence rate and often enjoys a low complexity per iteration similar as the standard stochastic gradient method. Computational experiments are presented for the special case of sparse online learning using L1-regularization.

Subject Area: Optimization

M50 Generalization Errors and Learning Curves for Regression with Multi-task Gaussian Processes

KIAN MING CHAI University of Edinburgh k.m.a.chai@ed.ac.uk

We provide some insights into how task correlations in multi-task Gaussian process (GP) regression affect the generalization error and the learning curve. We analyze the asymmetric two-task case, where a secondary task is to help the learning of a primary task. Within this setting, we give bounds on the generalization error and the learning curve of the primary task. Our approach admits intuitive understandings of the multi-task GP by relating it to single-task GPs. For the case of one-dimensional input-space under optimal sampling with data only for the secondary task, the limitations of multi-task GP can be quantified explicitly. **Subject Area:** Theory

M51 DUOL: A Double Updating Approach for Online Learning

Peilin Zhao	ZHAO0106@ntu.edu.sg
Steven Chu-Hong Hoi	chhoi@ntu.edu.sg
Nanyang Technological University	
Rong Jin	rongjin@cse.msu.edu
Michigan State University	

In most online learning algorithms, the weights assigned to the misclassified examples (or support vectors) remain unchanged during the entire learning process. This is clearly insufficient since when a new misclassified example is added to the pool of support vectors, we generally expect it to affect the weights for the existing support vectors. In this paper, we propose a new online learning method, termed "Double Updating Online Learning", or "DUOL" for short. Instead of only assigning a fixed weight to the misclassified example received in current trial, the proposed online learning algorithm also tries to update the weight for one of the existing support vectors. We show that the mistake bound can be significantly improved by the proposed online learning method. Encouraging experimental results show that the proposed technique is in general considerably more effective than the state-of-the-art online learning algorithms. **Subject Area:** Theory

M52 Monte Carlo Sampling for Regret Minimization in Extensive Games

Marc Lanctot	lanctot@ualberta.ca
University of Alberta	
Kevin Waugh	waugh@cs.cmu.edu
Carnegie Mellon University	
Martin Zinkevich	maz@yahoo-inc.com
Yahoo!	
Michael Bowling	bowling@cs.ualberta.ca
University of Alberta	

Sequential decision-making with multiple agents and imperfect information is commonly modeled as an extensive game. One efficient method for computing Nash equilibria in large, zero-sum, imperfect information games is counterfactual regret minimization (CFR). In the domain of poker, CFR has proven effective, particularly when using a domainspecific augmentation involving chance outcome sampling. In this paper, we describe a general family of domain independent CFR sample-based algorithms called Monte Carlo counterfactual regret minimization (MCCFR) of which the original and poker-specific versions are special cases. We start by showing that MCCFR performs the same regret updates as CFR on expectation. Then, we introduce two sampling schemes: *outcome sampling* and *external sampling*, showing that both have bounded overall regret with high probability. Thus, they can compute an approximate equilibrium using self-play. Finally, we prove a new tighter bound on the regret for the original CFR algorithm and relate this new bound to MCCFR's bounds. We show empirically that, although the sample-based algorithms require more iterations, their lower cost per iteration can lead to dramatically faster convergence in various games.

Subject Area: Theory

M53 Beyond Convexity: Online Submodular Minimization

Elad Hazan	ehazan@cs.princeton.edu
IBM	
Satyen Kale	skale@yahoo-inc.com
Yahoo! Research	

We consider an online decision problem over a discrete space in which the loss function is submodular. We give algorithms which are computationally efficient and are Hannanconsistent in both the full information and bandit settings. **Subject Area:** Theory

M54 Asymptotically Optimal Regularization in Smooth Parametric Models

Percy Liang	pliang@eecs.berkeley.edu
University of California	
Francis Bach	francis.bach@mines.org
Ecole Normale Superieure	
Guillaume Bouchard	guillaume.bouchard@xrce.xerox.com
Xerox Research Centre Europe	
Michael Jordan	jordan@cs.berkeley.edu
University of California	

Many types of regularization schemes have been employed in statistical learning, each one motivated by some assumption about the problem domain. In this paper, we present a unified asymptotic analysis of smooth regularizers, which allows us to see how the validity of these assumptions impacts the success of a particular regularizer. In addition, our analysis motivates an algorithm for optimizing regularization parameters, which in turn can be analyzed within our framework. We apply our analysis to several examples, including hybrid generative-discriminative learning and multi-task learning. **Subject Area:** Theory

M55 Compressed Least-Squares Regression

Odalric Maillard Remi Munos INRIA odalric.maillard@inria.fr remi.munos@inria.fr

We consider the problem of learning, from K input data, a regression function in a function space of high dimension N using projections onto a random subspace of lower dimension M. From any linear approximation algorithm using empirical risk minimization (possibly penalized), we provide bounds on the excess risk of the estimate computed in the projected subspace (compressed domain) in terms of the excess risk of the estimate built in the high-dimensional space (initial domain). We apply the analysis to the ordinary Least-Squares regression and show that by choosing $M = O(\sqrt{K})$, the estimation error (for the quadratic loss) of the "Compressed Least Squares Regression" is $O(1/\sqrt{K})$ up to logarithmic factors. We also discuss the numerical complexity of several algorithms (both in initial and compressed domains) as a function of N, K, and M. **Subject Area:** Theory

M56 Robust Nonparametric Regression with Metric-Space Valued Output

MATTHIAS HEIN hein@cs.uni-sb.de Saarland University

Motivated by recent developments in manifold-valued regression we propose a family of nonparametric kernel-smoothing estimators with metric-space valued output including a robust median type estimator and the classical Frechet mean. Depending on the choice of the output space and the chosen metric the estimator reduces to partially well-known procedures for multi-class classification, multivariate regression in Euclidean space, regression with manifold-valued output and even some cases of structured output learning. In this paper we focus on the case of regression with manifold-valued input and output. We show pointwise and Bayes consistency for all estimators in the family for the case of manifold-valued output and illustrate the robustness properties of the estimator with experiments.

Subject Area: Theory

M57 On Invariance in Hierarchical Models

JAKE BOUVRIE jvb@mit.edu MIT LORENZO ROSASCO lrosasco@mit.edu Massachusetts Institute of Technology TOMASO POGGIO tp@ai.mit.edu MIT A goal of central importance in the study of hierarchical models for object recognition – and indeed the visual cortex – is that of understanding quantitatively the trade-off between invariance and selectivity, and how invariance and discrimination properties contribute towards providing an improved representation useful for learning from data. In this work we provide a general group-theoretic framework for characterizing and understanding invariance in a family of hierarchical models. We show that by taking an algebraic perspective, one can provide a concise set of conditions which must be met to establish invariance, as well as a constructive prescription for meeting those conditions. Analyses in specific cases of particular relevance to computer vision and text processing are given, yielding insight into how and when invariance can be achieved. We find that the minimal sets of transformations intrinsic to the hierarchical model needed to support a particular invariance can be clearly described, thereby encouraging efficient computational implementations.

Subject Area: Theory

M58 Fast Learning from Non-i.i.d. Observations

 INGO STEINWART
 ingo@lanl.gov

 Los Alamos National Laboratory
 ANDREAS CHRISTMANN

 Andreas.Christmann@uni-bayreuth.de

 University of Bayreuth

We prove an oracle inequality for generic regularized empirical risk minimization algorithms learning from α -mixing processes. To illustrate this oracle inequality, we use it to derive learning rates for some learning methods including least squares SVMs. Since the proof of the oracle inequality uses recent localization ideas developed for independent and identically distributed (i.i.d.) processes, it turns out that these learning rates are close to the optimal rates known in the i.i.d. case. **Subject Area:** Theory

Spotlight presentation, Monday, 6:30.

M59 Periodic Step Size Adaptation for Single Pass On-line Learning

Chun-Nan Hsu	chunnan@iis.sinica.edu.tw
Yu-Ming Chang	porter@iis.sinica.edu.tw
Hanshen Huang	hanshen@iis.sinica.edu.tw
IIS, Academia Sinica	
Yuh-Jye Lee	yuh-jye@mail.ntust.edu.tw
Department of CSIE, National	Taiwan University of Science and Technology

It has been established that the second-order stochastic gradient descent (2SGD) method can potentially achieve generalization performance as well as empirical optimum in a single pass (i.e., epoch) through the training examples. However, 2SGD requires computing the inverse of the Hessian matrix of the loss function, which is prohibitively expensive. This paper presents Periodic Step-size Adaptation (PSA), which approximates the Jacobian matrix of the mapping function and explores a linear relation between the Jacobian and Hessian to approximate the Hessian periodically and achieve near-optimal results in experiments on a wide variety of models and tasks. **Subject Area:** Theory

M60 Nonparametric Bayesian Texture Learning and Synthesis

LONG ZHU leozhu@csail.mit.edu Massachusetts Institute of Technology YUANAHAO CHEN yhchen4@gmail.com UCLA BILL FREEMAN billf@mit.edu ANTONIO TORRALBA torralba@csail.mit.edu Massachusetts Institute of Technology

We present a nonparametric Bayesian method for texture learning and synthesis. A texture image is represented by a 2D-Hidden Markov Model (2D-HMM) where the hidden states correspond to the cluster labeling of textons and the transition matrix encodes their spatial layout (the compatibility between adjacent textons). 2D-HMM is coupled with the Hierarchical Dirichlet process (HDP) which allows the number of textons and the complexity of transition matrix grow as the input texture becomes irregular. The HDP makes use of Dirichlet process prior which favors regular textures by penalizing the model complexity. This framework (HDP-2D-HMM) learns the texton vocabulary and their spatial layout jointly and automatically. The HDP-2D-HMM results in a compact representation of textures which allows fast texture synthesis with comparable rendering quality over the state-of-the-art image-based rendering methods. We also show that HDP-2D-HMM can be applied to perform image segmentation and synthesis.

Subject Area: Vision

M61 The "tree-dependent components" of natural scenes are edge filters

DANIEL ZORAN	daniez@cs.huji.ac.il
Hebrew University of Jerusalem	
YAIR WEISS	yweiss@cs.huji.ac.il
Hebrew University	

We propose a new model for natural image statistics. Instead of minimizing dependency between components of natural images, we maximize a simple form of dependency in the form of tree-dependency. By learning filters and tree structures which are best suited for natural images we observe that the resulting filters are edge filters, similar to the famous ICA on natural images results. Calculating the likelihood of the model requires estimating the squared output of pairs of filters connected in the tree. We observe that after learning, these pairs of filters are predominantly of similar orientations but different phases, so their joint energy resembles models of complex cells. **Subject Area:** Vision

M62 A Biologically Plausible Model for Rapid Natural Scene Identification

Sennay Ghebreab	s.ghebreab@uva.nl
Steven Scholte	h.s.scholte@uva.nl
VICTOR LAMME	v.a.f.lamme@uva.nl
Arnold Smeulders	a.w.m.smeulders@uva.nl
University of Amsterdam	

Contrast statistics of the majority of natural images conform to a Weibull distribution. This property of natural images may facilitate efficient and very rapid extraction of a scene's visual gist. Here we investigate whether a neural response model based on the Weibull contrast distribution captures visual information that humans use to rapidly identify natural scenes. In a learning phase, we measure EEG activity of 32 subjects viewing brief flashes of 800 natural scenes. From these neural measurements and the contrast statistics of the natural image stimuli, we derive an across subject Weibull response model. We use this model to predict the responses to a large set of new scenes and estimate which scene the subject viewed by finding the best match between the model predictions and the observed EEG responses. In almost 90 percent of the cases our model accurately predicts the observed scene. Moreover, in most failed cases, the scene mistaken for the observed scene is visually similar to the observed scene itself. These results suggest that Weibull contrast statistics of natural images contain a considerable amount of scene gist information to warrant rapid identification of natural images. Subject Area: Vision

M63 Learning models of object structure

JOSEPH SCHLECHT Kobus Barnard University of Arizona schlecht@cs.arizona.edu kobus@cs.arizona.edu

We present an approach for learning stochastic geometric models of object categories from single view images. We focus here on models expressible as a spatially contiguous assemblage of blocks. Model topologies are learned across groups of images, and one or more such topologies is linked to an object category (e.g. chairs). Fitting learned topologies to an image can be used to identify the object class, as well as detail its geometry. The latter goes beyond labeling objects, as it provides the geometric structure We learn the models using joint statistical inference over of particular instances. structure parameters, camera parameters, and instance parameters. These produce an image likelihood through a statistical imaging model. We use trans-dimensional sampling to explore topology hypotheses, and alternate between Metropolis-Hastings and stochastic dynamics to explore instance parameters. Experiments on images of furniture objects such as tables and chairs suggest that this is an effective approach for learning models that encode simple representations of category geometry and the statistics thereof, and support inferring both category and geometry on held out single view images.

Subject Area: Vision

M64 Dirichlet-Bernoulli Alignment: A Generative Model for Multi-Class Multi-Label Multi-Instance Corpora

Shuang-Hong Yang shy@gatech.edu Hongyuan Zha Georgia Tech BAO-GANG HU hubg@nlpr.ia.ac.cn Chinese Academy of Sciences

zha@cc.gatech.edu

86

We propose Dirichlet-Bernoulli Alignment (DBA), a generative model for corpora in which each pattern (e.g., a document) contains a set of instances (e.g., paragraphs in the document) and belongs to multiple classes. By casting predefined classes as latent Dirichlet variables (i.e., instance level labels), and modeling the multi-label of each pattern as Bernoulli variables conditioned on the weighted empirical average of topic assignments, DBA automatically aligns the latent topics discovered from data to human-defined classes. DBA is useful for both pattern classification and instance disambiguation, which are tested on text classification and named entity disambiguation for web search queries respectively.

Subject Area: Supervised Learning

M65 Toward Provably Correct Feature Selection in Arbitrary Domains

DIMITRIS MARGARITIS Iowa State University dmarg@cs.iastate.edu

In this paper we address the problem of provably correct feature selection in arbitrary domains. An optimal solution to the problem is a Markov boundary, which is a minimal set of features that make the probability distribution of a target variable conditionally invariant to the state of all other features in the domain. While numerous algorithms for this problem have been proposed, their theoretical correctness and practical behavior under arbitrary probability distributions is unclear. We address this by introducing the Markov Boundary Theorem that precisely characterizes the properties of an ideal Markov boundary, and use it to develop algorithms that learn a more general boundary that can capture complex interactions that only appear when the values of multiple features are considered together. We introduce two algorithms: an exact, provably correct one as well a more practical randomized anytime version, and show that they perform well on artificial as well as benchmark and real-world data sets. Throughout the paper we make minimal assumptions that consist of only a general set of axioms that hold for every probability distribution, which gives these algorithms universal applicability.

Subject Area: Supervised Learning

M66 AUC optimization and the two-sample problem

Stéphan Clémençon	stephan.clemencon@telecom-paristech.fr
Telecom ParisTech	
NICOLAS VAYATIS	nicolas.vayatis@cmla.ens-cachan.fr
ENS Cachan	
Marine Depecker	marine.depecker@telecom-paristech.fr
Telecom ParisTech	

The purpose of the paper is to explore the connection between multivariate homogeneity tests and AUC optimization. The latter problem has recently received much attention in the statistical learning literature. From the elementary observation that, in the two-sample problem setup, the null assumption corresponds to the situation where the area under the optimal ROC curve is equal to 1/2, we propose a two-stage testing method based on data splitting. A nearly optimal scoring function in the AUC sense is first learnt from one of the two half-samples. Data from the remaining half-sample are then projected onto the real line and eventually ranked according to the scoring function computed at the first stage. The last step amounts to performing a standard Mann-Whitney Wilcoxon test in the one-dimensional framework. We show that the learning step of the procedure does not affect the consistency of the test as well as its properties in terms of power, provided the ranking produced is accurate enough in the AUC sense. The results of a numerical experiment are eventually displayed in order to show the efficiency of the method. **Subject Area:** Supervised Learning

M67 Learning to Rank by Optimizing NDCG Measure

HAMED VALIZADEGAN	valizade@msu.edu
Rong Jin	rongjin@cse.msu.edu
Michigan State University	
Ruofei Zhang	rzhang@yahoo-inc.com
Jianchang Mao	jmao@yahoo-inc.com
Yahoo!	

Learning to rank is a relatively new field of study, aiming to learn a ranking function from a set of training data with relevancy labels. The ranking algorithms are often evaluated using Information Retrieval measures, such as Normalized Discounted Cumulative Gain [1] and Mean Average Precision [2]. Until recently, most learning to rank algorithms were not using a loss function related to the above mentioned evaluation measures. The main difficulty in direct optimization of these measures is that they depend on the ranks of documents, not the numerical values output by the ranking We propose a probabilistic framework that addresses this challenge by function. optimizing the expectation of NDCG over all the possible permutations of documents. A relaxation strategy is used to approximate the average of NDCG over the space of permutation, and a bound optimization approach is proposed to make the computation efficient. Extensive experiments show that the proposed algorithm outperforms state-of-the-art ranking algorithms on several benchmark data sets. Subject Area: Supervised Learning

Spotlight presentation, Monday, 6:30.

M68 On Learning Rotations

RAMAN ARORA rmnarora@gmail.com Univ. of Wisconsin-Madison

An algorithm is presented for online learning of rotations. The proposed algorithm involves matrix exponentiated gradient updates and is motivated by the Von Neumann divergence. The additive updates are skew-symmetric matrices with trace zero which comprise the Lie algebra of the rotation group. The orthogonality and unit determinant of the matrix parameter are preserved using matrix logarithms and exponentials and the algorithm lends itself to interesting interpretations in terms of the computational topology of the compact Lie groups. The stability and the computational complexity of the algorithm are discussed. **Subject Area:** Supervised Learning

Spotlight presentation, Monday, 6:30.

M69 Zero-shot Learning with Semantic Output Codes

Mark Palatucci	mpalatuc@cs.cmu.edu
Carnegie Mellon University	
Dean Pomerleau	dean.a.pomerleau@intel.com
Intel Research	
Geoffrey Hinton	hinton@cs.toronto.edu
University of Toronto	
Tom Mitchell	tom.mitchell@cs.cmu.edu
Carnegie Mellon University	

We consider the problem of zero-shot learning, where the goal is to learn a classifier $f: X \to Y$ that must predict novel values of Y that were omitted from the training set. To achieve this, we define the notion of a semantic output code classifier (SOC) which utilizes a knowledge base of semantic properties of Y to extrapolate to novel classes. We provide a formalism for this type of classifier and study its theoretical properties in a PAC framework, showing conditions under which the classifier can accurately predict novel classes. As a case study, we build a SOC classifier for a neural decoding task and show that it can often predict words that people are thinking about from functional magnetic resonance images (fMRI) of their neural activity, even without training examples for those words.

Subject Area: Supervised Learning

M70 Ranking Measures and Loss Functions in Learning to Rank

Wei Chen	chenwei@amss.ac.cn
Chinese Academy of Sciences	
Tie-Yan Liu	tyliu@microsoft.com
Microsoft Research Asia	
Yanyan Lan	lanyanyan@amss.ac.cn
Zhi-Ming Ma	mazm@amt.ac.cn
Chinese Academy of Sciences	
Hang Li	hangli@microsoft.com
Microsoft Research Asia	

Learning to rank has become an important research topic in machine learning. While most learning-to-rank methods learn the ranking function by minimizing the loss functions, it is the ranking measures (such as NDCG and MAP) that are used to evaluate the performance of the learned ranking function. In this work, we reveal the relationship between ranking measures and loss functions in learning-to-rank methods, such as Ranking SVM, RankBoost, RankNet, and ListMLE. We show that these loss functions are upper bounds of the measure-based ranking errors. As a result, the minimization of these loss functions will lead to the maximization of the ranking measures. The key to obtaining this result is to model ranking as a sequence of classification tasks, and define a so-called essential loss as the weighted sum of the classification errors of individual tasks in the sequence. We have proved that the essential loss is both an upper bound of the measure-based ranking errors, and a lower bound of the loss functions in the aforementioned methods. Our proof technique also suggests a way to modify existing loss functions to make them tighter bounds of the measure-based ranking errors. Experimental results on benchmark datasets show that the modifications can lead to better ranking performance, demonstrating the correctness of our analysis. Subject Area: Supervised Learning

M71 Nash Equilibria of Static Prediction Games

MICHAEL BRÜCKNER TOBIAS SCHEFFER University of Potsdam mibrueck@cs.uni-potsdam.de scheffer@cs.uni-potsdam.de

The standard assumption of identically distributed training and test data can be violated when an adversary can exercise some control over the generation of the test data. In a prediction game, a learner produces a predictive model while an adversary may alter the distribution of input data. We study single-shot prediction games in which the cost functions of learner and adversary are not necessarily antagonistic. We identify conditions under which the prediction game has a unique Nash equilibrium, and derive algorithms that will find the equilibrial prediction models. In a case study, we explore properties of Nash-equilibrial prediction models for email spam filtering empirically. **Subject Area:** Supervised Learning

M72 Sparsistent Learning of Varying-coefficient Models with Structural Changes

Mladen Kolar	mladenk@cs.cmu.edu
Le Song	lesong@cs.cmu.edu
Eric Xing	epxing@cs.cmu.edu
Carnegie Mellon University	

To estimate the changing structure of a varying-coefficient varying-structure (VCVS) model remains an important and open problem in dynamic system modelling, which includes learning trajectories of stock prices, or uncovering the topology of an evolving gene network. In this paper, we investigate sparsistent learning of a sub-family of this model — piecewise constant VCVS models. We analyze two main issues in this problem: inferring time points where structural changes occur and estimating model structure (i.e., model selection) on each of the constant segments. We propose a two-stage adaptive procedure, which first identifies jump points of structural changes and then identifies relevant covariates to a response on each of the segments. We provide an asymptotic analysis of the procedure, showing that with the increasing sample size, number of structural changes, and number of variables, the true model can be consistently selected. We demonstrate the performance of the method on synthetic data and apply it to the brain computer interface dataset. We also consider how this applies to structure estimation of time-varying probabilistic graphical models.

Spotlight presentation, Monday, 6:30.

M73 Multiple Incremental Decremental Learning of Support Vector Machines

MASAYUKI KARASUYAMA krsym@ics.nitech.ac.jp ICHIRO TAKEUCHI takeuchi.ichiro@nitech.ac.jp Nagoya Institute of Technology We propose a multiple incremental decremental algorithm of Support Vector Machine (SVM). Conventional single cremental decremental SVM can update the trained model efficiently when single data point is added to or removed from the training set. When we add and/or remove multiple data points, this algorithm is time-consuming because we need to repeatedly apply it to each data point. The roposed algorithm is computationally more efficient when multiple data points are added and/or removed simultaneously. The single incremental decremental algorithm is built on an optimization technique called parametric programming. We extend the idea and introduce multi-parametric programming for developing the proposed algorithm. Experimental results on synthetic and real data sets indicate that the proposed algorithm can significantly reduce the computational cost of multiple incremental decremental operation. Our approach is especially useful for online SVM learning in which we need to remove old data points and add new data points in a short amount of time. Subject Area: Supervised Learning

M74 On the Algorithmics and Applications of a Mixed-norm based Kernel Learning Formulation

saketh@cse.iitb.ac.in
gdinesh@csa.iisc.ernet.in
<pre>sraman@csa.iisc.ernet.in</pre>
chiru@csa.iisc.ernet.in
lore
abental@ie.technion.ac.il
krr@ee.iisc.ernet.in
lore

Motivated from real world problems, like object categorization, we study a particular mixed-norm regularization for Multiple Kernel Learning (MKL). It is assumed that the given set of kernels are grouped into distinct components where each component is crucial for the learning task at hand. The formulation hence employs l_{∞} regularization for promoting combinations at the component level and l_1 regularization for promoting sparsity among kernels in each component. While previous attempts have formulated this as a non-convex problem, the formulation given here is an instance of non-smooth convex optimization problem which admits an efficient Mirror-Descent (MD) based procedure. The MD procedure optimizes over product of simplexes, which is not a well-studied case in literature. Results on real-world datasets show that the new MKL formulation is well-suited for object categorization tasks and that the MD based algorithm outperforms state-of-the-art MKL solvers like simpleMKL in terms of computational effort.

Subject Area: Supervised Learning

MONDAY, POSTER SESSION

M75 Graph-based Consensus Maximization among Multiple Supervised and Unsupervised Models

Jing Gao	jinggao3@illinois.edu
Feng Liang	liangf@illinois.edu
UIUC	
Wei Fan	weifan@us.ibm.com
IBM TJ Waston Research	
Yizhou Sun	sun22@illinois.edu
Jiawei Han	hanj@cs.uiuc.edu
UIUC	

Little work has been done to directly combine the outputs of multiple supervised and unsupervised models. However, it can increase the accuracy and applicability of ensemble methods. First, we can boost the diversity of classification ensemble by incorporating multiple clustering outputs, each of which provides grouping constraints for the joint label predictions of a set of related objects. Secondly, ensemble of supervised models is limited in applications which have no access to raw data but to the meta-level model outputs. In this paper, we aim at calculating a consolidated classification solution for a set of objects by maximizing the consensus among both supervised predictions and unsupervised grouping constraints. We seek a global optimal label assignment for the target objects, which is different from the result of traditional majority voting and model combination approaches. We cast the problem into an optimization problem on a bipartite graph, where the objective function favors smoothness in the conditional probability estimates over the graph, as well as penalizes deviation from initial labeling of supervised models. We solve the problem through iterative propagation of conditional probability estimates among neighboring nodes, and interpret the method as conducting a constrained embedding in a transformed space, as well as a ranking on the graph. Experimental results on three real applications demonstrate the benefits of the proposed method over existing alternatives. Subject Area: Supervised Learning

M76 Positive Semidefinite Metric Learning with Boosting

Chunhua Shen	chhshen@gmail.com
Junae Kim	junae.kim@nicta.com.au
NICTA	
Lei Wang	lei.wang@anu.edu.au
The Australian National University	У
Anton van den Hengel	anton.vandenhengel@adelaide.edu.au
University of Adelaide	

The learning of appropriate distance metrics is a critical problem in classification. In this work, we propose a boosting-based technique, termed BoostMetric, for learning a Mahalanobis distance metric. One of the primary difficulties in learning such a metric is to ensure that the Mahalanobis matrix remains positive semidefinite. Semidefinite programming is sometimes used to enforce this constraint, but does not scale well. BoostMetric is instead based on a key observation that any positive semidefinite matrix can be decomposed into a linear positive combination of trace-one rank-one matrices. BoostMetric thus uses rank-one positive semidefinite matrices as weak learners within an efficient and scalable boosting-based learning process. The resulting method is easy to implement, does not require tuning, and can accommodate various types of constraints.

Experiments on various datasets show that the proposed algorithm compares favorably to those state-of-the-art methods in terms of classification accuracy and running time. **Subject Area:** Supervised Learning

M77 Free energy score space

Alessandro Perina	alessandro.perina@gmail.com
Marco Cristani	marco.cristani@univr.it
Umberto Castellani	umberto.castellani@univr.it
Vittorio Murino	vittorio.murino@univr.it
University of Verona	
Nebojsa Jojic	jojic@microsoft.com
Microsoft Research	

Score functions induced by generative models extract fixed-dimension feature vectors from different-length data observations by subsuming the process of data generation, projecting them in highly informative spaces called score spaces. In this way, standard discriminative classifiers are proved to achieve higher performances than a solely generative or discriminative approach. In this paper, we present a novel score space that exploits the free energy associated to a generative model through a score function. This function aims at capturing both the uncertainty of the model learning and "local" compliance of data observations with respect to the generative process. Theoretical justifications and convincing comparative classification results on various generative models prove the goodness of the proposed strategy. **Subject Area:** Supervised Learning

M78 Conditional Neural Fields

Jian Peng	pengjian@tti-c.org
Liefeng Bo	blf0218@tti-c.org
Jinbo Xu	j3xu@tti-c.org
TTI-C	

Conditional random fields (CRF) are quite successful on sequence labeling tasks such as natural language processing and biological sequence analysis. CRF models use linear potential functions to represent the relationship between input features and outputs. However, in many real-world applications such as protein structure prediction and handwriting recognition, the relationship between input features and outputs is highly complex and nonlinear, which cannot be accurately modeled by a linear function. To model the nonlinear relationship between input features and outputs we propose Conditional Neural Fields (CNF), a new conditional probabilistic graphical model for sequence labeling. Our CNF model extends CRF by adding one (or possibly several) middle layer between input features and outputs. The middle layer consists of a number of hidden parameterized gates, each acting as a local neural network node or feature extractor to capture the nonlinear relationship between input features and outputs. Therefore, conceptually this CNF model is much more expressive than the linear CRF model. To better control the complexity of the CNF model, we also present a hyperparameter optimization procedure within the evidence framework. Experiments on two widely-used benchmarks indicate that this CNF model performs significantly better than a number of popular methods. In particular, our CNF model is the best among about ten machine learning methods for protein secondary tructure prediction and also among a few of the best methods for handwriting recognition. **Subject Area:** Supervised Learning

M79 Distribution-Calibrated Hierarchical Classification

OFER DEKEL Microsoft oferd@microsoft.com

While many advances have already been made on the topic of hierarchical classi- fication learning, we take a step back and examine how a hierarchical classifica- tion problem should be formally defined. We pay particular attention to the fact that many arbitrary decisions go into the design of the the label taxonomy that is provided with the training data, and that this taxonomy is often unbalanced. We correct this problem by using the data distribution to calibrate the hierarchical classification loss function. This distributionbased correction must be done with care, to avoid introducing unmanagable statical dependencies into the learning problem. This leads us off the beaten path of binomial-type estimation and into the uncharted waters of geometric-type estimation. We present a new calibrated definition of statistical risk for hierarchical classification, an unbiased geometric estimator for this risk, and a new algorithmic reduction from hierarchical classifi- cation to cost-sensitive classification.

Subject Area: Supervised Learning

M80 Learning with Compressible Priors

Volkan Cevher Rice University volkan@rice.edu

We describe probability distributions, dubbed compressible priors, whose independent and identically distributed (iid) realizations result in compressible signals. A signal is compressible when sorted magnitudes of its coefficients exhibit a power-law decay so that the signal can be well-approximated by a sparse signal. Since compressible signals live close to sparse signals, their intrinsic information can be stably embedded via simple nonadaptive linear projections into a much lower dimensional space whose dimension grows logarithmically with the ambient signal dimension. By using order statistics, we show that N-sample iid realizations of generalized Pareto, Student's t, log-normal, Frechet, and loglogistic distributions are compressible, i.e., they have a constant expected decay rate, which is independent of N. In contrast, we show that generalized Gaussian distribution with shape parameter q is compressible only in restricted cases since the expected decay rate of its N-sample iid realizations decreases with N as $1/[q \log(N/q)]$. We use compressible priors as a scaffold to build new iterative sparse signal recovery algorithms based on Bayesian inference arguments. We show how tuning of these algorithms explicitly depends on the parameters of the compressible prior of the signal, and how to learn the parameters of the signal's compressible prior on the fly during recovery. Subject Area: Supervised Learning

M81 A Fast, Consistent Kernel Two-Sample Test

ARTHUR GRETTONarthur@tuebingen.mpg.deCarnegie Mellon University and Max Planck InstituteKENJI FUKUMIZUfukumizu@ism.ac.jpThe Institute of Statistical MathematicsZAID HARCHAOUIcarnegie Mellon UniversityBHARATH SRIPERUMBUDURbharathsv@ucsd.eduUC San Diego

A kernel embedding of probability distributions into reproducing kernel Hilbert spaces (RKHS) has recently been proposed, which allows the comparison of two probability measures P and Q based on the distance between their respective embeddings: for a sufficiently rich RKHS, this distance is zero if and only if P and Q coincide. In using this distance as a statistic for a test of whether two samples are from different distributions, a major difficulty arises in computing the significance threshold, since the empirical statistic has as its null distribution (where P=Q) an infinite weighted sum of χ^2 random variables. The main result of the present work is a novel, consistent estimate of this null distribution, computed from the eigenspectrum of the Gram matrix on the aggregate sample from P and Q. This estimate may be computed faster than a previous consistent estimate based on fitting a parametric family with the low order moments of the test statistic: unlike the present work, this heuristic has no guarantee of being accurate or consistent. We verify the performance of our null distribution estimate on both an artificial example and on high dimensional multivariate data.

Subject Area: Supervised Learning

Spotlight presentation, Monday, 6:30.

M82 Improving Existing Fault Recovery Policies

GUY SHANI CHRISTOPHER MEEK Microsoft Research shanigu@cs.bgu.ac.il
cmeek@microsoft.com

Automated recovery from failures is a key component in the management of large data centers. Such systems typically employ a hand-made controller created by an expert. While such controllers capture many important aspects of the recovery process, they are often not systematically optimized to reduce costs such as server downtime. In this paper we explain how to use data gathered from the interactions of the hand-made controller with the system, to create an optimized controller. We suggest learning an indefinite horizon Partially Observable Markov Decision Process, a model for decision making under uncertainty, and solve it using a point-based algorithm. We describe the complete process, starting with data gathering, model learning, model checking procedures, and computing a policy. While our paper focuses on a specific domain, our method is applicable to other systems that use a hand-coded, imperfect controllers.

Subject Area: Control and Reinforcement Learning

Henning Sprekeler	henning.sprekeler@epfl.ch
Guillaume Hennequin	guillaume.hennequin@epfl.ch
Wulfram Gerstner	wulfram.gerstner@epfl.ch
Ecole Polytechnique Federale de	e Lausanne

Although it is widely believed that reinforcement learning is a suitable tool for describing behavioral learning, the mechanisms by which it can be implemented in networks of spiking neurons are not fully understood. Here, we show that different learning rules emerge from a policy gradient approach depending on which features of the spike trains are assumed to influence the reward signals, i.e., depending on which neural code is in effect. We use the framework of Williams (1992) to derive learning rules for arbitrary neural codes. For illustration, we present policy-gradient rules for three different example codes - a spike count code, a spike timing code and the most general "full spike train" code - and test them on simple model problems. In addition to classical synaptic learning, we derive learning rules for intrinsic parameters that control the excitability of the neuron. The spike count learning rule has structural similarities with established Bienenstock-Cooper-Munro rules. If the distribution of the relevant spike train features belongs to the natural exponential family, the learning rules have a characteristic shape that raises interesting prediction problems.

Subject Area: Control and Reinforcement Learning

M84 A Generalized Natural Actor-Critic Algorithm

TETSURO MORIMURA	tetsuro@jp.ibm.com
IBM Research - Tokyo	
Eiji Uchibe	uchibe@oist.jp
Junichiro Yoshimoto	jun-y@oist.jp
Kenji Doya	doya@oist.jp
Okinawa Institute of Science and	l Technology

Policy gradient Reinforcement Learning (RL) algorithms have received much attention in seeking stochastic policies that maximize the average rewards. In addition, extensions based on the concept of the Natural Gradient (NG) show promising learning efficiency because these regard metrics for the task. Though there are two candidate metrics, Kakade's Fisher Information Matrix (FIM) and Morimura's FIM, all RL algorithms with NG have followed the Kakade's approach. In this paper, we describe a generalized Natural Gradient (gNG) by linearly interpolating the two FIMs and propose an efficient implementation for the gNG learning based on a theory of the estimating function, generalized Natural Actor-Critic (gNAC). The gNAC algorithm involves a near optimal auxiliary function to reduce the variance of the gNG estimates. Interestingly, the gNAC can be regarded as a natural extension of the current state-of-the-art NAC algorithm, as long as the interpolating parameter is appropriately selected. Numerical experiments showed that the proposed gNAC algorithm can estimate gNG efficiently and outperformed the NAC algorithm.

Subject Area: Control and Reinforcement Learning

M85 Convergent Temporal-Difference Learning with Arbitrary Smooth Function Approximation

Hamid Maei	maei@cs.ualberta.ca
Csaba Szepesvari	szepesva@cs.ualberta.ca
University of Alberta	
Shalabh Bhatnagar	shalabh@cs.ualberta.ca
Indian Institute of Science,	Bangalore, India
Doina Precup	dprecup@cs.mcgill.ca
McGill University	
David Silver	silver@cs.ualberta.ca
RICH SUTTON	<pre>sutton@cs.ualberta.ca</pre>
University of Alberta	

We introduce the first temporal-difference learning algorithms that converge with smooth value function approximators, such as neural networks. Conventional temporal-difference (TD) methods, such as $TD(\lambda)$, Q-learning and Sarsa have been used successfully with function approximation in many applications. However, it is well known that off-policy sampling, as well as nonlinear function approximation, can cause these algorithms to become unstable (i.e., the parameters of the approximator may diverge). Sutton et al (2009a,b) solved the problem of off-policy learning with linear TD algorithms by introducing a new objective function, related to the Bellman-error, and algorithms that perform stochastic gradient-descent on this function. In this paper, we generalize their work to nonlinear function approximation. We present a Bellman error objective function and two gradient-descent TD algorithms that optimize it. We prove the asymptotic almost-sure convergence of both algorithms for any finite Markov decision process and any smooth value function approximator, under usual stochastic approximation conditions. The computational complexity per iteration scales linearly with the number of parameters of the approximator. The algorithms are incremental and are guaranteed to converge to locally optimal solutions.

Subject Area: Control and Reinforcement Learning

Spotlight presentation, Monday, 6:30.

M86 Solving Stochastic Games

LIAM MAC DERMED	liam@cc.gatech.edu
Charles Isbell	isbell@cc.gatech.edu
Georgia Inst. of Technology	

Solving multi-agent reinforcement learning problems has proven difficult because of the lack of tractable algorithms. We provide the first approximation algorithm which solves stochastic games to within ϵ relative error of the optimal game-theoretic solution, in time polynomial in $1/\epsilon$. Our algorithm extends Murray's and Gordon's (2007) modified Bellman equation which determines the *set* of all possible achievable utilities; this provides us a truly general framework for multi-agent learning. Further, we empirically validate our algorithm and find the computational cost to be orders of magnitude less than what the theory predicts.

Subject Area: Control and Reinforcement Learning

M87 White Functionals for Anomaly Detection in Dynamical Systems

Marco Cuturi	mcuturi@princeton.edu
Princeton University	
Jean-Philippe Vert	Jean-Philippe.Vert@Mines-ParisTech.fr
ParisTech	
Alexandre D'Aspremont	aspremon@princeton.edu
Princeton University	

We propose new methodologies to detect anomalies in discrete-time processes taking values in a set. The method is based on the inference of functionals whose evaluations on successive states visited by the process have low autocorrelations. Deviations from this behavior are used to flag anomalies. The candidate functionals are estimated in a subset of a reproducing kernel Hilbert space associated with the set where the process takes values. We provide experimental results which show that these techniques compare favorably with other algorithms.

Subject Area: Supervised Learning

Tuesday, December 8th

Oral Session (8:30–10:10am): "Information Theory and Estimation"

8:30am Invited Talk: Relative Entropy

SERGIO VERDU verdu@princeton.edu Princeton University

An overview of relative entropy (aka Kulback-Leibler divergence, etc.) and its multiple appearances in information theory, probability, and statistics, including recent results by the speaker.

9:30am Multi-Label Prediction via Compressed Sensing

Daniel Hsu	djhsu@cs.ucsd.edu
University of California	
Sham Kakade	sham@tti-c.org
Toyota Technological Institute	
John Langford	jl@hunch.net
Yahoo! Research	
Tong Zhang	tzhang@stat.rutgers.edu
Rutgers University	

We consider multi-label prediction problems with large output spaces under the assumption of output sparsity – that the target (label) vectors have small support. We develop a general theory for a variant of the popular error correcting output code scheme, using ideas from compressed sensing for exploiting this sparsity. The method can be regarded as a simple reduction from multi-label regression problems to binary regression problems. We show that the number of subproblems need only be logarithmic in the total number of possible labels, making this approach radically more efficient than others. We also state and prove robustness guarantees for this method in the form of regret transform bounds (in general), and also provide a more detailed analysis for the linear prediction setting.

Subject Area: Theory

9:50am A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers

SAHAND NEGAHBANsahand_n@eecs.berkeley.eduUC, Berkeleypradeepr@stat.berkeley.eduPRADEEP RAVIKUMARpradeepr@stat.berkeley.eduMARTIN WAINWRIGHTwainwrig@eecs.berkeley.eduBIN YUbinyu@stat.berkeley.eduUniversity of Californiabinyu@stat.berkeley.edu

The estimation of high-dimensional parametric models requires imposing some structure on the models, for instance that they be sparse, or that matrix structured parameters have low rank. A general approach for such structured parametric model estimation is to use regularized M-estimation procedures, which regularize a loss function that measures goodness of fit of the parameters to the data with some regularization function that encourages the assumed structure. In this paper, we aim to provide a unified analysis of such regularized M-estimation procedures. In particular, we report the convergence rates of such estimators in any metric norm. Using just our main theorem, we are able to rederive some of the many existing results, but also obtain a wide range of novel convergence rates results. Our analysis also identifies key properties of loss and regularization functions, such as restricted strong convexity, and decomposability, that ensure the corresponding regularized M-estimators have good convergence rates. **Subject Area:** Theory

Oral Session (10:40am–12:00pm): "Images and Codes"

10:40am An Additive Latent Feature Model for Transparent Object Recognition

Mario Fritz	mfritz@eecs.berkeley.edu
UC Berkeley EECS and ICSI	
Michael Black	black@cs.brown.edu
Brown University	
Gary Bradski	bradski@willowgarage.com
Willow Garage	
Sergey Karayev	sergeyk@eecs.berkeley.edu
UC Berkeley	
Trevor Darrell	<pre>trevor@eecs.berkeley.edu</pre>
UC Berkeley EECS and ICSI	

Existing methods for recognition of object instances and categories based on quantized local features can perform poorly when local features exist on transparent surfaces, such as glass or plastic objects. There are characteristic patterns to the local appearance of transparent objects, but they may not be well captured by distances to individual examples or by a local pattern codebook obtained by vector quantization. The appearance of a transparent patch is determined in part by the refraction of a background pattern through a transparent medium: the energy from the background usually dominates the patch appearance. We model transparent local patch appearance using an additive model of latent factors: background factors due to scene content, and factors which capture a local edge energy distribution characteristic of the refraction. We implement our method using a novel LDA-SIFT formulation which performs LDA prior to any vector quantization step; we discover latent topics which are characteristic of particular transparent patches and quantize the SIFT space into 'transparent visual words' according to the latent topic dimensions. No knowledge of the background scene is required at test time; we show examples recognizing transparent glasses in a domestic environment. **Subject Area:** Vision

11:00am Semi-Supervised Learning in Gigantic Image Collections

ROB FERGUS	fergus@cs.nyu.edu
New York University	
YAIR WEISS	yweiss@cs.huji.ac.il
Hebrew University	
Antonio Torralba	torralba@csail.mit.edu
Massachusetts Institute of Techno	logv

With the advent of the Internet it is now possible to collect hundreds of millions of images. These images come with varying degrees of label information. "Clean labels" can be manually obtained on a small fraction, "noisy labels" may be extracted automatically from surrounding text, while for most images there are no labels at all. Semi-supervised learning is a principled framework for combining these different label sources. However, it scales polynomially with the number of images, making it impractical for use on gigantic collections with hundreds of millions of images and thousands of classes. In this paper we show how to utilize recent results in machine learning to obtain highly efficient approximations for semi-supervised learning that are linear in the number of images. Specifically, we use the convergence of the eigenvectors of the normalized graph Laplacian to eigenfunctions of weighted Laplace-Beltrami operators. We combine this with a label sharing framework obtained from Wordnet to propagate label information to classes lacking manual annotations. Our algorithm enables us to apply semi-supervised learning to a database of 80 million images with 74 thousand classes.

11:20am Non-Parametric Bayesian Dictionary Learning for Sparse Image Representations

Mingyuan Zhou	mz1@ee.duke.edu
Duke ECE	
Haojun Chen	hc44@ee.duke.edu
Duke University	
John Paisley	jwp4@ee.duke.edu
Lu Ren	lr@ee.duke.edu
Duke ECE	
Guillermo Sapiro	guille@umn.edu
University of Minnesota	
LAWRENCE CARIN	lcarin@ee.duke.edu
Duke ECE	

Non-parametric Bayesian techniques are considered for learning dictionaries for sparse image representations, with applications in denoising, inpainting and compressive sensing (CS). The beta process is employed as a prior for learning the dictionary, and this nonparametric method naturally infers an appropriate dictionary size. The Dirichlet process and a probit stick-breaking process are also considered to exploit structure within an image. The proposed method can learn a sparse dictionary in situ; training images may be exploited if available, but they are not required. Further, the noise variance need not be known, and can be non-stationary. Another virtue of the proposed method is that sequential inference can be readily employed, thereby allowing scaling to large images. Several example results are presented, using both Gibbs and variational Bayesian inference, with comparisons to other state-of-the-art approaches.

Subject Area: Probabilistic Models and Methods

11:40am Locality-sensitive binary codes from shift-invariant kernels

MAXIM RAGINSKY m.raginsky@duke.edu Duke University SVETLANA LAZEBNIK lazebnik@cs.unc.edu University of North Carolina

This paper addresses the problem of designing binary codes for high-dimensional data such that vectors that are similar in the original space map to similar binary strings. We introduce a simple distribution-free encoding scheme based on random projections, such that the expected Hamming distance between the binary codes of two vectors is related to the value of a shift-invariant kernel (e.g., a Gaussian kernel) between the vectors. We present a full theoretical analysis of the convergence properties of the proposed scheme, and report favorable experimental performance as compared to a recent state-of-the-art method, spectral hashing.

Subject Area: Unsupervised and Semi-supervised Learning

Publication Models Debate (12:30pm-1:45pm)

12:30pm Debate on Future Publication Models for the NIPS Community

YOSHUA BENGIO University of Montreal bengioy@iro.umontreal.ca

yann@cs.nyu.edu
pereira@google.com
koller@cs.stanford.edu
saul@cs.ucsd.edu
С
tgd@cs.orst.edu

Leon Bottou	leon@bottou.org
NEC Laboratories	
Zoubin Ghahramani	zoubin@eng.cam.ac.uk
Cambridge University	
Terry Sejnowski	terry@salk.edu
Salk Institute	
Wolfgang Maass	maass@igi.tugraz.at
Universität Graz	

Our community is vibrant but many questions, complaints, and suggestions for change have been voiced about the way in which the science we produce is evaluated and disseminated. This lunchtime debate and discussion is an opportunity for us all to explore possible changes to the current publication model in our community. The panel members are senior pillars of NIPS with editorial experience. Below are some of the questions that they would like to discuss with you. We would like this discussion to continue beyond the 75 minutes of the Tuesday debate, maybe electronically.

Issues with the current model:

- The conference/journal model in many areas of computer science like machine learning is very different from what happens in other communities: why is it that way in CS? as discussed by Lance Fortnow (http://cacm.acm.org/magazines/2009/8/34492-viewpoint-time-for-computer-science-to-grow-up/fulltext), is that the model we want? are there alternatives?
- Is author-anonymity a barrier to faster availability of papers? (because for conferences with anonymous submission the paper could otherwise have been on a public archive from the time the work is done to when the final paper is published). What about the 'rich get richer' problem if we do not have author anonymity?
- Reviewers are often feeling overloaded and not always very motivated.
- Should we question the current practice of confidentiality and anonymity of the reviews? i.e. publishing the reviews might yield better reviews, even if anonymous.
- There is a 'reviewer crunch' in July when most of the community has to review so many papers quickly.
- The differences between the CS publication model and the neuroscience publication model is a problem for the multi-disciplinary nature of NIPS, yielding a loss of the computational neuroscience contributions over the years.

Proposals for changes:

- Do the new electronic tools, archives, collaborative filtering etc. provide the opportunity for new and better ways of doing things?
- With modern technologies publications delays could be zero, but evaluation delays remain substantial. How do we leverage the above to make scientific communication more efficient?
- How do we reward the reviewers to motivate them better, e.g. by allowing the reviews to be published and cited?
- Can we change the submission and evaluation process to improve the quality of the selection of papers and the reviewer crunch? Maybe journals and conferences can team-up in order to find better solutions?

• How do we reconcile and ideally disentangle the conflicting roles of efficient retrieval, score-keeping, and stimulation that conferences and journals play?

Your opinion matters, please come and express it, participate in the debate. Do not hesitate to bring your sandwich!

Location: Stanley Room (34th floor)

Oral Session (2:00–3:20pm): "Deep Learning and Network Models"

2:00pm Invited Talk: Deep Learning with Multiplicative Interactions

hinton@cs.toronto.edu

GEOFFREY HINTON University of Toronto

Deep networks can be learned efficiently from unlabeled data. The layers of representation are learned one at a time using a simple learning module that has only one layer of latent variables. The values of the latent variables of one module form the data for training the next module. The most commonly used modules are Restricted Boltzmann Machines or autoencoders with a sparsity penalty on the hidden activities. Although deep networks have been quite successful for tasks such as object recognition, information retrieval, and modeling motion capture data, the simple learning modules do not have multiplicative interactions which are very useful for some types of data. The talk will show how a thirdorder energy function can be factorized to yield a simple learning module that retains advantageous properties of a Restricted Boltzmann Machine such as very simple exact inference and a very simple learning rule based on pair-wise statistics. The new module contains multiplicative interactions that are useful for a variety of unsupervised learning tasks. Researchers at the University of Toronto have been using this type of module to extract oriented energy from image patches and dense flow fields from image sequences. The new module can also be used to allow the style of a motion to blend autoregressive models of motion capture data. Finally, the new module can be used to combine an eveposition with a feature-vector to allow a system that has a variable resolution retina to integrate information about shape over many fixations.

Geoffrey Hinton received his PhD in Artificial Intelligence from Edinburgh in 1978. He did postdoctoral work at the University of California San Diego and spent five years as a faculty member in the Computer Science Department at Carnegie-Mellon. He then became a fellow of the Canadian Institute for Advanced Research and moved to the Department of Computer Science at the University of Toronto. He spent three years from 1998 until 2001 setting up the Gatsby Computational Neuroscience Unit at University College London and then returned to the University of Toronto where he is the Raymond Reiter Distinguished Professor of Artificial Intelligence. He is a fellow of the Royal Society and an honorary foreign member of the American Academy of Arts. He has been awarded the Rumelhart prize and the Research Excellence award of the International Joint Conference on Artificial Intelligence. Geroffrey Hinton was one of the researchers who introduced the back-propagation algorithm that has been widely used for practical applications. His other main contributions to neural network research include Boltzmann machines, distributed representations, time-delay neural nets, mixtures of experts, Helmholtz machines, products of experts, and deep learning on unlabeled data by stacking simple learning modules.

3:00pm Discriminative Network Models of Schizophrenia

Guillermo Cecchi	gcecchi@us.ibm.com
Irina Rish	rish@us.ibm.com
IBM T.J. Watson Research Center	r
Benjamin Thyreau	benjamin.thyreau@gmail.com
Neurospin, CEA-INSERM	
Bertrand Thirion	bertrand.thirion@inria.fr
INRIA	
MARION PLAZE	marion.plaze@gmail.com
MADIE LAUDE DAILLEDE MADT	INOT ml poillone@pol op_hop_porig fr
AP-HP	inoi mi.paillereepsi.ap-nop-paris.ii
CATHERINE MARTELLI	martelli@cea.fr
Centre Hospitalier Paul Brousse	
Jean-Luc Martinot	jean-luc.martinot@cea.fr
INSERM - CEA - Univ. Paris Su	d
Jean-Baptiste Poline	jbpoline@cea.fr
Neurospin, CEA-INSERM	

Schizophrenia is a complex psychiatric disorder that has eluded a characterization in terms of local abnormalities of brain activity, and is hypothesized to affect the collective, "emergent" working of the brain. We propose a novel data-driven approach to capture emergent features using functional brain networks [Eguiluzet al] extracted from fMRI data, and demonstrate its advantage over traditional region-of-interest (ROI) and local, task-specific linear activation analyzes. Our results suggest that schizophrenia is indeed associated with disruption of global, emergent brain properties related to its functioning as a network, which cannot be explained by alteration of local activation patterns. Moreover, further exploitation of interactions by sparse Markov Random Field classifiers shows clear gain over linear methods, such as Gaussian Naive Bayes and SVM, allowing to reach 86% accuracy (over 50% baseline - random guess), which is quite remarkable given that it is based on a single fMRI experiment using a simple auditory task. **Subject Area:** Neuroscience

Spotlights (3:20–3:40pm)

- Kernels and learning curves for Gaussian process regression on random graphs
 PETER SOLLICH, Kings College London, MATTHEW URRY, King's College London, and CAMILLE COTI, INRIA Saclay-Ile de France.
 Subject Area: Probabilistic Models and Methods See abstract, page 110.
- Robust Value Function Approximation Using Bilinear Programming MAREK PETRIK, University of Massachusetts Amherst, and SHLOMO ZILBERSTEIN, University of Massachusetts at Amherst.
 Subject Area: Control and Reinforcement Learning See abstract, page 147.

- Analysis of SVM with Indefinite Kernels
 YIMING YING, COLIN CAMPBELL, University of Bristol, and MARK GIROLAMI, University
 of Glasgow.

 Subject Area: Supervised Learning
 See abstract, page 141.
- Boosting with Spatial Regularization ZHEN XIANG, Princeton University, YONGXIN XI, URI HASSON and PETER RAMADGE, Princeton University.
 Subject Area: Supervised Learning See abstract, page 142.
- Abstraction and Relational learning CHARLES KEMP and ALAN JERN, Carnegie Mellon University. Subject Area: Cognitive Science See abstract, page 122.
- Probabilistic Relational PCA WU-JUN LI, DIT-YAN YEUNG, HKUST, and ZHIHUA ZHANG, Zhejiang University. Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 111.
- Decoupling Sparsity and Smoothness in the Discrete Hierarchical Dirichlet Process CHONG WANG and DAVID BLEI, Princeton University. Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 112.
- A joint maximum-entropy model for binary neural population patterns and continuous signals
 SEBASTIAN GERWINN, PHILIPP BERENS and MATTHIAS BETHGE, MPI for Biological Cybernetics & University of Tübingen.
 Subject Area: Neuroscience See abstract, page 137.
- Efficient Bregman Range Search LAWRENCE CAYTON, Max Planck Institute for Biological Cybernetics. Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 112.
- Graph Zeta Function in the Bethe Free Energy and Loopy Belief Propagation YUSUKE WATANABE and KENJI FUKUMIZU, The Institute of Statistical Mathematics. Subject Area: Probabilistic Models and Methods See abstract, page 118.

- Online Learning of Assignments MATTHEW STREETER, Google, DANIEL GOLOVIN, Carnegie Mellon University, and ANDREAS KRAUSE, California Institute of Technology. Subject Area: Theory See abstract, page 127.
- Linear-time Algorithms for Pairwise Statistical Problems
 PARIKSHIT RAM, Georgia Institute of Technolog, DONGRYEOL LEE, Georgia Institute of
 Technology, WILLIAM MARCH, Georgia Tech, and ALEXANDER GRAY, Georgia Institute of
 Technology.

 Subject Area: Theory
 See abstract, page 129.

Oral Session (4:10–5:10pm): "Cognitive Science and EEG Analysis"

4:10pm Differential Use of Implicit Negative Evidence in Generative and Discriminative Language Learning

ANNE HSU showen@berkeley.edu University of California Berkeley THOMAS GRIFFITHS tom_griffiths@berkeley.edu University of California, Berkeley

A classic debate in cognitive science revolves around understanding how children learn complex linguistic rules, such as those governing restrictions on verb alternations, without negative evidence. Traditionally, formal learnability arguments have been used to claim that such learning is impossible without the aid of innate language-specific knowledge. However, recently, researchers have shown that statistical models are capable of learning complex rules from only positive evidence. These two kinds of learnability analyses differ in their assumptions about the role of the distribution from which linguistic input is generated. The former analyses assume that learners seek to identify grammatical sentences in a way that is robust to the distribution from which the sentences are generated, analogous to discriminative approaches in machine learning. The latter assume that learners are trying to estimate a generative model, with sentences being sampled from that model. We show that these two learning approaches differ in their use of implicit negative evidence – the absence of a sentence – when learning verb alternations, and demonstrate that human learners can produce results consistent with the predictions of both approaches, depending on the context in which the learning problem is presented.

Subject Area: Cognitive Science

4:30pm Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model

ED VULevul@mit.eduMICHAEL FRANKmcfrank@mit.eduMassachusetts Institute of TechnologyGEORGE ALVAREZGEORGE ALVAREZalvarez@wjh.harvard.eduDept. of Psychology, Harvard UniversityJOSHUA TENENBAUMJOSHUA TENENBAUMjbt@mit.eduMassachusetts Institute of Technology

Multiple object tracking is a task commonly used to investigate the architecture of human visual attention. Human participants show a distinctive pattern of successes and failures in tracking experiments that is often attributed to limits on an object system, a tracking module, or other specialized cognitive structures. Here we use a computational analysis of the task of object tracking to ask which human failures arise from cognitive limitations and which are consequences of inevitable perceptual uncertainty in the tracking task. We find that many human performance phenomena, measured through novel behavioral experiments, are naturally produced by the operation of our ideal observer model (a Rao-Blackwelized particle filter). The tradeoff between the speed and number of objects being tracked, however, can only arise from the allocation of a flexible cognitive resource, which can be formalized as either memory or attention. **Subject Area:** Cognitive Science

4:50pm Optimizing Multi-Class Spatio-Spectral Filters via Bayes Error Estimation for EEG Classification

WENMING ZHENG wenming_zheng@seu.edu.cn Southeast University, P.R. China ZHOUCHEN LIN zhoulin@microsoft.com Microsoft Research Asia, P.R. China

The method of common spatio-spectral patterns (CSSPs) is an extension of common spatial patterns (CSPs) by utilizing the technique of delay embedding to alleviate the adverse effects of noises and artifacts on the electroencephalogram (EEG) classification. Although the CSSPs method has shown to be more powerful than the CSPs method in the EEG classification, this method is only suitable for two-class EEG classification problems. In this paper, we generalize the two-class CSSPs method to multi-class cases. To this end, we first develop a novel theory of multi-class Bayes error estimation and then present the multi-class CSSPs (MCSSPs) method based on this Bayes error theoretical framework. By minimizing the estimated closed-form Bayes error, we obtain the optimal spatio-spectral filters of MCSSPs. To demonstrate the effectiveness of the proposed method, we conduct extensive experiments on the data set of BCI competition 2005. The experimental results show that our method significantly outperforms the previous multi-class CSPs (MCSPs) methods in the EEG classification. **Subject Area:** Neuroscience
Spotlights (5:10–5:30pm)

- Indian Buffet Processes with Power-law Behavior YEE WHYE TEH and DILAN GORUR, Gatsby Computational Neuroscience Unit, UCL. Subject Area: Probabilistic Models and Methods See abstract, page 119.
- Submodularity Cuts and Applications
 YOSHINOBU KAWAHARA, Osaka University, KIYOHITO NAGANO, Tokyo Institute of
 Technology, KOJI TSUDA, Advanced Industrial Science and Technology, and JEFF BILMES,
 University of Washington.

 Subject Area: Optimization
 See abstract, page 128.
- Statistical Models of Linear and Nonlinear Contextual Interactions in Early Visual Processing RUBEN COEN-CAGLI, Albert Einstein College of Medicine, PETER DAYAN, University College London, and ODELIA SCHWARTZ, Albert Einstein College of Medicine.
 Subject Area: Neuroscience See abstract, page 139.
- STDP enables spiking neurons to detect hidden causes of their inputs BERNHARD NESSLER, MICHAEL PFEIFFER and WOLFGANG MAASS, Graz University of Technology.

Subject Area: Neuroscience See abstract, page 139.

- 3D Object Recognition with Deep Belief Nets VINOD NAIR and GEOFFREY HINTON, University of Toronto. Subject Area: Supervised Learning See abstract, page 144.
- Maximum likelihood trajectories for continuous-time Markov chains THEODORE PERKINS, Ottawa Hospital Research Inst..
 Subject Area: Probabilistic Models and Methods See abstract, page 120.
- Time-Varying Dynamic Bayesian Networks LE SONG, MLADEN KOLAR and ERIC XING, Carnegie Mellon University. Subject Area: Applications See abstract, page 125.
- Construction of Nonparametric Bayesian Models from Parametric Bayes Equations PETER ORBANZ, University of Cambridge. Subject Area: Theory See abstract, page 131.

- Training Factor Graphs with Reinforcement Learning for Efficient MAP Inference MICHAEL WICK, KHASHAYAR ROHANIMANESH, SAMEER SINGH and ANDREW MCCALLUM, University of Massachusetts, Amherst. Subject Area: Supervised Learning See abstract, page 146.
- Riffled Independence for Ranked Data JONATHAN HUANG and CARLOS GUESTRIN, Carnegie Mellon University. Subject Area: Probabilistic Models and Methods See abstract, page 121.
- A Smoothed Approximate Linear Program
 VIJAY DESAI, Columbia University, VIVEK FARIAS, MIT Sloan, and CIAMAC MOALLEMI,
 Columbia University.
 Subject Area: Control and Reinforcement Learning
 See abstract, page 147.
- Randomized Pruning: Efficiently Calculating Expectations in Large Dynamic Programs
 ALEXANDRE BOUCHARD-CÔTÉ, UC Berkeley, SLAV PETROV, Google Research, and DAN KLEIN, University of California Berkeley.
 Subject Area: Applications See abstract, page 126.
- Subject independent EEG-based BCI decoding SIAMAC FAZLI, Technical University Berlin, CRISTIAN GROZEA, Fraunhofer Institute FIRST, MARTON DANOCZY, TU Berlin, BENJAMIN BLANKERTZ, Berlin Institute of Technology, FLORIN POPESCU, Fraunhofer FIRST, and KLAUS-ROBERT MULLER, Berlin Institute of Technology.
 Subject Area: Neuroscience See abstract, page 141.

Poster Session (7:30–11:59pm)

T1 Kernels and learning curves for Gaussian process regression on random graphs

Peter Sollich	peter.sollich@kcl.ac.uk
Kings College London	
Matthew Urry	matthew.urry@kcl.ac.uk
King's College London	
Camille Coti	coti@lri.fr
INRIA Saclay-Ile de France	

We investigate how well Gaussian process regression can learn functions defined on graphs, using large regular random graphs as a paradigmatic example. Random-walk based kernels are shown to have some surprising properties: within the standard approximation of a locally tree-like graph structure, the kernel does not become constant, i.e.neighbouring function values do not become fully correlated, when the lengthscale σ of the kernel is made large. Instead the kernel attains a non-trivial limiting form, which we calculate. The fully correlated limit is reached only once loops become relevant, and we estimate where the crossover to this regime occurs. Our main subject are learning curves of Bayes error versus training set size. We show that these are qualitatively well predicted by a simple approximation using only the spectrum of a large tree as input, and generically scale with n/V, the number of training examples per vertex. We also explore how this behaviour changes once kernel lengthscales are large enough for loops to become important. **Subject Area:** Probabilistic Models and Methods

Spotlight presentation, Tuesday, 3:20.

T2 Locality-sensitive binary codes from shift-invariant kernels

MAXIM RAGINSKY, Duke University, and SVETLANA LAZEBNIK, University of North Carolina.

Subject Area: Unsupervised and Semi-supervised Learning Oral presentation, Tuesday, 11:40am. See abstract, page 102.

T3 Probabilistic Relational PCA

WU-JUN LI	liwujun@cse.ust.hk
Dit-Yan Yeung	dyyeung@cse.ust.hk
HKUST	
Zhihua Zhang	zhzhang@gmail.com
Zhejiang University	

One crucial assumption made by both principal component analysis (PCA) and probabilistic PCA (PPCA) is that the instances are independent and identically distributed (i.i.d.). However, this common i.i.d. assumption is unreasonable for relational data. In this paper, by explicitly modeling covariance between instances as derived from the relational information, we propose a novel probabilistic dimensionality reduction method, called probabilistic relational PCA (PRPCA), for relational data analysis. Although the i.i.d. assumption is no longer adopted in PRPCA, the learning algorithms for PRPCA can still be devised easily like those for PPCA which makes explicit use of the i.i.d. assumption. Experiments on real-world data sets show that PRPCA can effectively utilize the relational information to dramatically outperform PCA and achieve state-of-the-art performance. Subject Area: Unsupervised and Semi-supervised Learning

Spotlight presentation, Tuesday, 3:20.

T4 Decoupling Sparsity and Smoothness in the Discrete Hierarchical Dirichlet Process

CHONG WANG DAVID BLEI Princeton University chongw@cs.princeton.edu blei@cs.princeton.edu

We present a nonparametric hierarchical Bayesian model of document collections that decouples sparsity and smoothness in the component distributions (i.e., the "topics"). In the sparse topic model (STM), each topic is represented by a bank of selector variables that determine which terms appear in the topic. Thus each topic is associated with a subset of the vocabulary, and topic smoothness is modeled on this subset. We develop an efficient Gibbs sampler for the STM that includes a general-purpose method for sampling from a Dirichlet mixture with a combinatorial number of components. We demonstrate the STM on four real-world datasets. Compared to traditional approaches, the empirical results show that STMs give better predictive performance with simpler inferred models. **Subject Area:** Unsupervised and Semi-supervised Learning

Spotlight presentation, Tuesday, 3:20.

T5 Efficient Bregman Range Search

LAWRENCE CAYTON lcayton@tuebingen.mpg.de Max Planck Institute for Biological Cybernetics

We develop an algorithm for efficient range search when the notion of dissimilarity is given by a Bregman divergence. The range search task is to return all points in a potentially large database that are within some specified distance of a query. It arises in many learning algorithms such as locally-weighted regression, kernel density estimation, neighborhood graph-based algorithms, and in tasks like outlier detection and information retrieval. In metric spaces, efficient range search-like algorithms based on spatial data structures have been deployed on a variety of statistical tasks. Here we describe the first algorithm for range search for an arbitrary Bregman divergence. This broad class of dissimilarity measures includes the relative entropy, Mahalanobis distance, Itakura-Saito divergence, and a variety of matrix divergences. Metric methods cannot be directly applied since Bregman divergences do not in general satisfy the triangle inequality. We derive geometric properties of Bregman divergences that yield an efficient algorithm for range search based on a recently proposed space decomposition for Bregman divergences. **Subject Area:** Unsupervised and Semi-supervised Learning

Spotlight presentation, Tuesday, 3:20.

T6 Learning Bregman Distance Functions and Its Application for Semi-Supervised Clustering

Lei Wu	wulei@ntu.edu.sg
Nanyang Technological University	
Rong Jin	rongjin@cse.msu.edu
Michigan State University	
Steven Chu-Hong Hoi	chhoi@ntu.edu.sg
Nanyang Technological University	
Jianke Zhu	zhuji@vision.ee.ethz.ch
ETH Zurich	
Nenghai Yu	ynh@ustc.edu.cn
University of Science and Technolo	ogy of China

Learning distance functions with side information plays a key role in many machine learning and data mining applications. Conventional approaches often assume a Mahalanobis distance function. These approaches are limited in two aspects: (i) they are computationally expensive (even infeasible) for high dimensional data because the size of the metric is in the square of dimensionality; (ii) they assume a fixed metric for the entire input space and therefore are unable to handle heterogeneous data. In this paper, we propose a novel scheme that learns nonlinear Bregman distance functions from side information using a non-parametric approach that is similar to support vector machines. The proposed scheme avoids the assumption of fixed metric because its local distance metric is implicitly derived from the Hessian matrix of a convex function that is used to generate the Bregman distance function. We present an efficient learning algorithm for the proposed scheme for distance function learning. The extensive experiments with semi-supervised clustering show the proposed technique (i) outperforms the state-of-the-art approaches for distance function learning, and (ii) is computationally efficient for high dimensional data.

Subject Area: Unsupervised and Semi-supervised Learning

T7 Linearly constrained Bayesian matrix factorization for blind source separation

MIKKEL SCHMIDT mns@imm.dtu.dk University of Cambridge

We present a general Bayesian approach to probabilistic matrix factorization subject to linear constraints. The approach is based on a Gaussian observation model and Gaussian priors with bilinear equality and inequality constraints. We present an efficient Markov chain Monte Carlo inference procedure based on Gibbs sampling. Special cases of the proposed model are Bayesian formulations of non-negative matrix factorization and factor analysis. The method is evaluated on a blind source separation problem. We demonstrate that our algorithm can be used to extract meaningful and interpretable features that are remarkably different from features extracted using existing related matrix factorization techniques.

Subject Area: Unsupervised and Semi-supervised Learning

T8 Matrix Completion from Noisy Entries

RAGHUNANDAN KESHAVAN ANDREA MONTANARI SEWOONG OH Stanford University raghuram@stanford.edu montanar@stanford.edu swoh@stanford.edu

Given a matrix M of low-rank, we consider the problem of reconstructing it from noisy observations of a small, random subset of its entries. The problem arises in a variety of applications, from collaborative filtering (the 'Netflix problem') to structure-from-motion and positioning. We study a low complexity algorithm introduced in [1], based on a combination of spectral techniques and manifold optimization, that we call here OPTSPACE. We prove performance guarantees that are order-optimal in a number of circumstances.

Subject Area: Unsupervised and Semi-supervised Learning

T9 Nonlinear Learning using Local Coordinate Coding

Kai Yu	kyu@sv.nec-labs.com
NEC Laboratories	
Tong Zhang	tongz@rci.rutgers.edu
Rutgers	
Yihong Gong	ygong@sv.nec-labs.com
NEC Laboratories America	

This paper introduces a new method for semi-supervised learning on high dimensional nonlinear manifolds, which includes a phase of unsupervised basis learning and a phase of supervised function learning. The learned bases provide a set of anchor points to form a local coordinate system, such that each data point x on the manifold can be locally approximated by a linear combination of its nearby anchor points, and the linear weights become its local coordinate coding. We show that a high dimensional nonlinear function can be approximated by a global linear function with respect to this coding scheme, and the approximation quality is ensured by the locality of such coding. The method turns a difficult nonlinear learning problem into a simple global linear learning problem, which overcomes some drawbacks of traditional local learning methods. **Subject Area:** Unsupervised and Semi-supervised Learning

T10 Semi-supervised Regression using Hessian energy with an application to semi-supervised dimensionality reduction

KWANG IN KIMkimki@cs.uni-sb.deSaarland Universityme@floriansteinke.netFLORIAN STEINKEme@floriansteinke.netSiemens Corporate Technology, Munich, GermanyMATTHIAS HEINhein@cs.uni-sb.deSaarland University

114

Semi-supervised regression based on the graph Laplacian suffers from the fact that the solution is biased towards a constant and the lack of extrapolating power. Outgoing from these observations we propose to use the second-order Hessian energy for semi-supervised regression which overcomes both of these problems, in particular, if the data lies on or close to a low-dimensional submanifold in the feature space, the Hessian energy prefers functions which vary "linearly" with respect to the natural parameters in the data. This property makes it also particularly suited for the task of semi-supervised dimensionality reduction where the goal is to find the natural parameters in the data based on a few labeled points. The experimental result suggest that our method is superior to semi-supervised regression using Laplacian regularization and standard supervised methods and is particularly suited for semi-supervised dimensionality reduction.

Subject Area: Unsupervised and Semi-supervised Learning

T11 Group Sparse Coding

SAMY BENGIO	bengio@google.com
Fernando Pereira	pereira@google.com
Yoram Singer	singer@google.com
Dennis Strelow	strelow@google.com
Google	

Bag-of-words document representations are often used in text, image and video processing. While it is relatively easy to determine a suitable word dictionary for text documents, there is no simple mapping from raw images or videos to dictionary terms. The classical approach builds a dictionary using vector quantization over a large set of useful visual descriptors extracted from a training set, and uses a nearest-neighbor algorithm to count the number of occurrences of each dictionary word in documents to be encoded. More robust approaches have been proposed recently that represent each visual descriptor as a sparse weighted combination of dictionary words. While favoring a sparse representation at the level of visual descriptors, those methods however do not ensure that images have sparse representation. In this work, we use mixed-norm regularization to achieve sparsity at the image level as well as a small overall dictionary. This approach can also be used to encourage using the same dictionary words for all the images in a class, providing a discriminative signal in the construction of image representations. Experimental results on a benchmark image classification dataset show that when compact image or dictionary representations are needed for computational efficiency, the proposed approach yields better mean average precision in classification. Subject Area: Unsupervised and Semi-supervised Learning

T12 Label Selection on Graphs

ANDREW GUILLORY JEFF BILMES University of Washington guillory@cs.washington.edu bilmes@ee.washington.edu We investigate methods for selecting sets of labeled vertices for use in predicting the labels of vertices on a graph. We specifically study methods which choose a single batch of labeled vertices (i.e. offline, non sequential methods). In this setting, we find common graph smoothness assumptions directly motivate simple label selection methods with interesting theoretical guarantees. These methods bound prediction error in terms of the smoothness of the true labels with respect to the graph. Some of these bounds give new motivations for previously proposed algorithms, and some suggest new algorithms which we evaluate. We show improved performance over baseline methods on several real world data sets. **Subject Area:** Unsupervised and Semi-supervised Learning

T13 Local Rules for Global MAP: When Do They Work?

Kyomin Jung	kyomin@kaist.edu
Pushmeet Kohli	pkohli@microsoft.com
Microsoft Devavrat Shah	devavrat@mit.edu
MIT	

We consider the question of computing Maximum A Posteriori (MAP) assignment in an arbitrary pair-wise Markov Random Field (MRF). We present a randomized iterative algorithm based on simple local updates. The algorithm, starting with an arbitrary initial assignment, updates it in each iteration by first, picking a random node, then selecting an (appropriately chosen) random local neighborhood and optimizing over this local neighborhood. Somewhat surprisingly, we show that this algorithm finds a near optimal assignment within $2n \ln n$ iterations on average and with high probability for any n node pair-wise MRF with geometry (i.e. MRF graph with polynomial growth) with the approximation error depending on (in a reasonable manner) the geometric growth rate of the graph and the average radius of the local neighborhood – this allows for a graceful tradeoff between the complexity of the algorithm and the approximation error. Through extensive simulations, we show that our algorithm finds extremely good approximate solutions for various kinds of MRFs with geometry.

Subject Area: Probabilistic Models and Methods

T14 Non-Parametric Bayesian Dictionary Learning for Sparse Image Representations

MINGYUAN ZHOU, Duke ECE, HAOJUN CHEN, Duke University, JOHN PAISLEY, LU REN, Duke ECE, GUILLERMO SAPIRO, University of Minnesota, and LAWRENCE CARIN, Duke ECE.

Subject Area: Probabilistic Models and Methods Oral presentation, Tuesday, 11:20am. See abstract, page 101.

JAMES PETTERSONjames.petterson@nicta.com.auTIBERIO CAETANOtiberio.caetano@nicta.com.auJULIAN MCAULEYjulian.mcauley@nicta.com.auNICTAjin.yu@nicta.com.auJIN YUjin.yu@nicta.com.au

We present a method for learning max-weight matching predictors in bipartite graphs. The method consists of performing maximum a posteriori estimation in exponential families with sufficient statistics that encode permutations and data features. Although inference is in general hard, we show that for one very relevant application - document ranking - exact inference is efficient. For general model instances, an appropriate sampler is readily available. Contrary to existing max-margin matching models, our approach is statistically consistent and, in addition, experiments with increasing sample sizes indicate superior improvement over such models. We apply the method to graph matching in computer vision as well as to a standard benchmark dataset for learning document ranking, in which we obtain state-of-the-art results, in particular improving on max-margin variants. The drawback of this method with respect to max-margin alternatives is its runtime for large graphs, which is high comparatively.

Subject Area: Probabilistic Models and Methods

T16 Sensitivity analysis in HMMs with application to likelihood maximization

Pierre-Arnaud Coquelin	pacoquelin@vekia.fr
Vekia	
Romain Deguest	rd2304@columbia.edu
Columbia University	
Remi Munos	remi.munos@inria.fr
INRIA	

This paper considers a sensitivity analysis in Hidden Markov Models with continuous state and observation spaces. We propose an Infinitesimal Perturbation Analysis (IPA) on the filtering distribution with respect to some parameters of the model. We describe a methodology for using any algorithm that estimates the filtering density, such as Sequential Monte Carlo methods, to design an algorithm that estimates its gradient. The resulting IPA estimator is proven to be asymptotically unbiased, consistent and has computational complexity linear in the number of particles. We consider an application of this analysis to the problem of identifying unknown parameters of the model given a sequence of observations. We derive an IPA estimator for the gradient of the log-likelihood, which may be used in a gradient method for the purpose of likelihood maximization. We illustrate the method with several numerical experiments. **Subject Area:** Probabilistic Models and Methods

T15 Exponential Family Graph Matching and Ranking

T17 A Gaussian Tree Approximation for Integer Least-Squares

Jacob Goldberger	goldbej@eng.biu.ac.il
Bar-Ilan U	
Amir Leshem	leshema@eng.biu.ac.il
Bar-Ilan University	

This paper proposes a new algorithm for the linear least squares problem where the unknown variables are constrained to be in a finite set. The factor graph that corresponds to this problem is very loopy; in fact, it is a complete graph. Hence, applying the Belief Propagation (BP) algorithm yields very poor results. The algorithm described here is based on an optimal tree approximation of the Gaussian density of the unconstrained linear system. It is shown that even though the approximation is not directly applied to the exact discrete distribution, applying the BP algorithm to the modified factor graph outperforms current methods in terms of both performance and complexity. The improved performance of the proposed algorithm is demonstrated on the problem of MIMO detection.

Subject Area: Probabilistic Models and Methods

T18 Conditional Random Fields with High-Order Features for Sequence Labeling

Nan Ye	g0701171@nus.edu.sg
National University of Singapo	
Wee Sun Lee	leews@comp.nus.edu.sg
National University of Singapore	
Hai Leong Chieu	chaileon@dso.org.sg
DSO National Laboratories	
Dan Wu	dwu@nus.edu.sg
Singapore MIT Alliance	

Dependencies among neighbouring labels in a sequence is an important source of information for sequence labeling problems. However, only dependencies between adjacent labels are commonly exploited in practice because of the high computational complexity of typical inference algorithms when longer distance dependencies are taken into account. In this paper, we show that it is possible to design efficient inference algorithms for a conditional random field using features that depend on long consecutive label sequences (high-order features), as long as the number of distinct label sequences in the features used is small. This leads to efficient learning algorithms for these conditional random fields. We show experimentally that exploiting dependencies using high-order features can lead to substantial performance improvements for some problems and discuss conditions under which high-order features can be effective. **Subject Area:** Probabilistic Models and Methods

118

T19 Graph Zeta Function in the Bethe Free Energy and Loopy Belief Propagation

Yusuke Watanabe	watay@ism.ac.jp
Kenji Fukumizu	fukumizu@ism.ac.jp
The Institute of Statistical	Mathematics

We propose a new approach to the analysis of Loopy Belief Propagation (LBP) by establishing a formula that connects the Hessian of the Bethe free energy with the edge zeta function. The formula has a number of theoretical implications on LBP. It is applied to give a sufficient condition that the Hessian of the Bethe free energy is positive definite, which shows non-convexity for graphs with multiple cycles. The formula clarifies the relation between the local stability of a fixed point of LBP and local minima of the Bethe free energy. We also propose a new approach to the uniqueness of LBP fixed point, and show various conditions of uniqueness.

Subject Area: Probabilistic Models and Methods

Spotlight presentation, Tuesday, 3:20.

T20 Indian Buffet Processes with Power-law Behavior

Yee Whye Teh	ywteh@gatsby.ucl.ac.uk
Dilan Gorur	dilan@gatsby.ucl.ac.uk
Gatsby Computational Neuros	cience Unit, UCL

The Indian buffet process (IBP) is an exchangeable distribution over binary matrices used in Bayesian nonparametric featural models. In this paper we propose a three-parameter generalization of the IBP exhibiting power-law behavior. We achieve this by generalizing the beta process (the de Finetti measure of the IBP) to the stable-beta process and deriving the IBP corresponding to it. We find interesting relationships between the stable-beta process and the Pitman-Yor process (another stochastic process used in Bayesian nonparametric models with interesting power-law properties). We show that our power-law IBP is a good model for word occurrences in documents with improved performance over the normal IBP. Subject Area: Probabilistic Models and Methods

Spotlight presentation, Tuesday, 5:10.

T21 Inter-domain Gaussian Processes for Sparse Inference using Inducing Features

MIGUEL LAZARO-GREDILLA ANIBAL FIGUEIRAS-VIDAL Universidad Carlos III Madrid

lazarox@gmail.com arfv@tsc.uc3m.es

We present a general inference framework for inter-domain Gaussian Processes (GPs), focusing on its usefulness to build sparse GP models. The state-of-the-art sparse GP model introduced by Snelson and Ghahramani in [1] relies on finding a small, representative pseudo data set of m elements (from the same domain as the n available data elements) which is able to explain existing data well, and then uses it to perform inference. This reduces inference and model selection computation time from $O(n^3)$ to $O(m^{2n})$, where m ii n. Inter-domain GPs can be used to find a (possibly more compact) representative set of features lying in a different domain, at the same computational cost. Being able to specify a different domain for the representative features allows to incorporate prior knowledge about relevant characteristics of data and detaches the functional form of the covariance and basis functions. We will show how previously existing models fit into this framework and will use it to develop two new sparse GP models. Tests on large, representative regression data sets suggest that significant improvement can be achieved, while retaining computational efficiency.

Subject Area: Probabilistic Models and Methods

T22 Bayesian Sparse Factor Models and DAGs Inference and Comparison

Ricardo Henao	rhenao@binf.ku.dk
Ole Winther	owi@imm.dtu.dk
Technical University of Denmark	

In this paper we present a novel approach to learn directed acyclic graphs (DAG) and factor models within the same framework while also allowing for model comparison between them. For this purpose, we exploit the connection between factor models and DAGs to propose Bayesian hierarchies based on spike and slab priors to promote sparsity, heavy-tailed priors to ensure identifiability and predictive densities to perform the model comparison. We require identifiability to be able to produce variable orderings leading to valid DAGs and sparsity to learn the structures. The effectiveness of our approach is demonstrated through extensive experiments on artificial and biological data showing that our approach outperform a number of state of the art methods.

Subject Area: Probabilistic Models and Methods

T23 Learning in Markov Random Fields using Tempered Transitions

RUSLAN SALAKHUTDINOV rsalakhu@cs.toronto.edu Department of Brain and Cognitive Sciences and CSAIL MIT

Markov random fields (MRF's), or undirected graphical models, provide a powerful framework for modeling complex dependencies among random variables. Maximum likelihood learning in MRF's is hard due to the presence of the global normalizing constant. In this paper we consider a class of stochastic approximation algorithms of Robbins-Monro type that uses Markov chain Monte Carlo to do approximate maximum likelihood learning. We show that using MCMC operators based on tempered transitions enables the stochastic approximation algorithm to better explore highly multimodal distributions, which considerably improves parameter estimates inlarge densely-connected MRF's. Our results on MNIST and NORB datasets demonstrate that we can successfully learn good generative models of high-dimensional, richly structured data and perform well on digit and object recognition tasks. Subject Area: Probabilistic Models and Methods

T24 Maximum likelihood trajectories for continuous-time Markov chains

THEODORE PERKINS tperkins@ohri.ca Ottawa Hospital Research Inst.

Continuous-time Markov chains are used to model systems in which transitions between states as well as the time the system spends in each state are random. Many computational problems related to such chains have been solved, including determining state distributions as a function of time, parameter estimation, and control. However, the problem of inferring most likely trajectories, where a trajectory is a sequence of states as well as the amount of time spent in each state, appears unsolved. We study three versions of this problem: (i) an initial value problem, in which an initial state is given and we seek the most likely trajectory until a given final time, (ii) a boundary value problem, in which initial and final states and times are given, and we seek the most likely trajectory connecting them, and (iii) trajectory inference under partial observability, analogous to finding maximum likelihood trajectories for hidden Markov models. We show that maximum likelihood trajectories are not always well-defined, and describe a polynomial time test for welldefinedness. When well-definedness holds, we show that each of the three problems can be solved in polynomial time, and we develop efficient dynamic programming algorithms for doing so.

Subject Area: Probabilistic Models and Methods

Spotlight presentation, Tuesday, 5:10.

T25 Riffled Independence for Ranked Data

JONATHAN HUANG CARLOS GUESTRIN Carnegie Mellon University jch1@cs.cmu.edu guestrin@cs.cmu.edu

Representing distributions over permutations can be a daunting task due to the fact that the number of permutations of n objects scales factorially in n. One recent way that has been used to reduce storage complexity has been to exploit probabilistic independence, but as we argue, full independence assumptions impose strong sparsity constraints on distributions and are unsuitable for modeling rankings. We identify a novel class of independence structures, called riffled independence, which encompasses a more expressive family of distributions while retaining many of the properties necessary for performing efficient inference and reducing sample complexity. In riffled independence, one draws two permutations independently, then performs the riffle shuffle, common in card games, to combine the two permutations to form a single permutation. In ranking, riffled independence corresponds to ranking disjoint sets of objects independently, then interleaving those rankings. We provide a formal introduction and present algorithms for using riffled independence within Fourier-theoretic frameworks which have been explored by a number of recent papers.

Subject Area: Probabilistic Models and Methods

Spotlight presentation, Tuesday, 5:10.

T26 Nonlinear directed acyclic structure learning with weakly additive noise models

Robert Tillman	rtillman@cmu.edu
Carnegie Mellon University	
Arthur Gretton	arthur.gretton@gmail.com
Carnegie Mellon and MPI for	Biological Cybernetics
Peter Spirtes	ps7z@andrew.cmu.edu
Carnegie Mellon	

The recently proposed *additive noise model* has advantages over previous structure learning algorithms, when attempting to recover some true data generating mechanism, since it (i) does not assume linearity or Gaussianity and (ii) can recover a unique DAG rather than an equivalence class. However, its original extension to the multivariate case required enumerating all possible DAGs, and for some special distributions, e.g. linear Gaussian, the model is invertible and thus cannot be used for structure learning. We present a new approach which combines a PC style search using recent advances in kernel measures of conditional dependence with local searches for additive noise models in substructures of the equivalence class. This results in a more computationally efficient approach that is useful for arbitrary distributions even when additive noise models are invertible. Experiments with synthetic and real data show that this method is more accurate than previous methods when data are nonlinear and/or non-Gaussian.

Subject Area: Probabilistic Models and Methods

T27 Abstraction and Relational learning

Charles Kemp	ckemp@cmu.edu
Alan Jern	ajern@cmu.edu
Carnegie Mellon University	

Many categories are better described by providing relational information than listing characteristic features. We present a hierarchical generative model that helps to explain how relational categories are learned and used. Our model learns abstract schemata that specify the relational similarities shared by members of a category, and our emphasis on abstraction departs from previous theoretical proposals that focus instead on comparison of concrete instances. Our first experiment suggests that our abstraction-based account can address some of the tasks that have previously been used to support comparison-based approaches. Our second experiment focuses on one-shot schema learning, a problem that raises challenges for comparison-based approaches but is handled naturally by our abstraction-based account. **Subject Area:** Cognitive Science

Spotlight presentation, Tuesday, 3:20.

T28 Human Rademacher Complexity

XIAOJIN ZHUjerryzhu@cs.wisc.eduTIMOTHY ROGERSttrogers@wisc.eduUniversity of Wisconsin-Madisonbgibson@cs.wisc.eduUniversity of Wisconsinbgibson@cs.wisc.edu

We propose to use Rademacher complexity, originally developed in computational learning theory, as a measure of human learning capacity. Rademacher complexity measures a learner's ability to fit random data, and can be used to bound the learner's true error based on the observed training sample error. We first review the definition of Rademacher complexity and its generalization bound. We then describe a "learning the noise" procedure to experimentally measure human Rademacher complexities. The results from empirical studies showed that: (i) human Rademacher complexity can be successfully measured, (ii) the complexity depends on the domain and training sample size in intuitive ways, (iii) human learning respects the generalization bounds, (iv) the bounds can be useful in predicting the danger of overfitting in human learning. Finally, we discuss the potential applications of human Rademacher complexity in cognitive science.

Subject Area: Cognitive Science

T29 Parallel Inference for Latent Dirichlet Allocation on Graphics Processing Units

FENG YANyan12@purdue.eduPurdue Universityningyixu@microsoft.comNINGYI XUningyixu@microsoft.comMicrosoft Research Asiaalanqi@cs.purdue.eduYUAN QIalanqi@cs.purdue.eduPurdue Universityalanqi@cs.purdue.edu

The recent emergence of Graphics Processing Units (GPUs) as general-purpose parallel computing devices provides us with new opportunities to develop scalable learning methods for massive data. In this work, we consider the problem of parallelizing two inference methods on GPUs for latent Dirichlet Allocation (LDA) models, collapsed Gibbs sampling (CGS) and collapsed variational Bayesian (CVB). To address limited memory constraints on GPUs, we propose a novel data partitioning scheme that effectively reduces the memory cost. Furthermore, the partitioning scheme balances the computational cost on each multiprocessor and enables us to easily avoid memory access conflicts. We also use data streaming to handle extremely large datasets. Extensive experiments showed that our parallel inference methods consistently produced LDA models with the same predictive power as sequential training methods did but with 26x speedup for CGS and 196x speedup for CVB on a GPU with 30 multiprocessors; actually the speedup is almost linearly scalable with the number of multiprocessors available. The proposed partitioning scheme and data streaming can be easily ported to many other models in machine learning.

Subject Area: Hardware

T30 Differential Use of Implicit Negative Evidence in Generative and Discriminative Language Learning

ANNE HSU, University of California Berkeley, and THOMAS GRIFFITHS, University of California, Berkeley. Subject Area: Cognitive Science Oral presentation, Tuesday, 4:10pm. See abstract, page 107.

T31 Adaptive Design Optimization in Experiments with People

DANIEL CAVAGNARO The Ohio State University MARK PITT JAY MYUNG Ohio State University cavagnaro.2@osu.edu

pitt.2@osu.edu myung.1@osu.edu

In cognitive science, empirical data collected from participants are the arbiters in model selection. Model discrimination thus depends on designing maximally informative experiments. It has been shown that adaptive design optimization (ADO) allows one to discriminate models as efficiently as possible in simulation experiments. In this paper we use ADO in a series of experiments with people to discriminate the Power, Exponential, and Hyperbolic models of memory retention, which has been a long-standing problem in cognitive science, providing an ideal setting in which to test the application of ADO for addressing questions about human cognition. Using an optimality criterion based on mutual information, ADO is able to find designs that are maximally likely to increase our certainty about the true model upon observation of the experiment outcomes. Results demonstrate the usefulness of ADO and also reveal some challenges in its implementation.

Subject Area: Cognitive Science

T32 Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model

ED VUL, MICHAEL FRANK, Massachusetts Institute of Technology, GEORGE ALVAREZ, Dept. of Psychology, Harvard University, and JOSHUA TENENBAUM, Massachusetts Institute of Technology.

Subject Area: Cognitive Science Oral presentation, Tuesday, 4:30pm. See abstract, page 107.

T33 Factor Modeling for Advertisement Targeting

YE CHEN	yechen1@ebay.com
EBay Research Labs	
Michael Kapralov	kapralov@stanford.edu
Stanford University	
DMITRY PAVLOV	dmitry-pavlov@yandex-team.ru
Yandex Labs	
John Canny	jfc@cs.berkeley.edu
University of California, Berkeley	

124

We adapt a probabilistic latent variable model, namely GaP (Gamma-Poisson), to ad targeting in the contexts of sponsored search (SS) and behaviorally targeted (BT) display advertising. We also approach the important problem of ad positional bias by formulating a one-latent-dimension GaP factorization. Learning from click-through data is intrinsically large scale, even more so for ads. We scale up the algorithm to terabytes of real-world SS and BT data that contains hundreds of millions of users and hundreds of thousands of features, by leveraging the scalability characteristics of the algorithm and the inherent structure of the problem including data sparsity and locality. Specifically, we demonstrate two somewhat orthogonal philosophies of scaling algorithms to large-scale problems, through the SS and BT implementations, respectively. Finally, we report the experimental results using Yahoo's vast datasets, and show that our approach substantially outperform the state-of-the-art methods in prediction accuracy. For BT in particular, the ROC area achieved by GaP is exceeding 0.95, while one prior approach using Poisson regression vielded 0.83. For computational performance, we compare a single-node sparse implementation with a parallel implementation using Hadoop MapReduce, the results are counterintuitive yet quite interesting. We therefore provide insights into the underlying principles of large-scale learning. Subject Area: Applications

T34 Clustering sequence sets for motif discovery

Jong Kyoung Kim Seungjin Choi POSTECH blkimjk@postech.ac.kr seungjin@postech.ac.kr

Most of existing methods for DNA motif discovery consider only a single set of sequences to find an over-represented motif. In contrast, we consider multiple sets of sequences where we group sets associated with the same motif into a cluster, assuming that each set involves a single motif. Clustering sets of sequences yields clusters of coherent motifs, improving signal-to-noise ratio or enabling us to identify multiple motifs. We present a probabilistic model for DNA motif discovery where we identify multiple motifs through searching for patterns which are shared across multiple sets of sequences. Our model infers cluster-indicating latent variables and learns motifs simultaneously, where these two tasks interact with each other. We show that our model can handle various motif discovery problems, depending on how to construct multiple sets of sequences. Experiments on three different problems for discovering DNA motifs emphasize the useful behavior and confirm the substantial gains over existing methods where only single set of sequences is considered.

Subject Area: Applications

T35 Time-Varying Dynamic Bayesian Networks

LE SONG MLADEN KOLAR ERIC XING Carnegie Mellon University lesong@cs.cmu.edu
mladenk@cs.cmu.edu
epxing@cs.cmu.edu

Directed graphical models such as Bayesian networks are a favored formalism to model the dependency structures in complex multivariate systems such as those encountered in biology and neural sciences. When the system is undergoing dynamic transformation, often a temporally rewiring network is needed for capturing the dynamic causal influences between covariates. In this paper, we propose a time-varying dynamic Bayesian network (TV-DBN) for modeling the structurally varying directed dependency structures underlying non-stationary biological/neural time series. This is a challenging problem due the non-stationarity and sample scarcity of the time series. We present a kernel reweighted ℓ_1 regularized auto-regressive procedure for learning the TV-DBN model. Our method enjoys nice properties such as computational efficiency and provable asymptotic consistency. Applying TV-DBN to time series measurements during yeast cell cycle and brain response to visual stimuli reveals interesting dynamics underlying the respective biological systems.

Subject Area: Applications

Spotlight presentation, Tuesday, 5:10.

T36 Randomized Pruning: Efficiently Calculating Expectations in Large Dynamic Programs

 ALEXANDRE BOUCHARD-CÔTÉ
 bouchard@cs.berkeley.edu

 UC Berkeley
 slav@petrovi.de

 SLAV PETROV
 slav@petrovi.de

 Google Research
 klein@cs.berkeley.edu

 DAN KLEIN
 klein@cs.berkeley.edu

 University of California Berkeley
 bouchard@cs.berkeley.edu

Pruning can massively accelerate the computation of feature expectations in large models. However, any single pruning mask will introduce bias. We present a novel approach which employs a randomized sequence of pruning masks. Formally, we apply auxiliary variable MCMC sampling to generate this sequence of masks, thereby gaining theoretical guarantees about convergence. Because each mask is generally able to skip large portions of an underlying dynamic program, our approach is particularly compelling for high-degree algorithms. Empirically, we demonstrate our method on bilingual parsing, showing decreasing bias as more masks are incorporated, and outperforming fixed tic-tac-toe pruning.

Subject Area: Applications

 $Spotlight\ presentation,\ Tuesday,\ 5:10.$

T38 Multi-Label Prediction via Compressed Sensing

DANIEL HSU, University of California, SHAM KAKADE, Toyota Technological Institute, JOHN LANGFORD, Yahoo! Research, and TONG ZHANG, Rutgers University. Subject Area: Theory Oral presentation, Tuesday, 9:30am. See abstract, page 99.

T39 Data-driven calibration of linear estimators with minimal penalties

Sylvain Arlot	sylvain.arlot@ens.fr
CNRS - ENS	
Francis Bach	francis.bach@mines.org
Ecole Normale Superieure	

This paper tackles the problem of selecting among several linear estimators in non-parametric regression; this includes model selection for linear regression, the choice of a regularization parameter in kernel ridge regression or spline smoothing, and the choice of a kernel in multiple kernel learning. We propose a new algorithm which first estimates consistently the variance of the noise, based upon the concept of minimal penalty which was previously introduced in the context of model selection. Then, plugging our variance estimate in Mallows' C_L penalty is proved to lead to an algorithm satisfying an oracle inequality. Simulation experiments with kernel ridge regression and multiple kernel learning show that the proposed algorithm often improves significantly existing calibration procedures such as 10-fold cross-validation or generalized cross-validation.

Subject Area: Theory

T40 Regularized Distance Metric Learning: Theory and Algorithm

Rong Jin	rongjin@cse.msu.edu
Michigan State University	
Shijun Wang	wangshi@cc.nih.gov
National Institutes of Health	
Yang Zhou	zhouyang@msu.edu
Michigan State University	

In this paper, we examine the generalization error of regularized distance metric learning. We show that with appropriate constraints, the generalization error of regularized distance metric learning could be independent from the dimensionality, making it suitable for handling high dimensional data. In addition, we present an efficient online learning algorithm for regularized distance metric learning. Our empirical studies with data classification and face recognition show that the proposed algorithm is (i) effective for distance metric learning when compared to the state-of-the-art methods, and (ii) efficient and robust for high dimensional data. **Subject Area:** Theory

T41 Online Learning of Assignments

Matthew Streeter	mstreeter@google.com
Google	
DANIEL GOLOVIN	dgolovin@cs.cmu.edu
Carnegie Mellon University	
Andreas Krause	krausea@caltech.edu
California Institute of Technology	

Which ads should we display in sponsored search in order to maximize our revenue? How should we dynamically rank information sources to maximize value of information? These applications exhibit strong diminishing returns: Selection of redundant ads and information sources decreases their marginal utility. We show that these and other problems can be formalized as repeatedly selecting an assignment of items to positions to maximize a sequence of monotone submodular functions that arrive one by one. We present an efficient algorithm for this general problem and analyze it in the no-regret model. Our algorithm is equipped with strong theoretical guarantees, with a performance ratio that converges to the optimal constant of 1-1/e. We empirically evaluate our algorithms on two real-world online optimization problems on the web: ad allocation with submodular utilities, and dynamically ranking blogs to detect information cascades.

Subject Area: Theory

Spotlight presentation, Tuesday, 3:20.

T42 A General Projection Property for Distribution Families

YAO-LIANG YU YUXI LI DALE SCHUURMANS CSABA SZEPESVARI University of Alberta yaoliang@cs.ualberta.ca yuxi@cs.ualberta.ca dale@cs.ualberta.ca szepesva@cs.ualberta.ca

We prove that linear projections between distribution families with fixed first and second moments are surjective, regardless of dimension. We further extend this result to families that respect additional constraints, such as symmetry, unimodality and log-concavity. By combining our results with classic univariate inequalities, we provide new worst-case analyses for natural risk criteria arising in different fields. One discovery is that portfolio selection under the worst-case value-at-risk and conditional value-at-risk criteria yields identical portfolios.

Subject Area: Optimization

T43 Submodularity Cuts and Applications

Yoshinobu Kawahara	kawahara@ar.sanken.osaka-u.ac.jp
Osaka University	
Kiyohito Nagano	nagano@is.titech.ac.jp
Tokyo Institute of Technology	
Koji Tsuda	koji.tsuda@aist.go.jp
Advanced Industrial Science and	Technology
Jeff Bilmes	bilmes@u.washington.edu
University of Washington	

Several key problems in machine learning, such as feature selection and active learning, can be formulated as submodular set function maximization. We present herein a novel algorithm for maximizing a submodular set function under a cardinality constraint — the algorithm is based on a cutting-plane method and is implemented as an iterative smallscale binary-integer linear programming procedure. It is well known that this problem is NP-hard, and the approximation factor achieved by the greedy algorithm is the theoretical limit for polynomial time. As for (non-polynomial time) exact algorithms that perform reasonably in practice, there has been very little in the literature although the problem is quite important for many applications. Our algorithm is guaranteed to find the exact solution in finite iterations, and it converges fast in practice due to the efficiency of the cutting-plane mechanism. Moreover, we also provide a method that produces successively decreasing upper-bounds of the optimal solution, while our algorithm provides successively increasing lower-bounds. Thus, the accuracy of the current solution can be estimated at any point, and the algorithm can be stopped early once a desired degree of tolerance is met. We evaluate our algorithm on sensor placement and feature selection applications showing good performance.

Subject Area: Optimization

Spotlight presentation, Tuesday, 5:10.

T44 Linear-time Algorithms for Pairwise Statistical Problems

Parikshit Ram	p.ram@gatech.edu
Georgia Institute of Technolog	
Dongryeol Lee	dongryel@cc.gatech.edu
Georgia Institute of Technology	
WILLIAM MARCH	march@gatech.edu
Georgia Tech	
Alexander Gray	agray@cc.gatech.edu
Georgia Institute of Technology	

Several key computational bottlenecks in machine learning involve pairwise distance computations, including all-nearest-neighbors (finding the nearest neighbor(s) for each point, e.g. in manifold learning) and kernel summations (e.g. in kernel density estimation or kernel machines). We consider the general, bichromatic case for these problems, in addition to the scientific problem of N-body potential calculation. In this paper we show for the first time O(N) worst case runtimes for practical algorithms for these problems based on the cover tree data structure (Beygelzimer, Kakade, Langford, 2006). Subject Area: Theory

Spotlight presentation, Tuesday, 3:20.

T45 On the Convergence of the Concave-Convex Procedure

Bharath Sriperumbudur	bharathsv@ucsd.edu
UC San Diego	
Gert Lanckriet	gert@ece.ucsd.edu
University of California	

The concave-convex procedure (CCCP) is a majorization-minimization algorithm that solves d.c. (difference of convex functions) programs as a sequence of convex programs. In machine learning, CCCP is extensively used in many learning algorithms like sparse support vector machines (SVMs), transductive SVMs, sparse principal component analysis, etc. Though widely used in many applications, the convergence behavior of CCCP has not gotten a lot of specific attention. Yuille and Rangarajan analyzed its convergence in their original paper, however, we believe the analysis is not complete. Although the convergence of CCCP can be derived from the convergence of the d.c. algorithm (DCA), their proof is more specialized and technical than actually required for the specific case of CCCP. In this paper, we follow a different reasoning and show how Zangwill's global convergence theory of iterative algorithms provides a natural framework to prove the convergence of CCCP, allowing a more elegant and simple proof. This underlines Zangwill's theory as a powerful and general framework to deal with the convergence issues of iterative algorithms, after also being used to prove the convergence of algorithms like expectation-maximization, generalized alternating minimization, etc. In this paper, we provide a rigorous analysis of the convergence of CCCP by addressing these questions: (i) When does CCCP find a local minimum or a stationary point of the d.c. program under consideration? (ii) When does the sequence generated by CCCP converge? We also present an open problem on the issue of local convergence of CCCP. Subject Area: Optimization

T46 Slow Learners are Fast

Martin Zinkevich	maz@yahoo-inc.com
Alex Smola	alex@smola.org
John Langford	jl@yahoo-inc.com
Yahoo!	

Online learning algorithms have impressive convergence properties when it comes to risk minimization and convex games on very large problems. However, they are inherently sequential in their design which prevents them from taking advantage of modern multi-core architectures. In this paper we prove that online learning with delayed updates converges well, thereby facilitating parallel online learning. **Subject Area:** Optimization

T47 Accelerated Gradient Methods for Stochastic Optimization and Online Learning

Chonghai Hu	hino.hu@gmail.com
JAMES KWOK	jamesk@cse.ust.hk
Weike Pan	weikep@cse.ust.hk
Hong Kong University of Science	and Technology

Regularized risk minimization often involves non-smooth optimization, either because of the loss function (e.g., hinge loss) or the regularizer (e.g., ℓ_1 -regularizer). Gradient descent methods, though highly scalable and easy to implement, are known to converge slowly on these problems. In this paper, we develop novel accelerated gradient methods for stochastic optimization while still preserving their computational simplicity and scalability. The proposed algorithm, called SAGE (Stochastic Accelerated GradiEnt), exhibits fast convergence rates on stochastic optimization with both convex and strongly convex objectives. Experimental results show that SAGE is faster than recent (sub)gradient methods including FOLOS, SMIDAS and SCD. Moreover, SAGE can also be extended for online learning, resulting in a simple but powerful algorithm. **Subject Area:** Optimization

T48 Noisy Generalized Binary Search

ROB NOWAK nowak@ece.wisc.edu University of Wisconsin,Madison

This paper addresses the problem of noisy Generalized Binary Search (GBS). GBS is a wellknown greedy algorithm for determining a binary-valued hypothesis through a sequence of strategically selected queries. At each step, a query is selected that most evenly splits the hypotheses under consideration into two disjoint subsets, a natural generalization of the idea underlying classic binary search. GBS is used in many applications, including fault testing, machine diagnostics, disease diagnosis, job scheduling, image processing, computer vision, and active learning. In most of these cases, the responses to queries can be noisy. Past work has provided a partial characterization of GBS, but existing noise-tolerant versions of GBS are suboptimal in terms of sample complexity. This paper presents the first optimal algorithm for noisy GBS and demonstrates its application to learning multidimensional threshold functions. **Subject Area:** Theory

T49 Thresholding Procedures for High Dimensional Variable Selection and Statistical Estimation

Shuheng Zhou ETH Zurich szhou@cs.cmu.edu

Given n noisy samples with p dimensions, where $n \ll p$, we show that the multi-stage thresholding procedures can accurately estimate a sparse vector $\beta \in \mathbb{R}^p$ in a linear model, under the restricted eigenvalue conditions (Bickel-Ritov-Tsybakov 09). Thus our conditions for model selection consistency are considerably weaker than what has been achieved in previous works. More importantly, this method allows very significant values of s, which is the number of non-zero elements in the true parameter β . For example, it works for cases where the ordinary Lasso would have failed. Finally, we show that if X obeys a uniform uncertainty principle and if the true parameter is sufficiently sparse, the Gauss-Dantzig selector (Candés-Tao 07) achieves the ℓ_2 loss within a logarithmic factor of the ideal mean square error one would achieve with an oracle which would supply perfect information about which coordinates are non-zero and which are above the noise level, while selecting a sufficiently sparse model. **Subject Area:** Theory

131

T50 Construction of Nonparametric Bayesian Models from Parametric Bayes Equations

PETER ORBANZ p.orbanz@eng.cam.ac.uk University of Cambridge

We consider the general problem of constructing nonparametric Bayesian models on infinite-dimensional random objects, such as functions, infinite graphs or infinite permutations. The problem has generated much interest in machine learning, where it is treated heuristically, but has not been studied in full generality in nonparametric Bayesian statistics, which tends to focus on models over probability distributions. Our approach applies a standard tool of stochastic process theory, the construction of stochastic processes from their finite-dimensional marginal distributions. The main contribution of the paper is a generalization of the classic Kolmogorov extension theorem to conditional probabilities. This extension allows a rigorous construction of nonparametric Bayesian models from systems of finite-dimensional, parametric Bayes equations. Using this approach, we show (i) how existence of a conjugate posterior for the nonparametric model can be guaranteed by choosing conjugate finite-dimensional models in the construction, (ii) how the mapping to the posterior parameters of the nonparametric model can be explicitly determined, and (iii) that the construction of conjugate models in essence requires the finite-dimensional models to be in the exponential family. As an application of our constructive framework, we derive a model on infinite permutations, the nonparametric Bayesian analogue of a model recently proposed for the analysis of rank data. Subject Area: Theory

Spotlight presentation, Tuesday, 5:10.

T51 Adapting to the Shifting Intent of Search Queries

UMAR SYED Princeton University ALEKSANDRS SLIVKINS NINA MISHRA Microsoft Research usyed@cs.princeton.edu slivkins@microsoft.com

ninam@microsoft.com

Search engines today present results that are often oblivious to recent shifts in intent. For example, the meaning of the query 'independence day' shifts in early July to a US holiday and to a movie around the time of the box office release. While no studies exactly quantify the magnitude of intent-shifting traffic, studies suggest that news events, seasonal topics, pop culture, etc account for 1/2 the search queries. This paper shows that the signals a search engine receives can be used to both determine that a shift in intent happened, as well as find a result that is now more relevant. We present a meta-algorithm that marries a classifier with a bandit algorithm to achieve regret that depends logarithmically on the number of query impressions, under certain assumptions. We provide strong evidence that this regret is close to the best achievable. Finally, via a series of experiments, we demonstrate that our algorithm outperforms prior approaches, particularly as the amount of intent-shifting traffic increases.

Subject Area: Theory

SAHAND NEGAHBAN, UC, Berkeley, PRADEEP RAVIKUMAR, MARTIN WAINWRIGHT and BIN YU, University of California. Subject Area: Theory Oral presentation, Tuesday, 9:50am. See abstract, page 99.

T53 Beyond Categories: The Visual Memex Model for Reasoning About Object Relationships

TOMASZ MALISIEWICZ ALYOSHA EFROS Carnegie Mellon University tomasz@cmu.edu efros@cs.cmu.edu

The use of context is critical for scene understanding in computer vision, where the recognition of an object is driven by both local appearance and the object's relationship to other elements of the scene (context). Most current approaches rely on modeling the relationships between object categories as a source of context. In this paper we seek to move beyond categories to provide a richer appearance-based model of context. We present an exemplar-based model of objects and their relationships, the Visual Memex, that encodes both local appearance and 2D spatial context between object instances. We evaluate our model on Torralba's proposed Context Challenge against a baseline category-based system. Our experiments suggest that moving beyond categories for context modeling appears to be quite beneficial, and may be the critical missing ingredient in scene understanding systems. **Subject Area:** Vision

T54 L₁-Penalized Robust Estimation for a Class of Inverse Problems Arising in Multiview Geometry

ARNAK DALALYAN RENAUD KERIVEN Université Paris-Est, ENPC dalalyan@certis.enpc.fr keriven@certis.enpc.fr

We propose a new approach to the problem of robust estimation in multiview geometry. Inspired by recent advances in the sparse recovery problem of statistics, our estimator is defined as a Bayesian maximum a posteriori with multivariate Laplace prior on the vector describing the outliers. This leads to an estimator in which the fidelity to the data is measured by the L_{∞} -norm while the regularization is done by the L_1 -norm. The proposed procedure is fairly fast since the outlier removal is done by solving one linear program (LP). An important difference compared to existing algorithms is that for our estimator it is not necessary to specify neither the number nor the proportion of the outliers. The theoretical results, as well as the numerical example reported in this work, confirm the efficiency of the proposed approach.

Subject Area: Vision

T55 An Additive Latent Feature Model for Transparent Object Recognition

MARIO FRITZ, UC Berkelev EECS and ICSI, MICHAEL BLACK, Brown University, GARY BRADSKI, Willow Garage, SERGEY KARAYEV, UC Berkeley, and TREVOR DARRELL, UC Berkeley EECS and ICSI. Subject Area: Vision

Oral presentation, Tuesday, 10:40am. See abstract, page 100.

T56 Semi-Supervised Learning in Gigantic Image Collections

ROB FERGUS, New York University, YAIR WEISS, Hebrew University, and ANTONIO TORRALBA, Massachusetts Institute of Technology. Subject Area: Vision Oral presentation, Tuesday, 11:00am. See abstract, page 101.

T57 Speaker Comparison with Inner Product Discriminant Functions

WILLIAM CAMPBELL wcampbell@ll.mit.edu MIT Lincoln Laboratory ZAHI KARAM MIT DOUGLAS STURIM MIT Lincoln Laboratory

zahi@mit.edu sturim@ll mit edu

Speaker comparison, the process of finding the speaker similarity between two speech signals, occupies a central role in a variety of applications—speaker verification, clustering, and identification. Speaker comparison can be placed in a geometric framework by casting the problem as a model comparison process. For a given speech signal, feature vectors are produced and used to adapt a Gaussian mixture model (GMM). Speaker comparison can then be viewed as the process of compensating and finding metrics on the space of adapted models. We propose a framework, inner product discriminant functions (IPDFs), which extends many common techniques for speaker comparison: support vector machines, joint factor analysis, and linear scoring. The framework uses inner products between the parameter vectors of GMM models motivated by several statistical methods. Compensation of nuisances is performed via linear transforms on GMM parameter vectors. Using the IPDF framework, we show that many current techniques are simple variations of each other. We demonstrate, on a 2006 NIST speaker recognition evaluation task, new scoring methods using IPDFs which produce excellent error rates and require significantly less computation than current techniques.

Subject Area: Speech and Signal Processing

00	0 0	01
LIANG SUN		sun.liang@asu.edu
Jun Liu		j.liu@asu.edu
Jianhui Chen		Jianhui.Chen@asu.edu
Jieping Ye		jieping.ye@asu.edu
Arizona State Universi	$\mathbf{t}\mathbf{v}$	

T58 Efficient Recovery of Jointly Sparse Vectors

We consider the reconstruction of sparse signals in the multiple measurement vector (MMV) model, in which the signal, represented as a matrix, consists of a set of jointly sparse vectors. MMV is an extension of the single measurement vector (SMV) model employed in standard compressive sensing (CS). Recent theoretical studies focus on the convex relaxation of the MMV problem based on the (2, 1)-norm minimization, which is an extension of the well-known 1-norm minimization employed in SMV. However, the resulting convex optimization problem in MMV is significantly much more difficult to solve than the one in SMV. Existing algorithms reformulate it as a second-order cone programming (SOCP) or semidefinite programming (SDP), which is computationally expensive to solve for problems of moderate size. In this paper, we propose a new (dual) reformulation of the convex optimization problem in MMV and develop an efficient algorithm based on the prox-method. Interestingly, our theoretical analysis reveals the close connection between the proposed reformulation and multiple kernel learning. Our simulation studies demonstrate the scalability of the proposed algorithm.

Subject Area: Speech and Signal Processing

T59 Extending Phase Mechanism to Differential Motion Opponency for Motion Pop-out

YICONG MENG	eeyicong@ust.hk
Bertram Shi	eebert@ee.ust.hk
Hong Kong University of Science	and Technology

We extend the concept of phase tuning, a ubiquitous mechanism in sensory neurons including motion and disparity detection neurons, to the motion contrast detection. We demonstrate that motion contrast can be detected by phase shifts between motion neuronal responses in different spatial regions. By constructing the differential motion opponency in response to motions in two different spatial regions, varying motion contrasts can be detected, where similar motion is detected by zero phase shifts and differences in motion by non-zero phase shifts. The model can exhibit either enhancement or suppression of responses by either different or similar motion in the surrounding. A primary advantage of the model is that the responses are selective to relative motion instead of absolute motion, which could model neurons found in neurophysiological experiments responsible for motion pop-out detection. **Subject Area:** Vision

T60 An Integer Projected Fixed Point Method for Graph Matching and MAP Inference

Marius Leordeanu	leordeanu@gmail.com
Carnegie Mellon University	
Martial Hebert	hebert@ri.cmu.edu
CMU	
Rahul Sukthankar	rahuls@cs.cmu.edu
Intel	

Graph matching and MAP inference are essential problems in computer vision and machine learning. We introduce a novel algorithm that can accommodate both problems and solve them efficiently. Recent graph matching algorithms are based on a general quadratic programming formulation, that takes in consideration both unary and second-order terms reflecting the similarities in local appearance as well as in the pairwise geometric relationships between the matched features. In this case the problem is NP-hard and a lot of effort has been spent in finding efficiently approximate solutions by relaxing the constraints of the original problem. Most algorithms find optimal continuous solutions of the modified problem, ignoring during the optimization the original discrete constraints. The continuous solution is quickly binarized at the end, but very little attention is put into this final discretization step. In this paper we argue that the stage in which a discrete solution is found is crucial for good performance. We propose an efficient algorithm, with climbing and convergence properties, that optimizes in the discrete domain the quadratic score, and it gives excellent results either by itself or by starting from the solution returned by any graph matching algorithm. In practice it outperforms state-or-the art algorithms and it also significantly improves their performance if used in combination. When applied to MAP inference, the algorithm is a parallel extension of Iterated Conditional Modes (ICM) with climbing and convergence properties that make it a compelling alternative to the sequential ICM. In our experiments on MAP inference our algorithm proved its effectiveness by outperforming ICM and Max-Product Belief Propagation.

Subject Area: Vision

T61 Filtering Abstract Senses From Image Search Results

Kate Saenko	saenko@mit.edu
MIT	
Trevor Darrell	<pre>trevor@eecs.berkeley.edu</pre>
UC Berkeley EECS and ICSI	

We propose an unsupervised method that, given a word, automatically selects non-abstract senses of that word from an online ontology and generates images depicting the corresponding entities. When faced with the task of learning a visual model based only on the name of an object, a common approach is to find images on the web that are associated with the object name, and then train a visual classifier from the search result. As words are generally polysemous, this approach can lead to relatively noisy models if many examples due to outlier senses are added to the model. We argue that images associated with an abstract word sense should be excluded when training a visual classifier to learn a model of a physical object. While image clustering can group together visually coherent sets of returned images, it can be difficult to distinguish whether an image cluster relates to a desired object or to an abstract sense of the word. We propose a method that uses both image features and the text associated with the images to relate latent topics to particular senses. Our model does not require any human supervision, and takes as input only the name of an object category. We show results of retrieving concrete-sense images in two available multimodal, multi-sense databases, as well as experiment with object classifiers trained on concrete-sense images returned by our method for a set of ten common office objects. **Subject Area:** Vision

T62 Optimizing Multi-Class Spatio-Spectral Filters via Bayes Error Estimation for EEG Classification

WENMING ZHENG, Southeast University, P.R. China, and ZHOUCHEN LIN, Microsoft Research Asia, P.R. China. Subject Area: Neuroscience Oral presentation, Tuesday, 4:50pm. See abstract, page 108.

T63 No evidence for active sparsification in the visual cortex

PIETRO BERKES BEN WHITE JOZSEF FISER Brandeis University

berkes@brandeis.edu blwhite@brandeis.edu fiser@brandeis.edu

The proposal that cortical activity in the visual cortex is optimized for sparse neural activity is one of the most established ideas in computational neuroscience. However, direct experimental evidence for optimal sparse coding remains inconclusive, mostly due to the lack of reference values on which to judge the measured sparseness. Here we analyze neural responses to natural movies in the primary visual cortex of ferrets at different stages of development, and of rats while awake and under different levels of anesthesia. In contrast with prediction from a sparse coding model, our data shows that population and lifetime sparseness decrease with visual experience, and increase from the awake to anesthetized state. These results suggest that the representation in the primary visual cortex is not actively optimized to maximize sparseness. **Subject Area:** Neuroscience

T64 A joint maximum-entropy model for binary neural population patterns and continuous signals

Sebastian Gerwinn	sgerwinn@tuebingen.mpg.de
Philipp Berens	berens@tuebingen.mpg.de
Matthias Bethge	mbethge@tuebingen.mpg.de
MPI for Biological Cybernetics &	z University of Tübingen

Second-order maximum-entropy models have recently gained much interest for describing the statistics of binary spike trains. Here, we extend this approach to take continuous stimuli into account as well. By constraining the joint second-order statistics, we obtain a joint Gaussian-Boltzmann distribution of continuous stimuli and binary neural firing patterns, for which we also compute marginal and conditional distributions. This model has the same computational complexity as pure binary models and fitting it to data is a convex problem. We show that the model can be seen as an extension to the classical spike-triggered average/covariance analysis and can be used as a non-linear method for extracting features which a neural population is sensitive to. Further, by calculating the posterior distribution of stimuli given an observed neural response, the model can be used to decode stimuli and yields a natural spike-train metric. Therefore, extending the framework of maximum-entropy models to continuous variables allows us to gain novel insights into the relationship between the firing patterns of neural ensembles and the stimuli they are processing.

Subject Area: Neuroscience

Spotlight presentation, Tuesday, 3:20.

T65 Unsupervised feature learning for audio classification using convolutional deep belief networks

HONGLAK LEE PETER PHAM YAN LARGMAN ANDREW NG Stanford University hllee@cs.stanford.edu ptpham@cs.stanford.edu yanl@cs.stanford.edu ang@cs.stanford.edu

In recent years, deep learning approaches have gained significant interest as a way of building hierarchical representations from unlabeled data. However, to our knowledge, these deep learning approaches have not been extensively studied for auditory data. In this paper, we apply convolutional deep belief networks to audio data and empirically evaluate them on various audio classification tasks. For the case of speech data, we show that the learned features correspond to phones/phonemes. In addition, our feature representations trained from unlabeled audio data show very good performance for multiple audio classification tasks. We hope that this paper will inspire more research on deep learning approaches applied to a wide range of audio recognition tasks. **Subject Area:** Speech and Signal Processing

T66 Hierarchical Mixture of Classification Experts Uncovers Interactions between Brain Regions

BANGPENG YAO Stanford University DIRK WALTHER DIANE BECK University of Illinois LI FEI-FEI Princeton University bangpeng@cs.stanford.edu
walther@uiuc.edu
dmbeck@uiuc.edu
feifeili@CS.Princeton.edu

The human brain can be described as containing a number of functional regions. For a given task, these regions, as well as the connections between them, play a key role in information processing in the brain. However, most existing multi-voxel pattern analysis approaches either treat multiple functional regions as one large uniform region or several independent regions, ignoring the connections between regions. In this paper, we propose to model such connections in an Hidden Conditional Random Field (HCRF) framework, where the classifier of one region of interest (ROI) makes predictions based on not only its voxels but also the classifier predictions from ROIs that it connects to. Furthermore, we propose a structural learning method in the HCRF framework to automatically uncover the connections between ROIs. Experiments on fMRI data acquired while human subjects viewing images of natural scenes show that our model can improve the top-level (the classifier combining information from all ROIs) and ROI-level prediction accuracy, as well as uncover some meaningful connections between ROIs. **Subject Area:** Neuroscience

T67 Statistical Models of Linear and Nonlinear Contextual Interactions in Early Visual Processing

 RUBEN COEN-CAGLI
 ruben.coencagli@einstein.yu.edu

 Albert Einstein College of Medicine

 PETER DAYAN
 dayan@gatsby.ucl.ac.uk

 University College London

 ODELIA SCHWARTZ
 odelia.schwartz@einstein.yu.edu

 Albert Einstein College of Medicine

A central hypothesis about early visual processing is that it represents inputs in a coordinate system matched to the statistics of natural scenes. Simple versions of this lead to Gabor-like receptive fields and divisive gain modulation from local surrounds; these have led to influential neural and psychological models of visual processing. However, these accounts are based on an incomplete view of the visual context surrounding each point. Here, we consider an approximate model of linear and non-linear correlations between the responses of spatially distributed Gabor-like receptive fields, which, when trained on an ensemble of natural scenes, unifies a range of spatial context effects. The full model accounts for neural surround data in primary visual cortex (V1), provides a statistical foundation for perceptual phenomena associated with Li's (2002) hypothesis that V1 builds a saliency map, and fits data on the tilt illusion. **Subject Area:** Neuroscience

Spotlight presentation, Tuesday, 5:10.

T68 Discriminative Network Models of Schizophrenia

GUILLERMO CECCHI, IRINA RISH, IBM T.J. Watson Research Center, BENJAMIN THYREAU, Neurospin, CEA-INSERM, BERTRAND THIRION, INRIA, MARION PLAZE, CEA-INSERM, MARIE-LAURE PAILLERE-MARTINOT, AP-HP, CATHERINE MARTELLI, Centre Hospitalier Paul Brousse, JEAN-LUC MARTINOT, INSERM - CEA - Univ. Paris Sud, and JEAN-BAPTISTE POLINE, Neurospin, CEA-INSERM. **Subject Area:** Neuroscience Oral presentation, Tuesday, 3:00pm. See abstract, page 104.

T69 STDP enables spiking neurons to detect hidden causes of their inputs

Bernhard Nessler	nessler@igi.tugraz.at
Michael Pfeiffer	pfeiffer@igi.tugraz.at
Wolfgang Maass	maass@igi.tugraz.at
Graz University of Technology	

The principles by which spiking neurons contribute to the astounding computational power of generic cortical microcircuits, and how spike-timing-dependent plasticity (STDP) of synaptic weights could generate and maintain this computational function, are unknown. We show here that STDP, in conjunction with a stochastic soft winner-take-all (WTA) circuit, induces spiking neurons to generate through their synaptic weights implicit internal models for subclasses (or "causes") of the high-dimensional spike patterns of hundreds of pre-synaptic neurons. Hence these neurons will fire after learning whenever the current input best matches their internal model. The resulting computational function of soft WTA circuits, a common network motif of cortical microcircuits, could therefore be a drastic dimensionality reduction of information streams, together with the autonomous creation of internal models for the probability distributions of their input patterns. We show that the autonomous generation and maintenance of this computational function can be explained on the basis of rigorous mathematical principles. In particular, we show that STDP is able to approximate a stochastic online Expectation-Maximization (EM) algorithm for modeling the input data. A corresponding result is shown for Hebbian learning in artificial neural networks. Subject Area: Neuroscience

Spotlight presentation, Tuesday, 5:10.

T70 A Neural Implementation of the Kalman Filter

ROBERT WILSON Princeton University LEIF FINKEL University of Pennsylvania rcw2@princeton.edu leif@neuroengineering.upenn.edu

There is a growing body of experimental evidence to suggest that the brain is capable of approximating optimal Bayesian inference in the face of noisy input stimuli. Despite this progress, the neural underpinnings of this computation are still poorly understood. In this paper we focus on the problem of Bayesian filtering of stochastic time series. In particular we introduce a novel neural network, derived from a line attractor architecture, whose dynamics map directly onto those of the Kalman Filter in the limit where the prediction error is small. When the prediction error is large we show that the network responds robustly to change-points in a way that is qualitatively compatible with the optimal Bayesian model. The model suggests ways in which probability distributions are encoded in the brain and makes a number of testable experimental predictions. **Subject Area:** Neuroscience

T71 Speeding up Magnetic Resonance Image Acquisition by Bayesian Multi-Slice Adaptive Compressed Sensing

MATTHIAS SEEGER mseeger@mmci.uni-saarland.de Saarland University and Max Planck Institute for Informatics

We show how to sequentially optimize magnetic resonance imaging measurement designs over stacks of neighbouring image slices, by performing convex variational inference on a large scale non-Gaussian linear dynamical system, tracking dominating directions of posterior covariance without imposing any factorization constraints. Our approach can be scaled up to high-resolution images by reductions to numerical mathematics primitives and parallelization on several levels. In a first study, designs are found that improve significantly on others chosen independently for each slice or drawn at random. **Subject Area:** Neuroscience

T72 Subject independent EEG-based BCI decoding

Siamac Fazli	fazli@cs.tu-berlin.de
Technical University Berlin	
Cristian Grozea	cristian.grozea@first.fraunhofer.de
Fraunhofer Institute FIRST	
MARTON DANOCZY	marton@cs.tu-berlin.de
TU Berlin	
Benjamin Blankertz	blanker@cs.tu-berlin.de
Berlin Institute of Technology	
FLORIN POPESCU	florin.popescu@first.fraunhofer.de
Fraunhofer FIRST	
Klaus-Robert Muller	krm@cs.tu-berlin.de
Berlin Institute of Technology	

In the quest to make Brain Computer Interfacing (BCI) more usable, dry electrodes have emerged that get rid of the initial 30 minutes required for placing an electrode cap. Another time consuming step is the required individualized adaptation to the BCI user, which involves another 30 minutes calibration for assessing a subject's brain signature. In this paper we aim to also remove this calibration proceedure from BCI setup time by means of machine learning. In particular, we harvest a large database of EEG BCI motor imagination recordings (83 subjects) for constructing a library of subject-specific spatio-temporal filters and derive a subject independent BCI classifier. Our offline results indicate that BCI-naïve users could start real-time BCI use with no prior calibration at only a very moderate performance loss. **Subject Area:** Neuroscience

Spotlight presentation, Tuesday, 5:10.

J	Analysis of SVM	with Indefinite Remeis
	YIMING YING	mathying@gmail.com
	Colin Campbell	C.Campbell@bris.ac.uk
	University of Bristol	
	Mark Girolami	girolami@dcs.gla.ac.uk
	University of Glasgow	

T73 Analysis of SVM with Indefinite Kernels

The recent introduction of indefinite SVM by Luss and d'Aspremont [15] has effectively demonstrated SVM classification with a non-positive semi-definite kernel (indefinite kernel). This paper studies the properties of the objective function introduced there. In particular, we show that the objective function is continuously differentiable and its gradient can be explicitly computed. Indeed, we further show that its gradient is Lipschitz continuous. The main idea behind our analysis is that the objective function is smoothed by the penalty term, in its saddle (min-max) representation, measuring the distance between the indefinite kernel matrix and the proxy positive semi-definite one. Our elementary result greatly facilitates the application of gradient-based algorithms. Based on our analysis, we further develop Nesterov's smooth optimization approach [16,17] for indefinite SVM which has an optimal convergence rate for smooth problems. Experiments on various benchmark datasets validate our analysis and demonstrate the efficiency of our proposed algorithms.

Subject Area: Supervised Learning

Spotlight presentation, Tuesday, 3:20.

T74 Boosting with Spatial Regularization

Zhen Xiang	zxiang@princeton.edu
Princeton University	
Yongxin Xi	yxi@princeton.edu
Uri Hasson	hasson@princeton.edu
Peter Ramadge	ramadge@princeton.edu
Princeton University	

By adding a spatial regularization kernel to a standard loss function formulation of the boosting problem, we develop a framework for spatially informed boosting. From this regularized loss framework we derive an efficient boosting algorithm that uses additional weights/priors on the base classifiers. We prove that the proposed algorithm exhibits a "grouping effect", which encourages the selection of all spatially local, discriminative base classifiers. The algorithm's primary advantage is in applications where the trained classifier is used to identify the spatial pattern of discriminative information, e.g. the voxel selection problem in fMRI. We demonstrate the algorithm's performance on various data sets.

Subject Area: Supervised Learning

Spotlight presentation, Tuesday, 3:20.

T75 Directed Regression

YI-HAO KAO BENJAMIN VAN ROY XIANG YAN Stanford University yihaokao@stanford.edu bvr@stanford.edu xyan@stanford.edu

When used to guide decisions, linear regression analysis typically involves estimation of regression coefficients via ordinary least squares and their subsequent use to make decisions. When there are multiple response variables and features do not perfectly capture their relationships, it is beneficial to account for the decision objective when computing regression coefficients. Empirical optimization does so but sacrifices performance when features are well-chosen or training data are insufficient. We propose directed regression, an efficient algorithm that combines merits of ordinary least squares and empirical optimization. We demonstrate through a computational study that directed regression can generate significant performance gains over either alternative. We also develop a theory that motivates the algorithm.

Subject Area: Supervised Learning

T76 Measuring Invariances in Deep Networks

IAN GOODFELLOW	ia3n@cs.stanford.edu
Quoc Le	quocle@stanford.edu
Andrew Saxe	asaxe@stanford.edu
Andrew Ng	ang@cs.stanford.edu
Stanford University	

For many computer vision applications, the ideal image feature would be invariant to multiple confounding image properties, such as illumination and viewing angle. Recently, deep architectures trained in an unsupervised manner have been proposed as an automatic method for extracting useful features. However, outside of using these learning algorithms in a classifier, they can be sometimes difficult to evaluate. In this paper, we propose a number of empirical tests that directly measure the degree to which these learned features are invariant to different image transforms. We find that deep autoencoders become invariant to increasingly complex image transformations with depth. This further justifies the use of "deep" vs. "shallower" representations. Our performance metrics agree with existing measures of invariance. Our evaluation metrics can also be used to evaluate future work in unsupervised deep learning, and thus help the development of future algorithms. **Subject Area:** Supervised Learning

T77 Nonparametric Greedy Algorithms for the Sparse Learning Problem

Han Liu	hanliu@cs.cmu.edu
XI CHEN	xichen@cs.cmu.edu
Carnegie Mellon University	

This paper studies the forward greedy strategy in sparse nonparametric regression. For additive models, we propose an algorithm called additive forward regression; for general multivariate regression, we propose an algorithm called generalized forward regression. Both of them simultaneously conduct estimation and variable selection in nonparametric settings for the high dimensional sparse learning problem. Our main emphasis is empirical: on both simulated and real data, these two simple greedy methods can clearly outperform several state-of-the-art competitors, including the LASSO, a nonparametric version of the LASSO called the sparse additive model (SpAM) and a recently proposed adaptive parametric forward-backward algorithm called the Foba. Some theoretical justifications are also provided.

Subject Area: Supervised Learning

T78 Learning Non-Linear Combinations of Kernels

CORINNA CORTES corinna@google.com Google Research MEHRYAR MOHRI mohri@cs.nyu.edu Courant Institute of Mathematical Sciences and Google Research AFSHIN ROSTAMIZADEH rostami@cs.nyu.edu Courant Institute, New York University

This paper studies the general problem of learning kernels based on a polynomial combination of base kernels. It analyzes this problem in the case of regression and the kernel ridge regression algorithm. It examines the corresponding learning kernel optimization problem, shows how that minimax problem can be reduced to a simpler minimization problem, and proves that the global solution of this problem always lies on the boundary. We give a projection-based gradient descent algorithm for solving the optimization problem, shown empirically to converge in few iterations. Finally, we report the results of extensive experiments with this algorithm using several publicly available datasets demonstrating the effectiveness of our technique. **Subject Area:** Supervised Learning

T79 3D Object Recognition with Deep Belief Nets

Vinod Nair	vnair@cs.toronto.edu
Geoffrey Hinton	hinton@cs.toronto.edu
University of Toronto	

We introduce a new type of Deep Belief Net and evaluate it on a 3D object recognition task. The top-level model is a third-order Boltzmann machine, trained using a hybrid algorithm that combines both generative and discriminative gradients. Performance is evaluated on the NORB database(normalized-uniform version), which contains stereo-pair images of objects under different lighting conditions and viewpoints. Our model achieves 6.5% error on the test set, which is close to the best published result for NORB (5.9%) using a convolutional neural net that has built-in knowledge of translation invariance. It substantially outperforms shallow models such as SVMs (11.6%). DBNs are especially suited for semi-supervised learning, and to demonstrate this we consider a modified version of the NORB recognition task in which additional unlabeled images are created by applying small translations to the images in the database. With the extra unlabeled data (and the same amount of labeled data as before), our model achieves 5.2% error, making it the
current best result for NORB. **Subject Area:** Supervised Learning

Spotlight presentation, Tuesday, 5:10.

T80 Grouped Orthogonal Matching Pursuit for Variable Selection and Prediction

AURELIE LOZANOaclozano@us.ibm.comIBM T.J. Watson Researchswirszcz@us.ibm.comGRZEGORZ SWIRSZCZswirszcz@us.ibm.comIBM Researchnabe@us.ibm.comNAOKI ABEnabe@us.ibm.comIBM T. J. Watson Research Center

We consider the problem of variable group selection for least squares regression, namely, that of selecting groups of variables for best regression performance, leveraging and adhering to a natural grouping structure within the explanatory variables. We show that this problem can be efficiently addressed by using a certain greedy style algorithm. More precisely, we propose the Group Orthogonal Matching Pursuit algorithm (Group-OMP), which extends the standard OMP procedure (also referred to as "forward greedy feature selection algorithm" for least squares regression) to perform stage-wise group variable selection. We prove that under certain conditions Group-OMP can identify the correct (groups of) variables. We also provide an upperbound on the l_{∞} norm of the difference between the estimated regression coefficients and the true coefficients. Experimental results on simulated and real world datasets indicate that Group-OMP compares favorably to Group Lasso, OMP and Lasso, both in terms of variable selection and prediction accuracy.

Subject Area: Supervised Learning

T81 Efficient and Accurate Lp-Norm Multiple Kernel Learning

Marius Kloft	mkloft@cs.tu-berlin.de
Technical University of Berlin	
Ulf Brefeld	brefeld@cs.tu-berlin.de
TU Berlin	
Soeren Sonnenburg	Soeren.Sonnenburg@tuebingen.mpg.de
Friedrich Miescher Laboratory, M	ax Planck Society
Pavel Laskov	pavel.laskov@uni-tuebingen.de
University of Tuebingen	
Klaus-Robert Müller	klaus-robert.mueller@tu-berlin.de
Technical University of Berlin	
Alexander Zien	Alexander.Zien@lifebiosystems.com
LIFE Biosystems	

Learning linear combinations of multiple kernels is an appealing strategy when the right choice of features is unknown. Previous approaches to multiple kernel learning (MKL) promote sparse kernel combinations and hence support interpretability. Unfortunately, L1-norm MKL is hardly observed to outperform trivial baselines in practical applications. To allow for robust kernel mixtures, we generalize MKL to arbitrary Lp-norms. We devise new insights on the connection between several existing MKL formulations and develop two efficient interleaved optimization strategies for arbitrary p.1. Empirically, we demonstrate that the interleaved optimization strategies are much faster compared to the traditionally used wrapper approaches. Finally, we apply Lp-norm MKL to real-world problems from computational biology, showing that non-sparse MKL achieves accuracies that go beyond the state-of-the-art.

Subject Area: Supervised Learning

Training Factor Graphs with Reinforcement T82Learning for Efficient MAP Inference

MICHAEL WICK	mwick@cs.umass.edu
Khashayar Rohanimanesh	khash@cs.umass.edu
SAMEER SINGH	sameer@cs.umass.edu
Andrew McCallum	mccallum@cs.umass.edu
University of Massachusetts, Amherst	

Large, relational factor graphs with structure defined by first-order logic or other languages give rise to notoriously difficult inference problems. Because unrolling the structure necessary to represent distributions over all hypotheses has exponential blow-up, solutions are often derived from MCMC. However, because of limitations in the design and parameterization of the jump function, these sampling-based methods suffer from local minima—the system must transition through lower-scoring configurations before arriving at a better MAP solution. This paper presents a new method of explicitly selecting fruitful downward jumps by leveraging reinforcement learning (RL). Rather than setting parameters to maximize the likelihood of the training data, parameters of the factor graph are treated as a log-linear function approximator and learned with temporal difference (TD); MAP inference is performed by executing the resulting policy on held out test data. Our method allows efficient gradient updates since only factors in the neighborhood of variables affected by an action need to be computed—we bypass the need to compute marginals entirely. Our method provides dramatic empirical success, producing new state-of-the-art results on a complex joint model of ontology alignment, with a 48% reduction in error over state-of-the-art in that domain. Subject Area: Supervised Learning

Spotlight presentation, Tuesday, 5:10.

T83 Fast, smooth and adaptive regression in metric spaces SAMORY KPOTUFE

UCSD CSE

skpotufe@cs.ucsd.edu

It was recently shown that certain nonparametric regressors can escape the curse of dimensionality in the sense that their convergence rates adapt to the intrinsic dimension of data ([BL:65, SK:77]). We prove some stronger results in more general settings. In particular, we consider a regressor which, by combining aspects of both tree-based regression and kernel regression, operates on a general metric space, yields a smooth function, and evaluates in time $O(\log n)$. We derive a tight convergence rate of the form $n^{-2/(2+d)}$ where d is the Assound dimension of the input space. Subject Area: Supervised Learning

T84 Robust Value Function Approximation Using Bilinear Programming

 MAREK PETRIK
 petrik@cs.umass.edu

 University of Massachusetts Amherst
 Shlomo@cs.umass.edu

 University of Massachusetts at Amherst
 shlomo@cs.umass.edu

Existing value function approximation methods have been successfully used in many applications, but they often lack useful a priori error bounds. We propose approximate bilinear programming, a new formulation of value function approximation that provides strong a priori guarantees. In particular, it provably finds an approximate value function that minimizes the Bellman residual. Solving a bilinear program optimally is NP hard, but this is unavoidable because the Bellman-residual minimization itself is NP hard. We, therefore, employ and analyze a common approximate algorithm for bilinear programs. The analysis shows that this algorithm offers a convergent generalization of approximate policy iteration. Finally, we demonstrate that the proposed approach can consistently minimize the Bellman residual on a simple benchmark problem. **Subject Area:** Control and Reinforcement Learning

Spotlight presentation, Tuesday, 3:20.

T85 The Infinite Partially Observable Markov Decision Process

FINALE DOSHI-VELEZ Cambridge University, MIT finale@mit.edu

The Partially Observable Markov Decision Process (POMDP) framework has proven useful in planning domains that require balancing actions that increase an agent's knowledge and actions that increase an agent's reward. Unfortunately, most POMDPs are complex structures with a large number of parameters. In many realworld problems, both the structure and the parameters are difficult to specify from domain knowledge alone. Recent work in Bayesian reinforcement learning has made headway in learning POMDP models; however, this work has largely focused on learning the parameters of the POMDP model. We define an infinite POMDP (iPOMDP) model that does not require knowledge of the size of the state space; instead, it assumes that the number of visited states will grow as the agent explores its world and explicitly models only visited states. We demonstrate the iPOMDP utility on several standard problems.

Subject Area: Control and Reinforcement Learning

T86 A Smoothed Approximate Linear Program

Vijay Desai	vvd2101@columbia.edu
Columbia University	
VIVEK FARIAS	vivekf@mit.edu
MIT Sloan	
CIAMAC MOALLEMI	ciamac@gsb.columbia.edu
Columbia University	

We present a novel linear program for the approximation of the dynamic programming cost-to-go function in high-dimensional stochastic control problems. LP approaches to approximate DP naturally restrict attention to approximations that are lower bounds to the optimal cost-to-go function. Our program – the 'smoothed approximate linear program' – relaxes this restriction in an appropriate fashion while remaining computationally tractable. Doing so appears to have several advantages: First, we demonstrate superior bounds on the quality of approximation to the optimal cost-to-go function afforded by our approach. Second, experiments with our approach on a challenging problem (the game of Tetris) show that the approach outperforms the existing LP approach (which has previously been shown to be competitive with several ADP algorithms) by an order of magnitude.

Subject Area: Control and Reinforcement Learning

Spotlight presentation, Tuesday, 5:10.

T87 Learning to Explore and Exploit in POMDPs

Chenghui Cai	cc88@duke.edu
Xuejun Liao	xjliao@ee.duke.edu
Duke University	
LAWRENCE CARIN	lcarin@ee.duke.edu
Duke ECE	

A fundamental objective in reinforcement learning is the maintenance of a proper balance between exploration and exploitation. This problem becomes more challenging when the agent can only partially observe the states of its environment. In this paper we propose a dual-policy method for jointly learning the agent behavior and the balance between exploration exploitation, in partially observable environments. The method subsumes traditional exploration, in which the agent takes actions to gather information about the environment, and active learning, in which the agent queries an oracle for optimal actions (with an associated cost for employing the oracle). The form of the employed exploration is dictated by the specific problem. Theoretical guarantees are provided concerning the optimality of the balancing of exploration and exploitation. The effectiveness of the method is demonstrated by experimental results on benchmark problems.

Subject Area: Control and Reinforcement Learning

T88 Compositionality of optimal control laws

Emanuel Todorov UW

todorov@cs.washington.edu

We present a theory of compositionality in stochastic optimal control, showing how task-optimal controllers can be constructed from certain primitives. The primitives are themselves feedback controllers pursuing their own agendas. They are mixed in proportion to how much progress they are making towards their agendas and how compatible their agendas are with the present task. The resulting composite control law is provably optimal when the problem belongs to a certain class. This class is rather general and yet has a number of unique properties - one of which is that the Bellman equation can be made linear even for non-linear or discrete dynamics. This gives rise to the compositionality developed here. In the special case of linear dynamics and Gaussian noise our framework yields analytical solutions (i.e. non-linear mixtures of linear-quadratic regulators) without requiring the final cost to be quadratic. More generally, a natural set of control primitives can be constructed by applying SVD to Green's function of the Bellman equation. We illustrate the theory in the context of human arm movements. The ideas of optimality and compositionality are both very prominent in the field of motor control, yet they are hard to reconcile. Our work makes this possible.

Subject Area: Control and Reinforcement Learning

T89 Constructing Topological Maps using Markov Random Fields and Loop-Closure Detection

Roy Anati	royanati@cis.upenn.edu
Kostas Daniilidis	kostas@cis.upenn.edu
University of Pennsylvania	

We present a system which constructs a topological map of an environment given a sequence of images. This system includes a novel image similarity score which uses dynamic programming to match images using both the appearance and relative positions of local features simultaneously. Additionally an MRF is constructed to model the probability of loop-closures. A locally optimal labeling is found using Loopy-BP. Finally we outline a method to generate a topological map from loop closure data. Results are presented on four urban sequences and one indoor sequence. **Subject Area:** Robotics

Demonstrations (7:30–11:59pm)

D1 Activity Recognition and Human Gait Recognition

JORDAN FRANK McGill University jordan.frank@cs.mcgill.ca

The technology to be demonstrated is an activity recognition system that runs on a standard consumer mobile phone (HTC Dream). The phone is on a belt clip attached to the waist of an individual, and is able to discern whether the person is walking, running, biking, walking up or down stairs, or standing still. Additionally, the device can be trained to detect a user from their walking gait. We will allow a number of users to provide training data, and then we will demonstrate the ability of the device to detect which person is carrying it from their gait.

D2 Automatically Adaptable Spike Sorting Software

JORGE GAMEZ DE LEON jorge.gamez@gmail.com U.N.A.M.

Numerous methods have been developed to detect and sort single unit neuron activity from extracellular recordings. Because of the high variability in the neuron waveform shapes, equipment and recording settings, there is no single best performance algorithm. In recent years, several wavelet based detection algorithms have been proposed as methods for detecting spikes in extracellular recordings. The characteristics of the wavelet family and scale used play an essential role on the detection performance. On previous works the wavelet selection is done by visual inspection of the wavelet waveform. We have created a Matlab based software library that can characterize a wavelet';s ability to detect events on a signal, it will generate Receiver Operating Characteristic curves (ROC) and the Area Under the Curve (AUC) as comparison measurement information. This can be used to find the best detecting wavelet for certain data. Our software library also includes a spike sorting algorithm composed of a wavelet based spike detection step and a t-distribution expectation maximization clustering.

D3 Car Detector trained with the Quantum Adiabatic Algorithm

HARTMUT NEVEN neven@google.com Google

We developed a novel method to train a binary classifier as a thresholded sum of weak classifiers. The training proceeds by solving a non-convex optimization over binary variables that strives to minimize the training error as well as the number of weak classifiers used. Such cost functions have not been studied before. Since this is formally a NP hard optimization problem it can not be solved efficiently with classical solvers. Therefore we employ a newly developed quantum processor that implements the quantum adiabatic algorithm to find good solutions to hard binary problems. Once the detector is trained this way it can run on classical hardware. We have been able to show on several datasets that the new approach outperforms the state of the art method AdaBoost.

D4 Open-access datasets for time series causality discovery validation

ISABELLE GUYON guyon@clopinet.com Clopinet

The CausalityWorkbench project provides an environment to test causal discovery algorithms. Via a web portal (http://clopinet.com/causality), we provide a number of resources, including a repository of datasets, models, and software packages, and a virtual laboratory allowing users to benchmark causal discovery algorithms by performing virtual experiments to study artificial causal systems. We regularly organize competitions. Our repository already includes several time dependent datasets from a variety of domains: system biology, neurosciences, physics, manufacturing, and marketing. We will invite new contributions and present our plan for upcoming evaluations of causal models for time series applications.

D5 Demonstration of the TreeRank Software

MARINE DEPECKER marine.depecker@telecom-paristech.fr Renault SA-Telecom ParisTech

The TreeRank software permits to build tree-based ranking rules from data with binary labels according to recently developed learning algorithms (see the list of references below). It is one of the sole software that performs ROC optimization. It also allows for reading ranking trees (in regards to prediction), plotting the corresponding tree graphics and visualizing their statistical performance on training/test data samples (in the ROC or PR spaces) and outputs a list of graphics which facilitate much interpretation of the prediction rules produced. TreeRank also performs model selection, automatically or at hand, through an easy-to-use graphic interface.

D6 The IID: A Natively Probabilistic Reconfigurable Computer

VIKASH MANSINGHKA vkm@mit.edu Massachusetts Institute of Technology

We demonstrate the IID, a natively probabilistic, reconfigurable digital computer based on stochastic digital circuits. The IID Mark 0 is implemented on top of a commodity array of Field Programmable Gate Arrays (FPGAs), and programmed using a generic toolchain based on the State-Density-Kernel abstractions from Blaise. It can be used to perform MCMC inference on factor graphs with hundreds of thousands of variables in real time. The technical novelty of the IID rests in the stochastic digital circuits which it simulates. These circuits enable the construction of massively parallel, low bit precision, fault tolerant machines that directly simulate Markov chains in hardware, leveraging small XOR-SHIFT RNGs to provide the necessary stochasticity. Due to space constraints, we refer the reader to FIXME cite for technical details about these circuits, including a detailed discussion of their capabilities and their novelty and an overview of the software tools used to program the IID to solve arbitrary discrete factor graphs. The key consequence of our approach is that the IID can be used to solve probabilistic graphical model problems with 2-3 orders of magnitude improvement in price/power/performance product and 3-6 orders of magnitude improvements in robustness to bit errors than is possible using conventional computer architectures.

D7 Vision Toolkit for Image Recognition

JAMIE NIEMASIK Numenta jniemasik@numenta.com

The Vision Toolkit is an easy-to-use graphical application that lets anyone train their own vision system, without knowledge of the underlying algorithms. The user drags in images (e.g. their own dataset or images from the web), organizes them into categories, and then the Toolkit trains an HTM (Hierarchical Temporal Memory) to recognize those categories. The user can experiment with their trained network by bringing in novel images, and they can upload it to our webserver with one click. Once on the webserver, they can use our web interface to recognize images from anywhere. They can even send images for recognition from any device using a simple API (e.g. an iPhone). Hierarchical Temporal Memory is Numenta's technology and is based on a theory of the neocortex. It comprises a deep hierarchy of nodes, each of which learns variable-order temporal sequences of spatial patterns. Inference is performed using Bayesian belief propagation. It also supports top-down attention, similarity search, prediction, and other useful features that we can demonstrate. The Vision Toolkit uses NuPIC, our platform for HTM development. The full NuPIC release allows complete customization of the HTM parameters, scales from single CPUs to clusters, and supports many sensory modalities.

Wednesday, December 9th

Oral Session (8:30–10:10am): "Neuroscience"

8:30am Invited Talk: The Rat Vibrissal Array as a Model Sensorimotor System

MITRA HARTMANN m-hartmann@northwestern.edu Northwestern University

The rat whisker (vibrissal) system has neural processing pathways that are analogous to human tactile pathways through the spinal cord. Rats use rhythmic (5-25 Hz) movements of their whiskers to tactually extract object features, including size, shape, orientation, and texture. In this talk, I will describe our efforts to characterize whisker mechanical properties and to determine the set of mechanical variables that could potentially be sensed by the rat. We have also developed a hardware array of artificial whiskers that makes use of these variables to successfully extract 3-dimensional object features. On the behavioral side, we have developed a laser light sheet to visualize whisker-object contact patterns during natural exploratory behavior and have used this technique to examine the relative roles of head and whisker movements in generating patterns of mechanosensory input across the array during natural exploratory sequences. Finally, we are developing a simulation environment that integrates our knowledge of whisker mechanics with our knowledge of head and whisker motion to enable the user to model the rat's interactions with various objects in the environment. Our goal is to predict the contact patterns –; and resulting forces and moments -; at each whisker base for a given exploratory sequence, and then to predict the resulting responses of trigeminal ganglion neurons.

Prof. Mitra Hartmann received a Bachelor of Science in Applied and Engineering Physics from Cornell University, and a PhD in Integrative Neuroscience from the California Institute of Technology. She was a postdoctoral scholar at the Jet Propulsion Laboratory in Pasadena, California in the Bio-Inspired Technology and Systems group, and joined the faculty at Northwestern University in 2003. Prof. Hartmann is presently an Associate Professor at Northwestern with a 50-50 joint appointment between the departments of Biomedical Engineering and Mechanical Engineering.

9:30am Functional network reorganization in motor cortex can be explained by reward-modulated Hebbian learning

Robert Legenstein	robert.legenstein@igi.tugraz.at
Graz University of Technology	
Steven Chase	schase@andrew.cmu.edu
Carnegie Mellon University	
Andrew Schwartz	abs21@pitt.edu
University of Pittsburgh	
Wolfgang Maass	maass@igi.tugraz.at
Graz University of Technology	

The control of neuroprosthetic devices from the activity of motor cortex neurons benefits from learning effects where the function of these neurons is adapted to the control task. It was recently shown that tuning properties of neurons in monkey motor cortex are adapted selectively in order to compensate for an erroneous interpretation of their activity. In particular, it was shown that the tuning curves of those neurons whose preferred directions had been misinterpreted changed more than those of other neurons. In this article, we show that the experimentally observed self-tuning properties of the system can be explained on the basis of a simple learning rule. This learning rule utilizes neuronal noise for exploration and performs Hebbian weight updates that are modulated by a global reward signal. In contrast to most previously proposed reward-modulated Hebbian learning rules, this rule does not require extraneous knowledge about what is noise and what is signal. The learning rule is able to optimize the performance of the model system within biologically realistic periods of time and under high noise levels. When the neuronal noise is fitted to experimental data, the model produces learning effects similar to those found in monkey experiments.

Subject Area: Neuroscience

9:50am Know Thy Neighbour: A Normative Theory of Synaptic Depression

JEAN-PASCAL PFISTER	jean-pascal.pfister@eng.cam.ac.uk
University of Cambridge	
Peter Dayan	dayan@gatsby.ucl.ac.uk
University College London	
Mate Lengyel	m.lengyel@eng.cam.ac.uk
University of Cambridge	

Synapses exhibit an extraordinary degree of short-term malleability, with release probabilities and effective synaptic strengths changing markedly over multiple timescales. From the perspective of a fixed computational operation in a network, this seems like a most unacceptable degree of added noise. We suggest an alternative theory according to which short term synaptic plasticity plays a normatively-justifiable role. This theory starts from the commonplace observation that the spiking of a neuron is an incomplete, digital, report of the analog quantity that contains all the critical information, namely its membrane potential. We suggest that one key task for a synapse is to solve the inverse problem of estimating the pre-synaptic membrane potential from the spikes it receives and prior expectations, as in a recursive filter. We show that short-term synaptic depression has canonical dynamics which closely resemble those required for optimal estimation, and that it indeed supports high quality estimation. Under this account, the local postsynaptic potential and the level of synaptic resources track the (scaled) mean and variance of the estimated presynaptic membrane potential. We make experimentally testable predictions for how the statistics of subthreshold membrane potential fluctuations and the form of spiking non-linearity should be related to the properties of short-term plasticity in any particular cell type.

Oral Session (10:40am–12:00pm): "Theory, Optimization and Games"

10:40am Efficient Learning using Forward-Backward Splitting

John Duchi UC Berkeley Yoram Singer Google jduchi@cs.berkeley.edu singer@google.com

We describe, analyze, and experiment with a new framework for empirical loss minimization with regularization. Our algorithmic framework alternates between two phases. On each iteration we first perform an *unconstrained* gradient descent step. We then cast and solve an instantaneous optimization problem that trades off minimization of a regularization term while keeping close proximity to the result of the first phase. This yields a simple yet effective algorithm for both batch penalized risk minimization and online learning. Furthermore, the two phase approach enables sparse solutions when used in conjunction with regularization functions that promote sparsity, such as ℓ_1 . We derive concrete and very simple algorithms for minimization of loss functions with ℓ_1, ℓ_2, ℓ_2^2 , and ℓ_{∞} regularization. We also show how to construct efficient algorithms for mixed-norm ℓ_1/ℓ_q regularization. We further extend the algorithms and give efficient implementations for very high-dimensional data with sparsity. We demonstrate the potential of the proposed framework in experiments with synthetic and natural datasets. **Subject Area:** Optimization

11:00am On Stochastic and Worst-case Models for Investing

ELAD HAZAN IBM SATYEN KALE Yahoo! Research ehazan@cs.princeton.edu skale@yahoo-inc.com

In practice, most investing is done assuming a probabilistic model of stock price returns known as the Geometric Brownian Motion (GBM). While it is often an acceptable approximation, the GBM model is not always valid empirically. This motivates a worst-case approach to investing, called universal portfolio management, where the objective is to maximize wealth relative to the wealth earned by the best fixed portfolio in hindsight. In this paper we tie the two approaches, and design an investment strategy which is universal in the worst-case, and yet capable of exploiting the mostly valid GBM model. Our method is based on new and improved regret bounds for online convex optimization with exp-concave loss functions.

Subject Area: Theory

11:20am Bootstrapping from Game Tree Search

Joel Veness	Joel.Veness@nicta.com.au
UNSW / NICTA	
David Silver	silver@cs.ualberta.ca
University of Alberta	
William Uther	William.Uther@nicta.com.au
NICTA / University of New South	Wales
Alan Blair	Alan.Blair@nicta.com.au
NICTA / UNSW	

In this paper we introduce a new algorithm for updating the parameters of a heuristic evaluation function, by updating the heuristic towards the values computed by an alphabeta search. Our algorithm differs from previous approaches to learning from search, such as Samuel's checkers player and the TD-Leaf algorithm, in two key ways. First, we update all nodes in the search tree, rather than a single node. Second, we use the outcome of a deep search, instead of the outcome of a subsequent search, as the training signal for the evaluation function. We implemented our algorithm in a chess program Meep, using a linear heuristic function. After initialising its weight vector to small random values, Meep was able to learn high quality weights from self-play alone. When tested online against human opponents, Meep played at a master level, the best performance of any chess program with a heuristic learned entirely from self-play. **Subject Area:** Control and Reinforcement Learning

11:40am Kernel Choice and Classifiability for RKHS Embeddings of Probability Distributions

Bharath Sriperumbudur	bharathsv@ucsd.edu
UC San Diego	
Kenji Fukumizu	fukumizu@ism.ac.jp
The Institute of Statistical Math	hematics
Arthur Gretton	arthur@tuebingen.mpg.de
Carnegie Mellon University and	Max Planck Institute
Gert Lanckriet	gert@ece.ucsd.edu
University of California	
Bernhard Schoelkopf	bernhard.schoelkopf@tuebingen.mpg.de
MPI for Biological Cybernetics	

Embeddings of probability measures into reproducing kernel Hilbert spaces have been proposed as a straightforward and practical means of representing and comparing probabilities. In particular, the distance between embeddings (the maximum mean discrepancy, or MMD) has several key advantages over many classical metrics on distributions, namely easy computability, fast convergence and low bias of finite sample estimates. An important requirement of the embedding RKHS is that it be characteristic: in this case, the MMD between two distributions is zero if and only if the distributions coincide. Three new results on the MMD are introduced in the present study. First, it is established that MMD corresponds to the optimal risk of a kernel classifier, thus forming a natural link between the distance between distributions and their ease of classification. An important consequence is that a kernel must be characteristic to guarantee classifiability between distributions in the RKHS. Second, the class of characteristic kernels is broadened to incorporate all bounded, continuous strictly positive definite kernels: these include non-translation invariant kernels and kernels on non-compact domains. Third, a generalization of the MMD is proposed for families of kernels, as the supremum over MMDs on a class of kernels (for instance the Gaussian kernels with different bandwidths). This extension is necessary to obtain a single distance measure if a large selection or class of characteristic kernels is potentially appropriate. This generalization is reasonable, given that it corresponds to the problem of learning the kernel by minimizing the risk of the corresponding kernel classifier. The generalized MMD is shown to have consistent finite sample estimates, and its performance is demonstrated on a homogeneity testing example. **Subject Area:** Theory

Oral Session (2:00–3:20pm): "Vision and Inference"

2:00pm Invited Talk: Learning and Inference in Low-Level Vision

YAIR WEISS yweiss@cs.huji.ac.il Hebrew University

Low level vision addresses the issues of labeling and organizing image pixels according to scene related properties - such as motion, contrast, depth and reflectance. I will describe our attempts to understand low-level vision in humans and machines as optimal inference given the statistics of the world. If time permits, I will discuss my favorite NIPS rejected papers.

Yair Weiss is an Associate Professor at the Hebrew University School of Computer Science and Engineering. He received his Ph.D. from MIT working with Ted Adelson on motion analysis and did postdoctoral work at UC Berkeley. Since 2005 he has been a fellow of the Canadian Institute for Advanced Research. With his students and colleagues he has co-authored award winning papers in NIPS (2002), ECCV (2006), UAI (2008) and CVPR (2009).

3:00pm An LP View of the M-best MAP problem

Menachem Fromer	fromer@cs.huji.ac.il
Amir Globerson	gamir@cs.huji.ac.il
Hebrew University	

We consider the problem of finding the M assignments with maximum probability in a probabilistic graphical model. We show how this problem can be formulated as a linear program (LP) on a particular polytope. We prove that, for tree graphs (and junction trees in general), this polytope has a particularly simple form and differs from the marginal polytope in a single inequality constraint. We use this characterization to provide an approximation scheme for non-tree graphs, by using the set of spanning trees over such graphs. The method we present puts the M-best inference problem in the context of LP relaxations, which have recently received considerable attention and have proven useful in solving difficult inference problems. We show empirically that our method often finds the provably exact M best configurations for problems of high tree width. **Subject Area:** Probabilistic Models and Methods

Spotlights (3:20–3:40pm)

- Quantification and the language of thought CHARLES KEMP, Carnegie Mellon University. Subject Area: Cognitive Science See abstract, page 176.
- Fast Image Deconvolution using Hyper-Laplacian Priors DILIP KRISHNAN and ROB FERGUS, New York University. Subject Area: Vision See abstract, page 170.
- Efficient Large-Scale Distributed Training of Conditional Maximum Entropy Models GIDEON MANN, RYAN MCDONALD, Google, MEHRYAR MOHRI, Google Research, NATHAN SILBERMAN, Google Inc., and DAN WALKER, New York University. Subject Area: Supervised Learning See abstract, page 163.
- A Data-Driven Approach to Modeling Choice VIVEK FARIAS, MIT Sloan, SRIKANTH JAGABATHULA and DEVAVRAT SHAH, MIT. Subject Area: Optimization See abstract, page 168.
- Statistical Analysis of Semi-Supervised Learning: The Limit of Infinite Unlabelled Data BOAZ NADLER, Weizmann Institute of Science, NATHAN SREBRO, Toyota Technological Institute at Chicago, and XUEYUAN ZHOU, University of Chicago. Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 185.
- Region-based Segmentation and Object Detection STEPHEN GOULD, TIANSHI GAO and DAPHNE KOLLER, Stanford University. Subject Area: Vision See abstract, page 171.
- Anomaly Detection with Score functions based on Nearest Neighbor Graphs MANQI ZHAO and VENKATESH SALIGRAMA, Boston University.
 Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 186.
- Adaptive Regularization of Weight Vectors
 KOBY CRAMMER, The Technion, ALEX KULESZA, University of Pennsylvania, and MARK DREDZE, Johns Hopkins University.

 Subject Area: Supervised Learning
 See abstract, page 165.

WEDNESDAY, SPOTLIGHTS

- Adaptive Regularization for Transductive Support Vector Machine ZENGLIN XU, CUHK, RONG JIN, Michigan State University, JIANKE ZHU, ETH Zurich, IRWIN KING, CUHK, MICHAEL LYU, The Chinese University of HK, and ZHIRONG YANG, CUHK,TKK.
 Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 186.
- Variational Gaussian-process factor analysis for modeling spatio-temporal data JAAKKO LUTTINEN and ALEXANDER ILIN, Helsinki University of Technology. Subject Area: Probabilistic Models and Methods See abstract, page 193.

 Predicting the Optimal Spacing of Study: A Multiscale Context Model of Memory MICHAEL MOZER, University of Colorado at Boulder, HAROLD PASHLER, UC San Diego, NICHOLAS CEPEDA, York University, ROBERT LINDSEY, University of Colorado, and ED VUL, Massachusetts Institute of Technology.
 Subject Area: Cognitive Science See abstract, page 177.

- Fast Graph Laplacian Regularized Kernel Learning via Semidefinite-Quadratic-Linear Programming XIAO-MING WU, ANTHONY MAN-CHO SO, ZHENGUO LI and SHUO-YEN ROBERT LI, The Chinese University of Hong Kong.
 Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 187.
- Learning Brain Connectivity of Alzheimer's Disease from Neuroimaging Data SHUAI HUANG, Arizona State University, JING LI, jing.li.8@asu.edu, LIANG SUN, JUN LIU, TERESA WU, Arizona State University, KEWEI CHEN, ADAM FLEISHER, ERIC REIMAN, Banner Alzheimer?s Institute and Banner PET Center, and JIEPING YE, Arizona State University.

Subject Area: Neuroscience See abstract, page 175.

Oral Session (4:10–5:10pm): "Probabilistic Models and Kernels"

4:10pm Reading Tea Leaves: How Humans Interpret Topic Models

Jonathan Chang	jcone@princeton.edu
Jordan Boyd-Graber	jbg@princeton.edu
Sean Gerrish	sgerrish@princeton.edu
Chong Wang	chongw@cs.princeton.edu
David Blei	blei@cs.princeton.edu
Princeton University	

Probabilistic topic models are a commonly used tool for analyzing text data, where the latent topic representation is used to perform qualitative evaluation of models and guide corpus exploration. Practitioners typically assume that the latent space is semantically meaningful, but this important property has lacked a quantitative evaluation. In this paper, we present new quantitative methods for measuring semantic meaning in inferred topics. We back these measures with large-scale user studies, showing that they capture aspects of the model that are undetected by measures of model quality based on held-out likelihood. Surprisingly, topic models which perform better on held-out likelihood may actually infer less semantically meaningful topics.

Subject Area: Unsupervised and Semi-supervised Learning

4:30pm Sharing Features among Dynamical Systems with Beta Processes

Emily Fox	ebfox@mit.edu
Massachusetts Institute of Teo	chnology
Erik Sudderth	sudderth@cs.brown.edu
Brown University	
Michael Jordan	jordan@eecs.berkeley.edu
UC Berkeley	
Alan Willsky	willsky@mit.edu
MIT	

We propose a Bayesian nonparametric approach to relating multiple time series via a set of latent, dynamical behaviors. Using a beta process prior, we allow data-driven selection of the size of this set, as well as the pattern with which behaviors are shared among time series. Via the Indian buffet process representation of the beta process' predictive distributions, we develop an exact Markov chain Monte Carlo inference method. In particular, our approach uses the sum-product algorithm to efficiently compute Metropolis-Hastings acceptance probabilities, and explores new dynamical behaviors via birth/death proposals. We validate our sampling algorithm using several synthetic datasets, and also demonstrate promising unsupervised segmentation of visual motion capture data.

Subject Area: Probabilistic Models and Methods

4:50pm Fast subtree kernels on graphs

NINO SHERVASHIDZE KARSTEN BORGWARDT MPIs Tuebingen nino.shervashidze@tuebingen.mpg.de karsten.borgwardt@tuebingen.mpg.de

In this article, we propose fast subtree kernels on graphs. On graphs with n nodes and m edges and maximum degree d, these kernels comparing subtrees of height h can be computed in O(m h), whereas the classic subtree kernel by Ramon & Gärtner scales as $O(n^2 \ 4^d \ h)$. Key to this efficiency is the observation that the Weisfeiler-Lehman test of isomorphism from graph theory elegantly computes a subtree kernel as a byproduct. Our fast subtree kernels can deal with labeled graphs, scale up easily to large graphs and outperform state-of-the-art graph kernels on several classification benchmark datasets in terms of accuracy and runtime.

Subject Area: Supervised Learning

Spotlights (5:10–5:30pm)

- Tracking Dynamic Sources of Malicious Activity at Internet Scale SHOBHA VENKATARAMAN, AT&T Labs – Research, AVRIM BLUM, CMU, DAWN SONG, UC Berkeley, SUBHABRATA SEN and OLIVER SPATSCHECK, AT&T Labs – Research. Subject Area: Theory See abstract, page 182.
- Skill Discovery in Continuous Reinforcement Learning Domains using Skill Chaining GEORGE KONIDARIS, University of Massachusetts, and ANDREW BARTO, University of Massachusetts Amherst.
 Subject Area: Control and Reinforcement Learning See abstract, page 179.
- Efficient Match Kernel between Sets of Features for Visual Recognition LIEFENG BO, Toyota Technological Institute at Chicago (TTI-C), and CRISTIAN SMINCHISESCU, U Bonn. Subject Area: Vision See abstract, page 171.
- Orthogonal Matching Pursuit From Noisy Random Measurements: A New Analysis ALYSON FLETCHER, University of California, Berkeley, and SUNDEEP RANGAN, Qualcomm, Inc..
 Subject Area: Supervised Learning See abstract, page 166.

• Learning to Hash with Binary Reconstructive Embeddings BRIAN KULIS, UC Berkeley EECS & ICSI, and TREVOR DARRELL, UC Berkeley EECS and ICSI.

Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 189.

- Perceptual Multistability as Markov Chain Monte Carlo Inference SAMUEL GERSHMAN, Princeton University, ED VUL and JOSHUA TENENBAUM, Massachusetts Institute of Technology.
 Subject Area: Cognitive Science See abstract, page 178.
- Information-theoretic lower bounds on the oracle complexity of convex optimization ALEKH AGARWAL, PETER BARTLETT, UC Berkeley, PRADEEP RAVIKUMAR, University of California, and MARTIN WAINWRIGHT, UC Berkeley.
 Subject Area: Theory See abstract, page 183.
- Hierarchical Learning of Dimensional Biases in Human Categorization KATHERINE HELLER, University of Cambridge, ADAM SANBORN and NICK CHATER, University College London.
 Subject Area: Cognitive Science See abstract, page 198.
- Lower bounds on minimax rates for nonparametric regression with additive sparsity and smoothness GARVESH RASKUTTI, UC Berkeley, MARTIN WAINWRIGHT and BIN YU, University of California.
 Subject Area: Theory See abstract, page 183.
- Neural Implementation of Hierarchical Bayesian Inference by Importance Sampling LEI SHI, UC Berkeley, and THOMAS GRIFFITHS, University of California, Berkeley. Subject Area: Neuroscience See abstract, page 175.
- Posterior vs Parameter Sparsity in Latent Variable Models JOAO GRACA, University of Lisbon, KUZMAN GANCHEV, University of Pennsylvania, BEN TASKAR, University of Pennsylvania, and FERNANDO PEREIRA, Google. Subject Area: Unsupervised and Semi-supervised Learning See abstract, page 189.
- Structural inference affects depth perception in the context of potential occlusion IAN STEVENSON and KONRAD KOERDING, Northwestern University. Subject Area: Vision See abstract, page 172.

WEDNESDAY, POSTER SESSION

• Bilinear classifiers for visual recognition HAMED PIRSIAVASH, DEVA RAMANAN and CHARLESS FOWLKES, UC Irvine. Subject Area: Vision See abstract, page 173.

Poster Session (7:30–11:59pm)

W1 Functional network reorganization in motor cortex can be explained by reward-modulated Hebbian learning

ROBERT LEGENSTEIN, Graz University of Technology, STEVEN CHASE, Carnegie Mellon University, ANDREW SCHWARTZ, University of Pittsburgh, and WOLFGANG MAASS, Graz University of Technology.
Subject Area: Neuroscience Oral presentation, Wednesday, 9:30am. See abstract, page 153.

W2 Efficient Large-Scale Distributed Training of Conditional Maximum Entropy Models

Conditional mathematic	Littiopy moucis
Gideon Mann	gmann@google.com
Ryan McDonald	ryanmcd@google.com
Google	
Mehryar Mohri	mohri@google.com
Google Research	
Nathan Silberman	nsilberman@google.com
Google Inc.	
Dan Walker	danwalkeriv@gmail.com
New York University	

Training conditional maximum entropy models on massive data requires significant time and computational resources. In this paper, we investigate three common distributed training strategies: distributed gradient, majority voting ensembles, and parameter mixtures. We analyze the worst-case runtime and resource costs of each and present a theoretical foundation for the convergence of parameters under parameter mixtures, the most efficient strategy. We present large-scale experiments comparing the different strategies and demonstrate that parameter mixtures over independent models use fewer resources and achieve comparable loss as compared to standard approaches. **Subject Area:** Supervised Learning

Spotlight presentation, Wednesday, 3:20.

W3 Ensemble Nystrom Method

SANJIV KUMARsanjivk@google.comGoogle Research NYmohri@cs.nyu.eduMEHRYAR MOHRImohri@cs.nyu.eduCourant Institute of Mathematical Sciences and Google ResearchAMEET TALWALKARameet@cs.nyu.eduCourant Institute (NYU)

A crucial technique for scaling kernel methods to very large data sets reaching or exceeding millions of instances is based on low-rank approximation of kernel matrices. We introduce a new family of algorithms based on mixtures of Nystrom approximations, ensemble Nystrom algorithms, that yield more accurate low-rank approximations than the standard Nystrom method. We give a detailed study of multiple variants of these algorithms based on simple averaging, an exponential weight method, or regression-based methods. We also present a theoretical analysis of these algorithms, including novel error bounds guaranteeing a better convergence rate than the standard Nystrom method. Finally, we report the results of extensive experiments with several data sets containing up to 1M points demonstrating the significant performance improvements gained over the standard Nystrom approximation. **Subject Area:** Supervised Learning

W4 Learning Label Embeddings for Nearest-Neighbor Multi-class Classification with an Application to Speech Recognition

NATASHA SINGH-MILLER	natashas@mit.edu
MIT	
Michael Collins	mcollins@csail.mit.edu
Massachusetts Institute of Technology	

We consider the problem of using nearest neighbor methods to provide a conditional probability estimate, P(y-a), when the number of labels y is large and the labels share some underlying structure. We propose a method for learning error-correcting output codes (ECOCs) to model the similarity between labels within a nearest neighbor framework. The learned ECOCs and nearest neighbor information are used to provide conditional probability estimates. We apply these estimates to the problem of acoustic modeling for speech recognition. We demonstrate an absolute reduction in word error rate (WER) of 0.9% (a 2.5% relative reduction in WER) on a lecture recognition task over a state-of-the-art baseline GMM model.

Subject Area: Supervised Learning

W5 From PAC-Bayes Bounds to KL Regularization

Pascal Germain	pascal.germain@ift.ulaval.ca
Alexandre Lacasse	alexandre.lacasse@ift.ulaval.ca
Laval University	
Francois Laviolette	Francois.Laviolette@ift.ulaval.ca
Mario Marchand	Mario.Marchand@ift.ulaval.ca
Université Laval	
SARA SHANIAN	<pre>sara.shanian@ift.ulaval.ca</pre>
Laval University	

We show that convex KL-regularized objective functions are obtained from a PAC-Bayes risk bound when using convex loss functions for the stochastic Gibbs classifier that upper-bound the standard zero-one loss used for the weighted majority vote. By restricting ourselves to a class of posteriors, that we call quasi uniform, we propose a simple coordinate descent learning algorithm to minimize the proposed KL-regularized cost function. We show that standard ℓ_p -regularized objective functions currently used, such as ridge regression and ℓ_p -regularized boosting, are obtained from a relaxation of the KL divergence between the quasi uniform posterior and the uniform prior. We present numerical experiments where the proposed learning algorithm generally outperforms ridge regression and AdaBoost.

 ${\bf Subject \ Area: \ Supervised \ Learning}$

W6 Fast subtree kernels on graphs

NINO SHERVASHIDZE and KARSTEN BORGWARDT, MPIs Tuebingen. Subject Area: Supervised Learning Oral presentation, Wednesday, 4:50pm. See abstract, page 160.

W7 Adaptive Regularization of Weight Vectors

Koby Crammer	koby@ee.technion.ac.il
The Technion	
Alex Kulesza	kulesza@cis.upenn.edu
University of Pennsylvania	
Mark Dredze	mdredze@cs.jhu.edu
Johns Hopkins University	

We present AROW, a new online learning algorithm that combines several properties of successful : large margin training, confidence weighting, and the capacity to handle non-separable data. AROW performs adaptive regularization of the prediction function upon seeing each new instance, allowing it to perform especially well in the presence of label noise. We derive a mistake bound, similar in form to the second order perceptron bound, which does not assume separability. We also relate our algorithm to recent confidence-weighted online learning techniques and empirically show that AROW achieves state-of-the-art performance and notable robustness in the case of non-separable data. **Subject Area:** Supervised Learning

Spotlight presentation, Wednesday, 3:20.

W8 Kernel Methods for Deep Learning

YOUNGMIN CHO yoc002@cs.ucsd.edu University of California, San Diego LAWRENCE SAUL saul@cs.ucsd.edu University of California We introduce a new family of positive-definite kernel functions that mimic the computation in large, multilayer neural nets. These kernel functions can be used in shallow architectures, such as support vector machines (SVMs), or in deep kernel-based architectures that we call multilayer kernel machines (MKMs). We evaluate SVMs and MKMs with these kernel functions on problems designed to illustrate the advantages of deep architectures. On several problems, we obtain better results than previous, leading benchmarks from both SVMs with Gaussian kernels as well as deep belief nets. **Subject Area:** Supervised Learning

W9 Lattice Regression

ERIC GARCIA MAYA GUPTA University of Washington

garciaer@ee.washington.edu gupta@ee.washington.edu

We present a new empirical risk minimization framework for approximating functions from training samples for low-dimensional regression applications where a lattice (lookup table) is stored and interpolated at run-time for an efficient hardware implementation. Rather than evaluating a fitted function at the lattice nodes without regard to the fact that samples will be interpolated, the proposed lattice regression approach estimates the lattice to minimize the interpolation error on the given training samples. Experiments show that lattice regression can reduce mean test error compared to Gaussian process regression for digital color management of printers, an application for which linearly interpolating a look-up table (LUT) is standard. Simulations confirm that lattice regression performs consistently better than the naive approach to learning the lattice, particularly when the density of training samples is low.

Subject Area: Supervised Learning

W10 Slow, Decorrelated Features for Pretraining Complex Cell-like Networks

JAMES BERGSTRA YOSHUA BENGIO University of Montreal bergstrj@iro.umontreal.ca yoshua.bengio@umontreal.ca

We introduce a new type of neural network activation function based on recent physiological rate models for complex cells in visual area V1. A single-hidden-layer neural network of this kind of model achieves 1.5% error on MNIST. We also introduce an existing criterion for learning slow, decorrelated features as a pretraining strategy for image models. This pretraining strategy results in orientation-selective features, similar to the receptive fields of complex cells. With this pretraining, the same single-hidden-layer model achieves better generalization error, even though the pretraining sample distribution is very different from the fine-tuning distribution. To implement this pretraining strategy, we derive a fast algorithm for online learning of decorrelated features such that each iteration of the algorithm runs in linear time with respect to the number of features.

Subject Area: Supervised Learning

W11 Orthogonal Matching Pursuit From Noisy Random Measurements: A New Analysis

 ALYSON FLETCHER
 alyson@eecs.berkeley.edu

 University of California, Berkeley
 sdrangan@yahoo.com

 SUNDEEP RANGAN
 sdrangan@yahoo.com

 Qualcomm, Inc.
 sdrangan@yahoo.com

Orthogonal matching pursuit (OMP) is a widely used greedy algorithm for recovering sparse vectors from linear measurements. A well-known analysis of Tropp and Gilbert shows that OMP can recover a k-sparse n-dimensional real vector from $m = 4k \log(n)$ noise-free random linear measurements with a probability that goes to one as n goes to infinity. This work shows strengthens this result by showing that a lower number of measurements, $m = 2k \log(n-k)$, is in fact sufficient for asymptotic recovery. Moreover, this number of measurements is also sufficient for detection of the sparsity pattern (support) of the vector with measurement errors provided the signal-to-noise ratio (SNR) scales to infinity. The scaling $m = 2k \log(n-k)$ exactly matches the number of measurements required by the more complex lasso for signal recovery. **Subject Area:** Supervised Learning

3

Spotlight presentation, Wednesday, 5:10.

W12 Sparse Estimation Using General Likelihoods and Non-Factorial Priors

David Wipf	david.wipf@mrsc.ucsf.edu
Srikantan Nagarajan	<pre>sri@radiology.ucsf.edu</pre>
UC San Francisco	

Finding maximally sparse representations from overcomplete feature dictionaries frequently involves minimizing a cost function composed of a likelihood (or data fit) term and a prior (or penalty function) that favors sparsity. While typically the prior is factorial, here we examine non-factorial alternatives that have a number of desirable properties relevant to sparse estimation and are easily implemented using an efficient, globally-convergent reweighted ℓ_1 minimization procedure. The first method under consideration arises from the sparse Bayesian learning (SBL) framework. Although based on a highly non-convex underlying cost function, in the context of canonical sparse estimation problems, we prove uniform superiority of this method over the Lasso in that, (i) it can never do worse, and (ii) for any dictionary and sparsity profile, there will always exist cases where it does better. These results challenge the prevailing reliance on strictly convex penalty functions for finding sparse solutions. We then derive a new non-factorial variant with similar properties that exhibits further performance improvements in empirical tests. For both of these methods, as well as traditional factorial analogs, we demonstrate the effectiveness of reweighted ℓ_1 -norm algorithms in handling more general sparse estimation problems involving classification, group feature selection, and non-negativity constraints. As a byproduct of this development, a rigorous reformulation of sparse Bayesian classification (e.g., the relevance vector machine) is derived that, unlike the original, involves no approximation steps and descends a well-defined objective function.

Subject Area: Supervised Learning

W13 Efficient Learning using Forward-Backward Splitting

JOHN DUCHI, UC Berkeley, and YORAM SINGER, Google. Subject Area: Optimization Oral presentation, Wednesday, 10:40am. See abstract, page 155.

W14 A Data-Driven Approach to Modeling Choice

VIVEK FARIAS	vivekf@mit.edu
MIT Sloan	
Srikanth Jagabathula	jskanth@mit.edu
Devavrat Shah	devavrat@mit.edu
MIT	

We visit the following fundamental problem: For a 'generic' model of consumer choice (namely, distributions over preference lists) and a limited amount of data on how consumers actually make decisions (such as marginal preference information), how may one predict revenues from offering a particular assortment of choices? This problem is central to areas within operations research, marketing and econometrics. We present a framework to answer such questions and design a number of tractable algorithms (from a data and computational standpoint) for the same.

Subject Area: Optimization

Spotlight presentation, Wednesday, 3:20.

W15 Unsupervised Detection of Regions of Interest Using Iterative Link Analysis

GUNHEE KIM gunhee@cs.cmu.edu Carnegie Mellon University ANTONIO TORRALBA torralba@csail.mit.edu Massachusetts Institute of Technology

This paper proposes a fast and scalable alternating optimization technique to detect regions of interest (ROIs) in cluttered Web images without labels. The proposed approach discovers highly probable regions of object instances by iteratively repeating the following two functions: (1) choose the exemplar set (i.e. small number of high ranked reference ROIs) across the dataset and (2) refine the ROIs of each image with respect to the exemplar set. These two subproblems are formulated as ranking in two different similarity networks of ROI hypotheses by link analysis. The experiments with the PASCAL 06 dataset show that our unsupervised localization performance is better than one of state-of-the-art techniques and comparable to supervised methods. Also, we test the scalability of our approach with five objects in Flickr dataset consisting of more than 200,000 images.

Subject Area: Vision

Srinivas Turaga	sturaga@mit.edu
MIT	
Kevin Briggman	Kevin.Briggman@mpimf-heidelberg.mpg.de
Moritz Helmstaedter	Moritz.Helmstaedter@mpimf-heidelberg.mpg.de
Winfried Denk	Winfried.Denk@mpimf-heidelberg.mpg.de
Max-Planck Institute for Medical	Research
Sebastian Seung	seung@mit.edu
MIT	

Images can be segmented by first using a classifier to predict an affinity graph that reflects the degree to which image pixels must be grouped together and then partitioning the graph to yield a segmentation. Machine learning has been applied to the affinity classifier to produce affinity graphs that are good in the sense of minimizing edge misclassification rates. However, this error measure is only indirectly related to the quality of segmentations produced by ultimately partitioning the affinity graph. We present the first machine learning algorithm for training a classifier to produce affinity graphs that are good in the sense of producing segmentations that directly minimize the Rand index, a well known segmentation performance measure. The Rand index measures segmentation performance by quantifying the classification of the connectivity of image pixel pairs after segmentation. By using the simple graph partitioning algorithm of finding the connected components of the thresholded affinity graph, we are able to train an affinity classifier to directly minimize the Rand index of segmentations resulting from the graph partitioning. Our learning algorithm corresponds to the learning of maximin affinities between image pixel pairs, which are predictive of the pixel-pair connectivity. Subject Area: Vision

W17 Evaluating multi-class learning strategies in a generative hierarchical framework for object detection

Sanja Fidler	sanja.fidler@fri.uni-lj.si
Marko Boben	marko.boben@fri.uni-lj.si
Ales Leonardis	ales.leonardis@fri.uni-lj.si
University of Ljubljana	

Multiple object class learning and detection is a challenging problem due to the large number of object classes and their high visual variability. Specialized detectors usually excel in performance, while joint representations optimize sharing and reduce inference time — but are complex to train. Conveniently, sequential learning of categories cuts down training time by transferring existing knowledge to novel classes, but cannot fully exploit the richness of shareability and might depend on ordering in learning. In hierarchical frameworks these issues have been little explored. In this paper, we show how different types of multi-class learning can be done within one generative hierarchical framework and provide a rigorous experimental analysis of various object class learning strategies as the number of classes grows. Specifically, we propose, evaluate and compare three important types of multi-class learning: 1.) independent training of individual categories, 2.) joint training of classes, 3.) sequential learning of classes. We explore and compare their computational behavior (space and time) and detection performance as a function of the number of learned classes on several recognition data sets. **Subject Area:** Vision

W18 Hierarchical Modeling of Local Image Features through L_v-Nested Symmetric Distributions

Fabian Sinz	fabee@tuebingen.mpg.de
MPI for Biological Cybernetics &	University of Tübingen
Eero Simoncelli	eero.simoncelli@nyu.edu
New York University	
Matthias Bethge	mbethge@tuebingen.mpg.de
MPI for Biological Cybernetics &	University of Tübingen

We introduce a new family of distributions, called L_p -nested symmetric distributions, whose densities access the data exclusively through a hierarchical cascade of L_p -norms. This class generalizes the family of spherically and L_p -spherically symmetric distributions which have recently been successfully used for natural image modeling. Similar to those distributions it allows for a nonlinear mechanism to reduce the dependencies between its variables. With suitable choices of the parameters and norms, this family also includes the Independent Subspace Analysis (ISA) model, which has been proposed as a means of deriving filters that mimic complex cells found in mammalian primary visual cortex. L_p -nested distributions are easy to estimate and allow us to explore the variety of models between ISA and the L_p -spherically symmetric models. Our main findings are that, without a preprocessing step of contrast gain control, the independent subspaces of ISA are in fact more dependent than the individual filter coefficients within a subspace and, with contrast gain control, where ISA finds more than one subspace, the filter responses were almost independent anyway. **Subject Area:** Vision

W19 Fast Image Deconvolution using Hyper-Laplacian Priors

DILIP KRISHNAN ROB FERGUS New York University dilip@cs.nyu.edu fergus@cs.nyu.edu

The heavy-tailed distribution of gradients in natural scenes have proven effective priors for a range of problems such as denoising, deblurring and super-resolution. However, the use of sparse distributions makes the problem non-convex and impractically slow to solve for multi-megapixel images. In this paper we describe a deconvolution approach that is several orders of magnitude faster than existing techniques that use hyper-Laplacian priors. We adopt an alternating minimization scheme where one of the two phases is a non-convex problem that is separable over pixels. This per-pixel sub-problem may be solved with a lookup table (LUT). Alternatively, for two specific values of α , 1/2 and 2/3 an analytic solution can be found, by finding the roots of a cubic and quartic polynomial, respectively. Our approach (using either LUTs or analytic formulae) is able to deconvolve a 1 megapixel image in less than ~3 seconds, achieving comparable quality to existing methods such as iteratively reweighted least squares (IRLS) that take ~20 minutes. Furthermore, our method is quite general and can easily be extended to related image processing problems, beyond the deconvolution application demonstrated. **Subject Area:** Vision

Spotlight presentation, Wednesday, 3:20.

W20 The Ordered Residual Kernel for Robust Motion Subspace Clustering

Tat-Jun Chin	tatjun@gmail.com
Hanzi Wang	hwang@cs.adelaide.edu.au
David Suter	dsuter@cs.adelaide.edu.au
The University of Adelaide	

We present a novel and highly effective approach for multi-body motion segmentation. Drawing inspiration from robust statistical model fitting, we estimate putative subspace hypotheses from the data. However, instead of ranking them we encapsulate the hypotheses in a novel Mercer kernel which elicits the potential of two point trajectories to have emerged from the same subspace. The kernel permits the application of well-established statistical learning methods for effective outlier rejection, automatic recovery of the number of motions and accurate segmentation of the point trajectories. The method operates well under severe outliers arising from spurious trajectories or mistracks. Detailed experiments on a recent benchmark dataset (Hopkins 155) show that our method is superior to other state-of-the-art approaches in terms of recovering the number of motions, segmentation accuracy, robustness against gross outliers and computational efficiency.

Subject Area: Vision

W21 Region-based Segmentation and Object Detection

Stephen Gould	sgould@stanford.edu
Tianshi Gao	tianshig@stanford.edu
Daphne Koller	koller@cs.stanford.edu
Stanford University	

Object detection and multi-class image segmentation are two closely related tasks that can be greatly improved when solved jointly by feeding information from one task to the other. However, current state-of-the-art models use a separate representation for each task making joint inference clumsy and leaving classification of many parts of the scene ambiguous. In this work, we propose a hierarchical region-based approach to joint object detection and image segmentation. Our approach reasons about pixels, regions and objects in a coherent probabilistic model. Importantly, our model gives a single unified description of the scene. We explain every pixel in the image and enforce global consistency between all variables in our model. We run experiments on challenging vision datasets and show significant improvement over state-of-the-art object detection accuracy. **Subject Area:** Vision

Spotlight presentation, Wednesday, 3:20.

W22 Efficient Match Kernel between Sets of Features for Visual Recognition

LIEFENG BO liefengbo@gmail.com Toyota Technological Institute at Chicago (TTI-C) CRISTIAN SMINCHISESCU cristian.sminchisescu@ins.uni-bonn.de U Bonn In visual recognition, the images are frequently modeled as sets of local features (bags). We show that bag of words, a common method to handle such cases, can be viewed as a special match kernel, which counts 1 if two local features fall into the same regions partitioned by visual words and 0 otherwise. Despite its simplicity, this quantization is too coarse. It is, therefore, appealing to design match kernels that more accurately measure the similarity between local features. However, it is impractical to use such kernels on large datasets due to their significant computational cost. To address this problem, we propose an efficient match kernel (EMK), which maps local features to a low dimensional feature space, average the resulting feature vectors to form a set-level feature, then apply a linear classifier. The local feature maps are learned so that their inner products preserve, to the best possible, the values of the specified kernel function. EMK is linear both in the number of images and in the number of local features. We demonstrate that EMK is extremely efficient and achieves the current state of the art performance on three difficult real world datasets: Scene-15, Caltech-101 and Caltech-256. **Subject Area:** Vision

Spotlight presentation, Wednesday, 5:10.

W23 Segmenting Scenes by Matching Image Composites

BRYAN RUSSELLrussell@di.ens.frINRIA - ENS ParisALYOSHA EFROSefros@cs.cmu.eduCarnegie Mellon UniversityJOSEF SIVICjosef@di.ens.frINRIA ENSBILL FREEMANbillf@mit.eduMassachusetts Institute of TechnologyANDREW ZISSERMANaz@robots.ox.ac.ukUniversity of Oxford

In this paper, we investigate how similar images sharing the same global description can help with unsupervised scene segmentation in an image. In contrast to recent work in semantic alignment of scenes, we allow an input image to be explained by partial matches of similar scenes. This allows for a better explanation of the input scenes. We perform MRF-based segmentation that optimizes over matches, while respecting boundary information. The recovered segments are then used to re-query a large database of images to retrieve better matches for the target region. We show improved performance in detecting occluding boundaries over previous methods on data gathered from the LabelMe database.

Subject Area: Vision

W24 Structural inference affects depth perception in the context of potential occlusion

IAN STEVENSON KONRAD KOERDING Northwestern University i-stevenson@northwestern.edu kk@northwestern.edu In many domains, humans appear to combine perceptual cues in a near-optimal, probabilistic fashion: two noisy pieces of information tend to be combined linearly with weights proportional to the precision of each cue. Here we present a case where structural information plays an important role. The presence of a background cue gives rise to the possibility of occlusion, and places a soft constraint on the location of a target – in effect propelling it forward. We present an ideal observer model of depth estimation for this situation where structural or ordinal information is important and then fit the model to human data from a stereo-matching task. To test whether subjects are truly using ordinal cues in a probabilistic manner we then vary the uncertainty of the task. We find that the model accurately predicts shifts in subject's behavior. Our results indicate that the nervous system estimates depth ordering in a probabilistic fashion and estimates the structure of the visual scene during depth perception.

Spotlight presentation, Wednesday, 5:10.

W25 Who's Doing What: Joint Modeling of Names and Verbs for Simultaneous Face and Pose Annotation

JIE LUO Idiap / EPF Lausanne BARBARA CAPUTO Idiap Research Institute VITTORIO FERRARI ETH Zurich jluo@idiap.ch

bcaputo@idiap.ch

ferrari@vision.ee.ethz.ch

Given a corpus of news items consisting of images accompanied by text captions, we want to find out "who's doing what", i.e. associate names and action verbs in the captions to the face and body pose of the persons in the images. We present a joint model for simultaneously solving the image-caption correspondences and learning visual appearance models for the face and pose classes occurring in the corpus. These models can then be used to recognize people and actions in novel images without captions. We demonstrate experimentally that our joint 'face and pose' model solves the correspondence problem better than earlier models covering only the face, and that it can perform recognition of new uncaptioned images.

Subject Area: Vision

W26 Bilinear classifiers for visual recognition

HAMED PIRSIAVASH DEVA RAMANAN CHARLESS FOWLKES UC Irvine hpirsiav@ics.uci.edu dramanan@ics.uci.edu fowlkes@ics.uci.edu We describe an algorithm for learning bilinear SVMs. Bilinear classifiers are a discriminative variant of bilinear models, which capture the dependence of data on multiple factors. Such models are particularly appropriate for visual data that is better represented as a matrix or tensor, rather than a vector. Matrix encodings allow for more natural regularization through rank restriction. For example, a rank-one scanning-window classifier yields a separable filter. Low-rank models have fewer parameters and so are easier to regularize and faster to score at run-time. We learn low-rank models with bilinear classifiers. We also use bilinear classifiers for transfer learning by sharing linear factors between different classification tasks. Bilinear classifiers are trained with biconvex programs. Such programs are optimized with coordinate descent, where each coordinate step requires solving a convex program - in our case, we use a standard off-the-shelf SVM solver. We demonstrate bilinear SVMs on difficult problems of people detection in video sequences and action classification of video sequences, achieving state-of-the-art results in both.

Subject Area: Vision

Spotlight presentation, Wednesday, 5:10.

W27 Neurometric function analysis of population codes

Philipp Berens	berens@tuebingen.mpg.de
Sebastian Gerwinn	sgerwinn@tuebingen.mpg.de
Alexander Ecker	alexander.ecker@tuebingen.mpg.de
Matthias Bethge	mbethge@tuebingen.mpg.de
MPI for Biological Cybernetics &	University of Tübingen

The relative merits of different population coding schemes have mostly been analyzed in the framework of stimulus reconstruction using Fisher Information. Here, we consider the case of stimulus discrimination in a two alternative forced choice paradigm and compute neurometric functions in terms of the minimal discrimination error and the Jensen-Shannon information to study neural population codes. We first explore the relationship between minimum discrimination error, Jensen-Shannon Information and Fisher Information and show that the discrimination framework is more informative about the coding accuracy than Fisher Information as it defines an error for any pair of possible stimuli. In particular, it includes Fisher Information as a special case. Second, we use the framework to study population codes of angular variables. Specifically, we assess the impact of different noise correlations structures on coding accuracy in long versus short decoding time windows. That is, for long time window we use the common Gaussian noise approximation. To address the case of short time windows we analyze the Ising model with identical noise correlation structure. In this way, we provide a new rigorous framework for assessing the functional consequences of noise correlation structures for the representational accuracy of neural population codes that is in particular applicable to short-time population coding. Subject Area: Neuroscience

W28 Time-rescaling methods for the estimation and assessment of non-Poisson neural encoding models

JONATHAN PILLOW, University of Texas at Austin. Subject Area: Neuroscience Oral presentation, Thursday, 11:30am. See abstract, page 204.

W29 Know Thy Neighbour: A Normative Theory of Synaptic Depression

JEAN-PASCAL PFISTER, University of Cambridge, PETER DAYAN, University College London, and MATE LENGYEL, University of Cambridge. Subject Area: Neuroscience Oral presentation, Wednesday, 9:50am. See abstract, page 154.

W30 Learning Brain Connectivity of Alzheimer's Disease from Neuroimaging Data

Shuai Huang	shuang31@asu.edu
Arizona State University	
Jing Li	jing.li.8@asu.edu
jing.li.8@asu.edu	
Liang Sun	sun.liang@asu.edu
Jun Liu	j.liu@asu.edu
Teresa Wu	teresa.wu@asu.edu
Arizona State University	
Kewei Chen	kewei.chen@bannerhealth.com
Adam Fleisher	adam.fleisher@bannerhealth.com
Eric Reiman	eric.reiman@bannerhealth.com
Banner Alzheimer?s Institute and	Banner PET Center
JIEPING YE	jieping.ye@asu.edu
Arizona State University	

Recent advances in neuroimaging techniques provide great potentials for effective diagnosis of Alzheimer's disease (AD), the most common form of dementia. Previous studies have shown that AD is closely related to alternation in the functional brain network, i.e., the functional connectivity among different brain regions. In this paper, we consider the problem of learning functional brain connectivity from neuroimaging, which holds great promise for identifying image-based markers used to distinguish Normal Controls (NC), patients with Mild Cognitive Impairment (MCI), and patients with AD. More specifically, we study sparse inverse covariance estimation (SICE), also known as exploratory Gaussian graphical models, for brain connectivity modeling. In particular, we apply SICE to learn and analyze functional brain connectivity patterns from different subject groups, based on a key property of SICE, called the "monotone property" we established in this paper. Our experimental results on neuroimaging PET data of 42 AD, 116 MCI, and 67 NC subjects reveal several interesting connectivity patterns consistent with literature findings, and also some new patterns that can help the knowledge discovery of AD. **Subject Area:** Neuroscience

Spotlight presentation, Wednesday, 3:20.

W31 Neural Implementation of Hierarchical Bayesian Inference by Importance Sampling

Lei Shi	lshi.berkeley@gmail.com
UC Berkeley	
Thomas Griffiths	tom_griffiths@berkeley.edu
University of California, Berkeley	

The goal of perception is to infer the hidden states in the hierarchical process by which sensory data are generated. Human behavior is consistent with the optimal statistical solution to this problem in many tasks, including cue combination and orientation detection. Understanding the neural mechanisms underlying this behavior is of particular importance, since probabilistic computations are notoriously challenging. Here we propose a simple mechanism for Bayesian inference which involves averaging over a few feature detection neurons which fire at a rate determined by their similarity to a sensory stimulus. This mechanism is based on a Monte Carlo method known as importance sampling, commonly used in computer science and statistics. Moreover, a simple extension to recursive importance sampling can be used to perform hierarchical Bayesian inference. We identify a scheme for implementing importance sampling with spiking neurons, and show that this scheme can account for human behavior in cue combination and oblique effect.

Subject Area: Neuroscience

Spotlight presentation, Wednesday, 5:10.

W32 Exploring Functional Connectivities of the Human Brain using Multivariate Information Analysis

BARRY CHAIbwchai@cs.stanford.edu*StanfordDIRK WALTHERwalther@illinois.edu*DIANE BECKdmbeck@illinois.edu**University of IllinoisLI FEI-FEIfeifeili@CS.Princeton.edu**Princeton University

In this study, we present a method for estimating the mutual information for a localized pattern of fMRI data. We show that taking a multivariate information approach to voxel selection leads to a decoding accuracy that surpasses an univariate information approach and other standard voxel selection methods. Furthermore, we extend the multivariate mutual information theory to measure the functional connectivity between distributed brain regions. By jointly estimating the information shared by two sets of voxels we can reliably map out the connectivities in the human brain during experiment conditions. We validated our approach on a 6-way scene categorization fMRI experiment. The multivariate information analysis is able to find strong information flow between PPA and RSC, which confirms existing neuroscience studies on scenes. Furthermore, by exploring over the whole brain, our method identifies other interesting ROIs that share information with the PPA, RSC scene network, suggesting interesting future work for neuroscientists. (* and ** indicate equal contributions)

Subject Area: Neuroscience

W33 Quantification and the language of thought

CHARLES KEMP Carnegie Mellon University ckemp@cmu.edu

Many researchers have suggested that the psychological complexity of a concept is related to the length of its representation in a "language of thought." As yet, however, there are few concrete proposals about the nature of this language. This paper makes one such proposal: the language of thought allows first order quantification (quantification over objects) more readily than second-order quantification (quantification over features). To support this proposal we present behavioral results from a concept learning study inspired by the work of Shepard, Hovland and Jenkins.

Subject Area: Cognitive Science

Spotlight presentation, Wednesday, 3:20.

W34 Bayesian Belief Polarization

Alan Jern	ajern@cmu.edu
Kai-min Chang	kkchang@cs.cmu.edu
Charles Kemp	ckemp@cmu.edu
Carnegie Mellon University	

Situations in which people with opposing prior beliefs observe the same evidence and then strengthen those existing beliefs are frequently offered as evidence of human irrationality. This phenomenon, termed belief polarization, is typically assumed to be non-normative. We demonstrate, however, that a variety of cases of belief polarization are consistent with a Bayesian approach to belief revision. Simulation results indicate that belief polarization is not only possible but relatively common within the class of Bayesian models that we consider.

Subject Area: Cognitive Science

W35 Predicting the Optimal Spacing of Study: A Multiscale Context Model of Memory

MICHAEL MOZER	mozer@colorado.edu
University of Colorado at Boulder	
HAROLD PASHLER	hpashler@ucsd.edu
UC San Diego	
NICHOLAS CEPEDA	ncepeda@york.ca
York University	
Robert Lindsey	robert.lindsey@colorado.edu
University of Colorado	
Ed Vul	evul@mit.edu
Massachusetts Institute of Technology	

When individuals learn facts (e.g., foreign language vocabulary) over multiple study sessions, the temporal spacing of study has a significant impact on memory retention. Behavioral experiments have shown a nonmonotonic relationship between spacing and retention: short or long intervals between study sessions yield lower cued-recall accuracy than intermediate intervals. Appropriate spacing of study can double retention on educationally relevant time scales. We introduce a Multiscale Context Model (MCM) that is able to predict the influence of a particular study schedule on retention for specific material. MCM's prediction is based on empirical data characterizing forgetting of the material following a single study session. MCM is a synthesis of two existing memory models (Staddon, Chelaru, & Higa, 2002; Raaijmakers, 2003). On the surface, these models are unrelated and incompatible, but we show they share a core feature that allows them to be integrated. MCM can determine study schedules that maximize the durability of learning, and has implications for education and training. MCM can be cast either as a neural network with inputs that fluctuate over time, or as a cascade of leaky integrators. MCM is intriguingly similar to a Bayesian multiscale model of memory (Kording, Tenenbaum, Shadmehr, 2007), yet MCM is better able to account for human declarative memory.

Subject Area: Cognitive Science

Spotlight presentation, Wednesday, 3:20.

W36 Perceptual Multistability as Markov Chain Monte Carlo Inference

SAMUEL GERSHMAN	sjgershm@princeton.edu
Princeton University	
Ed Vul	evul@mit.edu
Joshua Tenenbaum	jbt@mit.edu
Massachusetts Institute of Te	echnology

While many perceptual and cognitive phenomena are well described in terms of Bayesian inference, the necessary computations are intractable at the scale of real-world tasks, and it remains unclear how the human mind approximates Bayesian inference algorithmically. We explore the proposal that for some tasks, humans use a form of Markov Chain Monte Carlo to approximate the posterior distribution over hidden variables. As a case study, we show how several phenomena of perceptual multistability can be explained as MCMC inference in simple graphical models for low-level vision. **Subject Area:** Cognitive Science

Spotlight presentation, Wednesday, 5:10.

W37 Multi-Step Dyna Planning for Policy Evaluation and Control

Hengshuai Yao	hengshua@cs.ualberta.ca
University of alberta	
RICH SUTTON	<pre>sutton@cs.ualberta.ca</pre>
University of Alberta	
Shalabh Bhatnagar	shalabh@csa.iisc.ernet.in
Indian Institute of Science	
Dongcui Diao	dongcui_diao@126.com
South China Normal University	
CSABA SZEPESVARI	szepesva@cs.ualberta.ca
University of Alberta	-

We extend Dyna planning architecture for policy evaluation and control in two significant aspects. First, we introduce a multi-step Dyna planning that projects the simulated state/feature many steps into the future. Our multi-step Dyna is based on a multi-step model, which we call the λ -model. The λ -model interpolates between the one-step model and an infinite-step model, and can be learned efficiently online. Second, we use for Dyna control a dynamic multi-step model that is able to predict the results of a sequence of greedy actions and track the optimal policy in the long run. Experimental results show that Dyna using the multi-step model evaluates a policy faster than using single-step models; Dyna control algorithms using the dynamic tracking model are much faster than model-free algorithms; further, multi-step Dyna control algorithms enable the policy and value function to converge much faster to their optima than single-step Dyna algorithms.

Subject Area: Control and Reinforcement Learning

W38 Bootstrapping from Game Tree Search

JOEL VENESS, UNSW / NICTA, DAVID SILVER, University of Alberta, WILLIAM UTHER, NICTA / University of New South Wales, and ALAN BLAIR, NICTA / UNSW. Subject Area: Control and Reinforcement Learning Oral presentation, Wednesday, 11:20am. See abstract, page 155.

W39 Skill Discovery in Continuous Reinforcement Learning Domains using Skill Chaining

GEORGE KONIDARIS gdk@cs.umass.edu University of Massachusetts ANDREW BARTO barto@cs.umass.edu University of Massachusetts Amherst

We introduce skill chaining, a skill discovery method for reinforcement learning agents in continuous domains, that builds chains of skills leading to an end-of-task reward. We demonstrate experimentally that it creates skills that result in performance benefits in a challenging continuous domain.

Subject Area: Control and Reinforcement Learning

Spotlight presentation, Wednesday, 5:10.

W40 Strategy Grafting in Extensive Games

Kevin Waugh	waugh@cs.ualberta.ca
Nolan Bard	nolan@cs.ualberta.ca
Michael Bowling	bowling@cs.ualberta.ca
University of Alberta	

Extensive games are often used to model the interactions of multiple agents within an environment. Much recent work has focused on increasing the size of an extensive game that can be feasibly solved. Despite these improvements, many interesting games are still too large for such techniques. A common approach for computing strategies in these large games is to first employ an abstraction technique to reduce the original game to an abstract game that is of a manageable size. This abstract game is then solved and the resulting strategy is used in the original game. Most top programs in recent AAAI Computer Poker Competitions use this approach. The trend in this competition has been that strategies found in larger abstract games tend to beat strategies found in smaller abstract games. These larger abstract games have more expressive strategy spaces and therefore contain better strategies. In this paper we present a new method for computing strategies in large games. This method allows us to compute more expressive strategies without increasing the size of abstract games that we are required to solve. We demonstrate the power of the approach experimentally in both small and large games, while also providing a theoretical justification for the resulting improvement.

Subject Area: Control and Reinforcement Learning

W41 Complexity of Decentralized Control: Special Cases

MARTIN ALLEN martin.allen@conncoll.edu Connecticut College SHLOMO ZILBERSTEIN shlomo@cs.umass.edu University of Massachusetts at Amherst

The worst-case complexity of general decentralized POMDPs, which are equivalent to partially observable stochastic games (POSGs) is very high, both for the cooperative and competitive cases. Some reductions in complexity have been achieved by exploiting independence relations in some models. We show that these results are somewhat limited: when these independence assumptions are relaxed in very small ways, complexity returns to that of the general case.

Subject Area: Control and Reinforcement Learning

W42 Discrete MDL Predicts in Total Variation

MARCUS HUTTER marcus.hutter@gmx.net Australian National University

The Minimum Description Length (MDL) principle selects the model that has the shortest code for data plus model. We show that for a countable class of models, MDL predictions are close to the true distribution in a strong sense. The result is completely general. No independence, ergodicity, stationarity, identifiability, or other assumption on the model class need to be made. More formally, we show that for any countable class of models, the distributions selected by MDL (or MAP) asymptotically predict (merge with) the true measure in the class in total variation distance. Implications for non-i.i.d. domains like time-series forecasting, discriminative learning, and reinforcement learning are discussed. **Subject Area:** Theory
ADAM KALAI adum@microsoft.com Microsoft VARUN KANADE vkanade@fas.harvard.edu Harvard University

W43 Potential-Based Agnostic Boosting

We prove strong noise-tolerance properties of a potential-based boosting algorithm, similar to MadaBoost (Domingo and Watanabe, 2000) and SmoothBoost (Servedio, 2003). Our analysis is in the agnostic framework of Kearns, Schapire and Sellie (1994), giving polynomial-time guarantees in presence of arbitrary noise. A remarkable feature of our algorithm is that it can be implemented without reweighting examples, by randomly relabeling them instead. Our boosting theorem gives, as easy corollaries, alternative derivations of two recent non-trivial results in computational learning theory: agnostically learning decision trees (Gopalan et al, 2008) and agnostically learning halfspaces (Kalai et al, 2005). Experiments suggest that the algorithm performs similarly to Madaboost.

Subject Area: Theory

W44 On Stochastic and Worst-case Models for Investing

ELAD HAZAN, IBM, and SATYEN KALE, Yahoo! Research. Subject Area: Theory Oral presentation, Wednesday, 11:00am. See abstract, page 155.

W45 Statistical Consistency of Top-k Ranking

FEN XIAustcfenxia@gmail.comInstitute of Automation, Chinese Academy of SciencesTIE-YAN LIUtyliu@microsoft.comHANG LIMicrosoft Research Asia

This paper is concerned with the consistency analysis on listwise ranking methods. Among various ranking methods, the listwise methods have competitive performances on benchmark datasets and are regarded as one of the state-of-the-art approaches. Most listwise ranking methods manage to optimize ranking on the whole list (permutation) of objects, however, in practical applications such as information retrieval, correct ranking at the top k positions is much more important. This paper aims to analyze whether existing listwise ranking methods are statistically consistent in the top-k setting. For this purpose, we define a top-k ranking framework, where the true loss (and thus the risks) are defined on the basis of top-k subgroup of permutations. This framework can include the permutation-level ranking framework proposed in previous work as a special case. Based on the new framework, we derive sufficient conditions for a listwise ranking method to be consistent with the top-k true loss, and show an effective way of modifying the surrogate loss functions in existing methods to satisfy these conditions. Experimental results show that after the modifications, the methods can work significantly better than their original versions.

Subject Area: Theory

W46 Learning from Multiple Partially Observed Views - an Application to Multilingual Text Categorization

 MASSIH AMINI
 Massih-Reza.Amini@cnrc-nrc.gc.ca

 National Research Councl Canad
 nicolas.usunier@lip6.fr

 NICOLAS USUNIER
 nicolas.usunier@lip6.fr

 Université Pierre et Marie Cur
 Cyril.Goutte@nrc-cnrc.gc.ca

 National Research Council Canada

We address the problem of learning classifiers when observations have multiple views, some of which may not be observed for all examples. We assume the existence of view generating functions which may complete the missing views in an approximate way. This situation corresponds for example to learning text classifiers from multilingual collections where documents are not available in all languages. In that case, Machine Translation (MT) systems may be used to translate each document in the missing languages. We derive a generalization error bound for classifiers learned on examples with multiple artificially created views. Our result uncovers a trade-off between the size of the training set, the number of views, and the quality of the view generating functions. As a consequence, we identify situations where it is more interesting to use multiple views for learning instead of classical single view learning. An extension of this framework is a natural way to leverage unlabeled multi-view data in semi-supervised learning. Experimental results on a subset of the Reuters RCV1/RCV2 collections support our findings by showing that additional views obtained from MT may significantly improve the classification performance in the cases identified by our trade-off.

Subject Area: Theory

W47 Tracking Dynamic Sources of Malicious Activity at Internet Scale

Shobha Venkataraman	shvenk@research.att.com
AT&T Labs – Research	
Avrim Blum	avrim@cs.cmu.edu
CMU	
Dawn Song	dawnsong@cs.berkeley.edu
UC Berkeley	
Subhabrata Sen	sen@research.att.com
OLIVER SPATSCHECK	spatsch@research.att.com
AT&T Labs – Research	

We formulate and address the problem of discovering dynamic malicious regions on the Internet. We model this problem as one of adaptively pruning a known decision tree, but with additional challenges: (1) severe space requirements, since the underlying decision tree has over 4 billion leaves, and (2) a changing target function, since malicious activity on the Internet is dynamic. We present a novel algorithm that addresses this problem, by putting together a number of different "experts" algorithms and online paging algorithms. We prove guarantees on our algorithm's performance as a function of the best possible pruning of a similar size, and our experiments show that our algorithm achieves high accuracy on large real-world data sets, with significant improvements over

existing approaches. Subject Area: Theory

Spotlight presentation, Wednesday, 5:10.

W48 Kernel Choice and Classifiability for RKHS Embeddings of Probability Distributions

BHARATH SRIPERUMBUDUR, UC San Diego, KENJI FUKUMIZU, The Institute of Statistical Mathematics, ARTHUR GRETTON, Carnegie Mellon University and Max Planck Institute, GERT LANCKRIET, University of California, and BERNHARD SCHOELKOPF, MPI for Biological Cybernetics.

Subject Area: Theory Oral presentation, Wednesday, 11:40am. See abstract, page 156.

W49 Information-theoretic lower bounds on the oracle complexity of convex optimization

Alekh Agarwal	alekh@cs.berkeley.edu
Peter Bartlett	bartlett@cs.berkeley.edy
UC Berkeley	
Pradeep Ravikumar	pradeepr@stat.berkeley.edu
University of California	
Martin Wainwright	wainwrig@cs.berkeley.edu
UC Berkeley	

Despite the large amount of literature on upper bounds on complexity of convex analysis, surprisingly little is known about the fundamental hardness of these problems. The extensive use of convex optimization in machine learning and statistics makes such an understanding critical to understand fundamental computational limits of learning and estimation. In this paper, we study the complexity of stochastic convex optimization in an oracle model of computation. We improve upon known results and obtain tight minimax complexity estimates for some function classes. We also discuss implications of these results to the understanding the inherent complexity of large-scale learning and estimation problems.

Subject Area: Theory

Spotlight presentation, Wednesday, 5:10.

W50 Lower bounds on minimax rates for nonparametric regression with additive sparsity and smoothness

GARVESH RASKUTTI UC Berkeley MARTIN WAINWRIGHT BIN YU University of California garveshr@stat.berkeley.edu wainwrig@eecs.berkeley.edu binyu@stat.berkeley.edu

This paper uses information-theoretic techniques to determine minimax rates for estimating nonparametric sparse additive regression models under high-dimensional scaling. We assume anadditive decomposition of the form $f^*(X_1, \ldots, X_p) = \sum_{j \in S} h_j(X_j)$, where each component function h_j lies in some Hilbert Space \mathcal{H} and $S \subset \{1, \ldots, p\}$ is an unknown subset with cardinality s = |S|. Given n i.i.d. observations of $f^*(X)$ corrupted with white Gaussian noise where the covariate vectors $(X_1, X_2, X_3, ..., X_p)$ are drawn with i.i.d. components from some distribution \mathbb{P} , we determine tight lower bounds on the minimax rate for estimating the regression function with respect to squared- $L^2(\mathbb{P})$ error. The main result shows that the minimax rates are $\max\left(\frac{s\log(p/s)}{n}, s\epsilon_n^2(\mathcal{H})\right)$. The first term reflects the difficulty of performing subset selection and is independent of the Hilbert space \mathcal{H} ; the second term $s\epsilon_n^2(\mathcal{H})$ is an s-dimensional estimation term, depending only on the low dimension s but not the ambient dimension p, that captures the difficulty of estimating a sum of s univariate functions in the Hilbert space \mathcal{H} . As a special case, if \mathcal{H} corresponds to the *m*-th order Sobolev space W^m of functions that are *m*-times differentiable, the *s*-dimensional estimation term takes the form $s\epsilon_n^2(\mathcal{H}) \simeq s n^{-2m/(2m+1)}$. The minimax rates are compared with rates achieved by an ℓ_1 -penalty based approach, it can be shown that a certain ℓ_1 -based approach achieves the minimax optimal rate. Subject Area: Theory

Spotlight presentation, Wednesday, 5:10.

W51 A Parameter-free Hedging Algorithm

Kamalika Chaudhuri	kamalika@soe.ucsd.edu
Yoav Freund	yfreund@cs.ucsd.edu
UC San Diego	
Daniel Hsu	djhsu@cs.ucsd.edu
University of California	

We study the problem of decision-theoretic online learning (DTOL). Motivated by practical applications, we focus on DTOL when the number of actions is very large. Previous algorithms for learning in this framework have a tunable learning rate parameter, and a major barrier to using online-learning in practical applications is that it is not understood how to set this parameter optimally, particularly when the number of actions is large. In this paper, we offer a clean solution by proposing a novel and completely parameter-free algorithm for DTOL. In addition, we introduce a new notion of regret, which is more natural for applications with a large number of actions. We show that our algorithm achieves good performance with respect to this new notion of regret; in addition, it also achieves performance close to that of the best bounds achieved by previous algorithms with optimally-tuned parameters, according to previous notions of regret. Subject Area: Theory

W52 Reading Tea Leaves: How Humans Interpret Topic Models

JONATHAN CHANG, JORDAN BOYD-GRABER, SEAN GERRISH, CHONG WANG and DAVID BLEI, Princeton University.

Subject Area: Unsupervised and Semi-supervised Learning

Oral presentation, Wednesday, 4:10pm. See abstract, page 160.

W53 Sufficient Conditions for Agnostic Active Learnable

LIWEI WANG Peking University wanglw@cis.pku.edu.cn

We study pool-based active learning in the presence of noise, i.e. the agnostic setting. Previous works have shown that the effectiveness of agnostic active learning depends on the learning problem and the hypothesis space. Although there are many cases on which active learning is very useful, it is also easy to construct examples that no active learning algorithm can have advantage. In this paper, we propose intuitively reasonable sufficient conditions under which agnostic active learning algorithm is strictly superior to passive supervised learning. We show that under some noise condition, if the classification boundary and the underlying distribution are smooth to a finite order, active learning achieves polynomial improvement in the label complexity; if the boundary and the distribution are infinitely smooth, the improvement is exponential. **Subject Area:** Unsupervised and Semi-supervised Learning

W54 Sparse Metric Learning via Smooth Optimization

YIMING YING mathying@gmail.com University of Bristol KAIZHU HUANG kzhuang@nlpr.ia.ac.cn NLPR, Institute of Automation, Chinese Academy of Sciences COLIN CAMPBELL C.Campbell@bris.ac.uk University of Bristol

In this paper we study the problem of learning a low-dimensional (sparse) distance matrix. We propose a novel metric learning model which can simultaneously conduct dimension reduction and learn a distance matrix. The sparse representation involves a mixed-norm regularization which is non-convex. We then show that it can be equivalently formulated as a convex saddle (min-max) problem. From this saddle representation, we develop an efficient smooth optimization approach for sparse metric learning although the learning model is based on a non-differential loss function. This smooth optimization approach has an optimal convergence rate of $O(1/\ell^2)$ for smooth problems where ℓ is the iteration number. Finally, we run experiments to validate the effectiveness and efficiency of our sparse metric learning model on various datasets.

Subject Area: Unsupervised and Semi-supervised Learning

W55 Statistical Analysis of Semi-Supervised Learning: The Limit of Infinite Unlabelled Data

BOAZ NADLERboaz.nadler@weizmann.ac.ilWeizmann Institute of Sciencenati@uchicago.eduNATHAN SREBROnati@uchicago.eduToyota Technological Institute at ChicagoChicagoXUEYUAN ZHOUzhouxy@uchicago.eduUniversity of Chicagozhouxy@uchicago.edu

We study the behavior of the popular Laplacian Regularization method for Semi-Supervised Learning at the regime of a fixed number of labeled points but a large number of unlabeled points. We show that in \mathbb{R}^d , $d \geq 2$, the method is actually not well-posed, and as the number of unlabeled points increases the solution degenerates to a noninformative function. We also contrast the method with the Laplacian Eigenvector method, and discuss the "smoothness" assumptions associated with this alternate method.

Subject Area: Unsupervised and Semi-supervised Learning

Spotlight presentation, Wednesday, 3:20.

W56 Manifold Regularization for SIR with Rate Root-n Convergence

Wei Bian	brian.weibian@gmail.com
NTU	
Dacheng Tao	dctao@ntu.edu.sg
Nanyang Technological University	

In this paper, we study the manifold regularization for the Sliced Inverse Regression (SIR). The manifold regularization improves the standard SIR in two aspects: 1) it encodes the local geometry for SIR and 2) it enables SIR to deal with transductive and semi-supervised learning problems. We prove that the proposed graph Laplacian based regularization is convergent at rate root-n. The projection directions of the regularized SIR are optimized by using a conjugate gradient method on the Grassmann manifold. Experimental results support our theory.

Subject Area: Unsupervised and Semi-supervised Learning

W57 Anomaly Detection with Score functions based on Nearest Neighbor Graphs

Manqi Zhao	mqzhao@bu.edu
Venkatesh Saligrama	saligrama@gmail.com
Boston University	

We propose a novel non-parametric adaptive anomaly detection algorithm for high dimensional data based on score functions derived from nearest neighbor graphs on n-point nominal data. Anomalies are declared whenever the score of a test sample falls below q, which is supposed to be the desired false alarm level. The resulting anomaly detector is shown to be asymptotically optimal in that it is uniformly most powerful for the specified false alarm level, q, for the case when the anomaly density is a mixture of the nominal and a known density. Our algorithm is computationally efficient, being linear in dimension and quadratic in data size. It does not require choosing complicated tuning parameters or function approximation classes and it can adapt to local structure such as local change in dimensionality. We demonstrate the algorithm on both artificial and real data sets in high dimensional feature spaces.

Subject Area: Unsupervised and Semi-supervised Learning

Spotlight presentation, Wednesday, 3:20.

W58 Adaptive Regularization for Transductive Support Vector Machine

Zenglin Xu	zlxu@cse.cuhk.edu.hk
CUHK	
Rong Jin	rongjin@cse.msu.edu
Michigan State University	
Jianke Zhu	zhuji@vision.ee.ethz.ch
ETH Zurich	
Irwin King	king@cse.cuhk.edu.hk
CUHK	
Michael Lyu	lyu@cse.cuhk.edu.hk
The Chinese University of HK	
Zhirong Yang	zhirong.yang@tkk.fi
CUHK,TKK	

We discuss the framework of Transductive Support Vector Machine (TSVM) from the perspective of the regularization strength induced by the unlabeled data. In this framework, SVM and TSVM can be regarded as a learning machine without regularization and one with full regularization from the unlabeled data, respectively. Therefore, to supplement this framework of the regularization strength, it is necessary to introduce data-dependant partial regularization. To this end, we reformulate TSVM into a form with controllable regularization strength, which includes SVM and TSVM as special cases. Furthermore, we introduce a method of adaptive regularization that is data dependant and is based on the smoothness assumption. Experiments on a set of benchmark data sets indicate the promising results of the proposed work compared with state-of-the-art TSVM algorithms.

Subject Area: Unsupervised and Semi-supervised Learning

Spotlight presentation, Wednesday, 3:20.

W59 Canonical Time Warping for Alignment of Human Behavior

FENG ZHOU FERNANDO DE LA TORRE Carnegie Mellon University zhfe990gmail.com
ftorre@cs.cmu.edu

Alignment of time series is an important problem to solve in many scientific disciplines. In particular, temporal alignment of two or more subjects performing similar activities is a challenging problem due to the large temporal scale difference between human actions as well as the inter/intra subject variability. In this paper we present canonical time warping (CTW), an extension of canonical correlation analysis (CCA) for spatio-temporal alignment of the behavior between two subjects. CTW extends previous work on CCA in two ways: (i) it combines CCA with dynamic time warping for temporal alignment; and (ii) it extends CCA to allow local spatial deformations. We show CTW's effectiveness in three experiments: alignment of synthetic data, alignment of motion capture data of two subjects performing similar actions, and alignment of two people with similar facial expressions. Our results demonstrate that CTW provides both visually and qualitatively better alignment than state-of-the-art techniques based on dynamic time warping. Subject Area: Unsupervised and Semi-supervised Learning

W60 Fast Graph Laplacian Regularized Kernel Learning via Semidefinite-Quadratic-Linear Programming

Xiao-Ming WU	wxm007@ie.cuhk.edu.hk
Anthony Man-Cho So	manchoso@se.cuhk.edu.hk
Zhenguo Li	zgli@ie.cuhk.edu.hk
Shuo-Yen Robert Li	bobli@ie.cuhk.edu.hk
The Chinese University of Hong Kong	

Kernel learning is a powerful framework for nonlinear data modeling. Using the kernel trick, a number of problems have been formulated as semidefinite programs (SDPs). These include Maximum Variance Unfolding (MVU) (Weinberger et al., 2004) in nonlinear dimensionality reduction, and Pairwise Constraint Propagation (PCP) (Li et al., 2008) in constrained clustering. Although in theory SDPs can be efficiently solved, the high computational complexity incurred in numerically processing the huge linear matrix inequality constraints has rendered the SDP approach unscalable. In this paper, we show that a large class of kernel learning problems can be reformulated as semidefinite-quadratic-linear programs (SQLPs), which only contain a simple positive semidefinite constraint, a second-order cone constraint and a number of linear constraints. These constraints are much easier to process numerically, and the gain in speedup over previous approaches is at least of the order $m^{2.5}$, where m is the matrix dimension. Experimental results are also presented to show the superb computational efficiency of our approach.

Subject Area: Unsupervised and Semi-supervised Learning

Spotlight presentation, Wednesday, 3:20.

W61 Replicated Softmax: an Undirected Topic Model

RUSLAN SALAKHUTDINOVrsalakhu@cs.toronto.eduDepartment of Brain and Cognitive Sciences and CSAIL MITGEOFFREY HINTONhinton@cs.toronto.eduUniversity of Toronto

We show how to model documents as bags of words using family of two-layer, undirected graphical models. Each member of the family has the same number of binary hidden units but a different number of "softmax" visible units. All of the softmax units in all of the models in the family share the same weights to the binary hidden units. We describe efficient inference and learning procedures for such a family. Each member of the family models the probability distribution of documents of a specific length as a product of topic-specific distributions rather than as a mixture and this gives much better generalization than Latent Dirichlet Allocation for modeling the log probabilities of held-out documents. The low-dimensional topic vectors learned by the undirected family are also much better than LDA topic vectors for retrieving documents that are similar to a query document. The learned topics are more general than those found by LDA because precision is achieved by intersecting many general topics rather than by selecting a single precise topic to generate each word.

Subject Area: Unsupervised and Semi-supervised Learning

SAMUEL ROTA BULÒ srotabul@dsi.unive.it Università Ca' Foscari di Venezia MARCELLO PELILLO pelillo@dsi.unive.it Università Ca Foscari di Venezia

Hypergraph clustering refers to the process of extracting maximally coherent groups from a set of objects using high-order (rather than pairwise) similarities. Traditional approaches to this problem are based on the idea of partitioning the input data into a user-defined number of classes, thereby obtaining the clusters as a by-product of the partitioning process. In this paper, we provide a radically different perspective to the problem. In contrast to the classical approach, we attempt to provide a meaningful formalization of the very notion of a cluster and we show that game theory offers an attractive and unexplored perspective that serves well our purpose. Specifically, we show that the hypergraph clustering problem can be naturally cast into a non-cooperative multi-player "clustering game", whereby the notion of a cluster is equivalent to a classical game-theoretic equilibrium concept. From the computational viewpoint, we show that the problem of finding the equilibria of our clustering game is equivalent to locally optimizing a polynomial function over the standard simplex, and we provide a discrete-time dynamics to perform this optimization. Experiments are presented which show the superiority of our approach over state-of-the-art hypergraph clustering techniques.

Subject Area: Unsupervised and Semi-supervised Learning

W63 Learning to Hash with Binary Reconstructive Embeddings

BRIAN KULISkulis@eecs.berkeley.eduUC Berkeley EECS & ICSItrevor@eecs.berkeley.eduUC Berkeley EECS and ICSItrevor@eecs.berkeley.edu

Fast retrieval methods are increasingly critical for many large-scale analysis tasks, and there have been several recent methods that attempt to learn hash functions for fast and accurate nearest neighbor searches. In this paper, we develop an algorithm for learning hash functions based on explicitly minimizing the reconstruction error between the original distances and the Hamming distances of the corresponding binary embeddings. We develop a scalable coordinate-descent algorithm for our proposed hashing objective that is able to efficiently learn hash functions in a variety of settings. Unlike existing methods such as semantic hashing and spectral hashing, our method is easily kernelized and does not require restrictive assumptions about the underlying distribution of the data. We present results over several domains to demonstrate that our method outperforms existing stateof-the-art techniques.

Subject Area: Unsupervised and Semi-supervised Learning

Spotlight presentation, Wednesday, 5:10.

W64 Posterior vs Parameter Sparsity in Latent Variable Models

Joao Graca	joao.graca@l2f.inesc-id.pt
University of Lisbon	
Kuzman Ganchev	kuzman@cis.upenn.edu
University of Pennsylvania	
Ben Taskar	taskar@cis.upenn.edu
University of Pennsylvania	
Fernando Pereira	pereira@google.com
Google	

In this paper we explore the problem of biasing unsupervised models to favor sparsity. We extend the posterior regularization framework [8] to encourage the model to achieve posterior sparsity on the unlabeled training data. We apply this new method to learn first-order HMMs for unsupervised part-of-speech (POS) tagging, and show that HMMs learned this way consistently and significantly out-performs both EM-trained HMMs, and HMMs with a sparsity-inducing Dirichlet prior trained by variational EM. We evaluate these HMMs on three languages — English, Bulgarian and Portuguese — under four conditions. We find that our method always improves performance with respect to both baselines, while variational Bayes actually degrades performance in most cases. We increase accuracy with respect to EM by 2.5%-8.7% absolute and we see improvements even in a semisupervised condition where a limited dictionary is provided. **Subject Area:** Unsupervised and Semi-supervised Learning

Spotlight presentation, Wednesday, 5:10.

W65 Occlusive Components Analysis

JÖRG LÜCKEluecke@fias.uni-frankfurt.deFIAS, Goethe-University FrankfurtRICHARD TURNERturner@gatsby.ucl.ac.ukGatsby Computational NeuroscienceUniversity College LondonMARC HENNIGEShenniges@fias.uni-frankfurt.deFIAS, Goethe-University Frankurt

We study unsupervised learning in a probabilistic generative model for occlusion. The model uses two types of latent variables: one indicates which objects are present in the image, and the other how they are ordered in depth. This depth order then determines how the positions and appearances of the objects present, specified in the model parameters, combine to form the image. We show that the object parameters can be learnt from an unlabelled set of images in which objects occlude one another. Exact maximum-likelihood learning is intractable. However, we show that tractable approximations to Expectation Maximization (EM) can be found if the training images each contain only a small number of objects on average. In numerical experiments it is shown that these approximations recover the correct set of object parameters. Experiments on a novel version of the bars test using colored bars, and experiments on more realistic data, show that the algorithm performs well in extracting the generating causes. Experiments based on the standard bars benchmark test for object learning show that the algorithm performs well in comparison to other recent component extraction approaches. The model and the learning algorithm thus connect research on occlusion with the research field of multiple-cause component

extraction methods. Subject Area: Probabilistic Models and Methods

W66 Bayesian Source Localization with the Multivariate Laplace Prior

Marcel Van Gerven	marcel.van.gerven@gmail.com
Botond Cseke	B.Cseke@cs.ru.nl
ICIS	
Robert Oostenveld	$\tt r.oostenveld@fcdonders.ru.nl$
DCCN	
Tom Heskes	t.heskes@science.ru.nl
Radboud University Nijmegen	

We introduce a novel multivariate Laplace (MVL) distribution as a sparsity promoting prior for Bayesian source localization that allows the specification of constraints between and within sources. We represent the MVL distribution as a scale mixture that induces a coupling between source variances instead of their means. Approximation of the posterior marginals using expectation propagation is shown to be very efficient due to properties of the scale mixture representation. The computational bottleneck amounts to computing the diagonal elements of a sparse matrix inverse. Our approach is illustrated using a mismatch negativity paradigm for which MEG data and a structural MRI have been acquired. We show that spatial coupling leads to sources which are active over larger cortical areas as compared with an uncoupled prior.

Subject Area: Probabilistic Models and Methods

W67 A Bayesian Model for Simultaneous Image Clustering, Annotation and Object Segmentation

Lan Du	lan.du@duke.edu
Duke University, ECE Department	5
Lu Ren	lr@ee.duke.edu
Duke ECE	
David Dunson	dunson@stats.duke.edu
Duke	
LAWRENCE CARIN	lcarin@ee.duke.edu
Duke ECE	

A non-parametric Bayesian model is proposed for processing multiple images. The analysis employs image features and, when present, the words associated with accompanying annotations. The model clusters the images into classes, and each image is segmented into a set of objects, also allowing the opportunity to assign a word to each object (localized labeling). Each object is assumed to be represented as a heterogeneous mix of components, with this realized via mixture models linking image features to object types. The number of image classes, number of object types, and the characteristics of the object-feature mixture models are inferred non-parametrically. To constitute spatially contiguous objects, a new logistic stick-breaking process is developed. Inference is performed efficiently via variational Bayesian analysis, with example results presented on two image databases.

Subject Area: Probabilistic Models and Methods

W68 Bayesian Nonparametric Models on Decomposable Graphs

FRANCOIS CARON Francois.Caron@inria.fr INRIA Bordeaux ARNAUD DOUCET arnaud@cs.ubc.ca Institute of Statistical Mathematics

Over recent years Dirichlet processes and the associated Chinese restaurant process (CRP) have found many applications in clustering while the Indian buffet process (IBP) is increasingly used to describe latent feature models. In the clustering case, we associate to each data point a latent allocation variable. These latent variables can share the same value and this induces a partition of the data set. The CRP is a prior distribution on such partitions. In latent feature models, we associate to each data point a potentially infinite number of binary latent variables indicating the possession of some features and the IBP is a prior distribution on the associated infinite binary matrix. These prior distributions are attractive because they ensure exchangeability (over samples). We propose here extensions of these models to decomposable graphs. These models have appealing properties and can be easily learned using Monte Carlo techniques. **Subject Area:** Probabilistic Models and Methods

W69 Efficient Moments-based Permutation Tests

Chunxiao Zhou	czhou4@gmail.edu
UIUC	
Huixia Judy Wang	wang@stat.ncsu.edu
North Carolina State University	
Yongmei Michelle Wang	ymw@illinois.edu
University of Illinois at Urbana-C	hampaign

In this paper, we develop an efficient moments-based permutation test approach to improve the system's efficiency by approximating the permutation distribution of the test statistic with Pearson distribution series. This approach involves the calculation of the first four moments of the permutation distribution. We propose a novel recursive method to derive these moments theoretically and analytically without any permutation. Experimental results using different test statistics are demonstrated using simulated data and real data. The proposed strategy takes advantage of nonparametric permutation tests and parametric Pearson distribution approximation to achieve both accuracy and efficiency. **Subject Area:** Probabilistic Models and Methods

W70 Localizing Bugs in Program Executions with Graphical Models

Laura Dietz	dietz@mpi-inf.mpg.de
Max Planck Institute for Infor	
VALENTIN DALLMEIER	dallmeier@st.cs.uni-sb.de
Andreas Zeller	zeller@st.cs.uni-sb.de
Saarland University	
Tobias Scheffer	<pre>scheffer@cs.uni-potsdam.de</pre>
University of Potsdam	

We devise a graphical model that supports the process of debugging software by guiding developers to code that is likely to contain defects. The model is trained using execution traces of passing test runs; it reflects the distribution over transitional patterns of code positions. Given a failing test case, the model determines the least likely transitional pattern in the execution trace. The model is designed such that Bayesian inference has a closed-form solution. We evaluate the Bernoulli graph model on data of the software projects AspectJ and Rhino.

Subject Area: Probabilistic Models and Methods

W71 Learning transport operators for image manifolds

BENJAMIN CULPEPPER bjc@cs.berkeley.edu UC Berkelev BRUNO OLSHAUSEN baolshausen@berkeley.edu University of California, Berkeley

We describe a method for learning a group of continuous transformation operators to traverse smooth nonlinear manifolds. The method is applied to model how natural images change over time and scale. The group of continuous transform operators is represented by a basis that is adapted to the statistics of the data so that the infinitesimal generator for a measurement orbit can be produced by a linear combination of a few basis elements. We illustrate how the method can be used to efficiently code time-varying images by describing changes across time and scale in terms of the learned operators.

Subject Area: Unsupervised and Semi-supervised Learning

W72 Variational Gaussian-process factor analysis for modeling spatio-temporal data

Jaakko Luttinen	jluttine@cis.hut.fi
Alexander Ilin	alexander.ilin@tkk.fi
Helsinki University of Technology	

We present a probabilistic latent factor model which can be used for studying spatiotemporal datasets. The spatial and temporal structure is modeled by using Gaussian process priors both for the loading matrix and the factors. The posterior distributions are approximated using the variational Bayesian framework. High computational cost of Gaussian process modeling is reduced by using sparse approximations. The model is used to compute the reconstructions of the global sea surface temperatures from a historical dataset. The results suggest that the proposed model can outperform the state-of-the-art reconstruction systems.

Subject Area: Probabilistic Models and Methods

Spotlight presentation, Wednesday, 3:20.

W73 An LP View of the M-best MAP problem

MENACHEM FROMER and AMIR GLOBERSON, Hebrew University. Subject Area: Probabilistic Models and Methods Oral presentation, Wednesday, 3:00pm. See abstract, page 157.

W74 Nonparametric Bayesian Models for Unsupervised Event Coreference Resolution

Cosmin Bejan	ady@hlt.utdallas.edu
University of Texas at Dallas	
Matthew Titsworth	matthew@languagecomputer.com
Andrew Hickl	andy@languagecomputer.com
Language Computer Corporation	
Sanda Harabagiu	sanda@hlt.utdallas.edu
University of Texas at Dallas	

We present a sequence of unsupervised, nonparametric Bayesian models for clustering complex linguistic objects. In this approach, we consider a potentially infinite number of features and categorical outcomes. We evaluate these models for the task of within- and cross-document event coreference on two corpora. All the models we investigated show significant improvements when compared against an existing baseline for this task. **Subject Area:** Probabilistic Models and Methods

W75 Sharing Features among Dynamical Systems with Beta Processes

EMILY FOX, Massachusetts Institute of Technology, ERIK SUDDERTH, Brown University, MICHAEL JORDAN, UC Berkeley, and ALAN WILLSKY, MIT. Subject Area: Probabilistic Models and Methods Oral presentation, Wednesday, 4:30pm. See abstract, page 160.

W76 Convex Relaxation of Mixture Regression with Efficient Algorithms

novi.quad@gmail.com
tiberio.caetano@nicta.com.au
john.lim@rsise.anu.edu.au
dale@cs.ualberta.ca

We develop a convex relaxation of maximum a posteriori estimation of a mixture of regression models. Although our relaxation involves a semidefinite matrix variable, we reformulate the problem to eliminate the need for general semidefinite programming. In particular, we provide two reformulations that admit fast algorithms. The first is a max-min spectral reformulation exploiting quasi-Newton descent. The second is a min-min reformulation consisting of fast alternating steps of closed-form updates. We evaluate the methods against Expectation-Maximization in a real problem of motion segmentation from video data.

Subject Area: Probabilistic Models and Methods

W77 FACTORIE: Probabilistic Programming via Imperatively Defined Factor Graphs

Andrew McCallum	mccallum@cs.umass.edu
KARL SCHULTZ	kschultz@cs.umass.edu
SAMEER SINGH	sameer@cs.umass.edu
University of Massachusetts,	Amherst

Discriminatively trained undirected graphical models have had wide empirical success, and there has been increasing interest in toolkits that ease their application to complex relational data. The power in relational models is in their repeated structure and tied parameters; at issue is how to define these structures in a powerful and flexible way. Rather than using a declarative language, such as SQL or first-order logic, we advocate using an imperative language to express various aspects of model structure, inference, and learning. By combining the traditional, declarative, statistical semantics of factor graphs with imperative definitions of their construction and operation, we allow the user to mix declarative and procedural domain knowledge, and also gain significant efficiencies. We have implemented such imperatively defined factor graphs in a system we call Factorie, a software library for an object-oriented, strongly-typed, functional language. In experimental comparisons to Markov Logic Networks on joint segmentation and coreference, we find our approach to be 3-15 times faster while reducing error by 20-25%-achieving a new state of the art.

Subject Area: Probabilistic Models and Methods

W78 Particle-based Variational Inference for Continuous Systems

Alexander Ihler Andrew Frank Padhraic Smyth UC Irvine ihler@ics.uci.edu
ajfrank@ics.uci.edu
smyth@ics.uci.edu

Since the development of loopy belief propagation, there has been considerable work on advancing the state of the art for approximate inference over distributions defined on discrete random variables. Improvements include guarantees of convergence, approximations that are provably more accurate, and bounds on the results of exact inference. However, extending these methods to continuous-valued systems has lagged behind. While several methods have been developed to use belief propagation on systems with continuous values, they have not as yet incorporated the recent advances for discrete variables. In this context we extend a recently proposed particle-based belief propagation algorithm to provide a general framework for adapting discrete message-passing algorithms to perform inference in continuous systems. The resulting algorithms behave similarly to their purely discrete counterparts, extending the benefits of these more advanced inference techniques to the continuous domain. **Subject Area:** Probabilistic Models and Methods

W79 Approximating MAP by Compensating for Structural Relaxations

Arthur Choi	aychoi@cs.ucla.edu
UCLA	
Adnan Darwiche	darwiche@cs.ucla.edu
University of California	

We introduce a new perspective on approximations to the maximum a posteriori (MAP) task in probabilistic graphical models, that is based on simplifying a given instance, and then tightening the approximation. First, we start with a structural relaxation of the original model. We then infer from the relaxation its deficiencies, and compensate for them. This perspective allows us to identify two distinct classes of approximations. First, we find that max-product belief propagation can be viewed as a way to compensate for a relaxation, based on a particular idealized case for exactness. We identify a second approach to compensation that is based on a more refined idealized case, resulting in a new approximation with distinct properties. We go on to propose a new class of algorithms that, starting with a relaxation, iteratively yields tighter approximations. **Subject Area:** Probabilistic Models and Methods

W80 Accelerating Bayesian Structural Inference for Non-Decomposable Gaussian Graphical Models

BABACK MOGHADDAM, Caltech, BENJAMIN MARLIN, EMTIYAZ KHAN and KEVIN MURPHY, University of British Columbia. Subject Area: Probabilistic Models and Methods Oral presentation, Thursday, 9:30am. See abstract, page 203.

W81 An Infinite Factor Model Hierarchy Via a Noisy-Or Mechanism

AARON COURVILLE DOUGLAS ECK YOSHUA BENGIO University of Montreal aaron.courville@gmail.com
eckdoug@iro.umontreal.ca
bengioy@iro.umontreal.ca

The Indian Buffet Process is a Bayesian nonparametric approach that models objects as arising from an infinite number of latent factors. Here we extend the latent factor model framework to two or more unbounded layers of latent factors. From a generative perspective, each layer defines a conditional *factorial* prior distribution over the binary latent variables of the layer below via a noisy-or mechanism. We explore the properties of the model with two empirical studies, one digit recognition task and one music tag data experiment.

Subject Area: Probabilistic Models and Methods

196

W82 Multi-Label Prediction via Sparse Infinite CCA

PIYUSH RAI HAL DAUME University of Utah piyush@cs.utah.edu hal@cs.utah.edu

Canonical Correlation Analysis (CCA) is a useful technique for modeling dependencies between two (or more) sets of variables. Building upon the recently suggested probabilistic interpretation of CCA, we propose a nonparametric, fully Bayesian framework that can automatically select the number of correlation components, and effectively capture the sparsity underlying the projections. In addition, given (partially) labeled data, our algorithm can also be used as a (semi)supervised dimensionality reduction technique, and can be applied to learn useful predictive features in the context of learning a set of related tasks. Experimental results demonstrate the efficacy of the proposed approach for both CCA as a stand-alone problem, and when applied to multi-label prediction.

Subject Area: Probabilistic Models and Methods

W83 Submanifold density estimation

ARKADAS OZAKIN ARKADAS@GMAIL.COM Georgia Tech Research Institute ALEXANDER GRAY agray@cc.gatech.edu Georgia Institute of Technology

Kernel density estimation is the most widely-used practical method for accurate nonparametric density estimation. However, long-standing worst-case theoretical results showing that its performance worsens exponentially with the dimension of the data have quashed its application to modern high-dimensional datasets for decades. In practice, it has been recognized that often such data have a much lower-dimensional intrinsic structure. We propose a small modification to kernel density estimation for estimating probability density functions on Riemannian submanifolds of Euclidean space. Using ideas from Riemannian geometry, we prove the consistency of this modified estimator and show that the convergence rate is determined by the intrinsic dimension of the submanifold. We conclude with empirical results demonstrating the behavior predicted by our theory.

Subject Area: Probabilistic Models and Methods

W84 Heterogeneous multitask learning with joint sparsity constraints

XIAOLIN YANG SEYOUNG KIM ERIC XING Carnegie Mellon University

xiaolin.yang@gmail.com
sssykim@cs.cmu.edu
epxing@cs.cmu.edu

Multitask learning addressed the problem of learning related tasks whose information can be shared each other. Traditional problem usually deal with homogeneous tasks such as regression, classification individually. In this paper we consider the problem learning multiple related tasks where tasks consist of both continuous and discrete outputs from a common set of input variables that lie in a high-dimensional space. All of the tasks are related in the sense that they share the same set of relevant input variables, but the amount of influence of each input on different outputs may vary. We formulate this problem as a combination of linear regression and logistic regression and model the joint sparsity as L1/Linf and L1/L2-norm of the model parameters. Among several possible applications, our approach addresses an important open problem in genetic association mapping, where we are interested in discovering genetic markers that influence multiple correlated traits jointly. In our experiments, we demonstrate our method in the scenario of association mapping, using simulated and asthma data, and show that the algorithm can effectively recover the relevant inputs with respect to all of the tasks. **Subject Area:** Applications

W85 Modeling the spacing effect in sequential category learning

HONGJING LU hongjing@ucla.edu Dept. of Psychology & Statistics,UCLA MATTHEW WEIDEN mweiden@ucla.edu Dept. of Psychology, UCLA ALAN YUILLE yuille@stat.ucla.edu UCLA

We develop a Bayesian sequential model for category learning. The sequential model updates two category parameters, the mean and the variance, over time. We define conjugate temporal priors to enable closed form solutions to be obtained. This model can be easily extended to supervised and unsupervised learning involving multiple categories. To model the spacing effect, we introduce a generic prior in the temporal updating stage to capture a learning preference, namely, less change for repetition and more change for variation. Finally, we show how this approach can be generalized to efficiently performmodel selection to decide whether observations are from one or multiple categories.

Subject Area: Cognitive Science

W86 Hierarchical Learning of Dimensional Biases in Human Categorization

KATHERINE HELLERkheller@gmail.comUniversity of Cambridgeasanborn@gatsby.ucl.ac.ukADAM SANBORNasanborn@gatsby.ucl.ac.ukNICK CHATERn.chater@ucl.ac.ukUniversity College London

Existing models of categorization typically represent to-be-classified items as points in a multidimensional space. While from a mathematical point of view, an infinite number of basis sets can be used to represent points in this space, the choice of basis set is psychologically crucial. People generally choose the same basis dimensions, and have a strong preference to generalize along the axes of these dimensions, but not "diagonally". What makes some choices of dimension special? We explore the idea that the dimensions used by people echo the natural variation in the environment. Specifically, we present a rational model that does not assume dimensions, but learns the same type of dimensional generalizations that people display. This bias is shaped by exposing the model to many categories with a structure hypothesized to be like those which children encounter. Our model can be viewed as a type of transformed Dirichlet process mixture model, where it is the learning of the base distribution of the Dirichlet process which allows dimensional generalization. The learning behaviour of our model captures the developmental shift from roughly "isotropic" for children to the axis-aligned generalization that adults show. **Subject Area:** Cognitive Science

Spotlight presentation, Wednesday, 5:10.

W87 Help or Hinder: Bayesian Models of Social Goal Inference

Tomer Ullman	tomerdo@gmail.com
Chris Baker	clbaker@mit.edu
Owen Macindoe	skrawl@gmail.com
Owain Evans	owain@mit.edu
Noah Goodman	ndg@mit.edu
Joshua Tenenbaum	jbt@mit.edu
Massachusetts Institute of Technology	

Everyday social interactions are heavily influenced by our snap judgments about others' goals. Even young infants can infer the goals of intentional agents from observing how they interact with objects and other agents in their environment: e.g., that one agent is 'helping' or 'hindering' another's attempt to get up a hill or open a box. We propose a model for how people can infer these social goals from actions, based on inverse planning in multiagent Markov decision problems (MDPs). The model infers the goal most likely to be driving an agent's behavior by assuming the agent acts approximately rationally given environmental constraints and its model of other agents present. We also present behavioral evidence in support of this model over a simpler, perceptual cue-based alternative. **Subject Area:** Cognitive Science

Demonstrations (7:30–11:59pm)

D8 An FPGA-based Multi-Algorithm Learning and Classification Engine

HARI CADAMBI cadambi@nec-labs.com NEC Laboratories America We demonstrate a prototype of a multi-algorithm learning and classification engine on a state-of-the-art FPGA. The architecture can accelerate training and classification for CNN, SVM and GVLQ. It can also do Supervised Semantic Indexing (SSI). We identify common underlying computations in these various algorithms. Specifically, we map them to variants of matrix-matrix and matrix-vector operations, and architect a many-core coprocessor to execute a massively parallelized version of these operations using fine-grained, lightweight threads. A 512-core prototype will be demonstrated on a Xilinx SX240T FPGA on an off-the-shelf PCI card.

D9 Curious George: An Attentive Semantic Robotic Vision System

ANKUR GUPTA agupta@cs.ubc.ca University of British Columbia

Our robot, Curious George, has participated in the Semantic Robot Vision Challenge (SRVC): http://www.semantic-robot-vision-challenge.org/ a contest for robots performing a fully automated visual scavenger hunt within a realistic environment. The contest features completely autonomous acquisition of training data from the World Wide Web (WWW) (sources such as Google Images for example), that is no human labeled imagery is provided. Also, the visual categories chosen by the contest organizers are announced only at the time of the contest, and they are drawn from an unrestricted set within indoor objects, so contestants are challenged with perhaps the most general vision task amongst major contests held in the Computer Vision community. The evaluation phase of the contest requires robots to use the visual appearance models they have learned from the web data to explore a realistic contest setting which contains many of the target objects, and to recognize as many objects as possible. Our entry, Curious George, has won the SRVC for both of the two years in which it has been held.

D10 Deformed Lattice Detection in Real-World Images Using a Novel Mean-Shift Belief

MINWOO PARK mipark@cse.psu.edu Penn State Univ

The software library is based on our paper, "Deformed Lattice Detection in Real-World Images Using Mean-Shift Belief Propagation", TPAMI special issue OCTOBER 2009 (Vol. 31, No. 10). It detects, automatically, various repetitive patterns as deformed 2D lattices in real photos with cluttered background. Examples include windows on buildings, fruits on display, draped cloth, or see-through wiry structures. The lattice detection problem is formulated as a spatial, multi-target tracking problem, solved within an MRF framework using a novel mean-shift belief propagation method. The model is evaluated and compared with state of the art existing methods on a wide range of images that contain near-regular textures (NRT).

D11 Monte: An Interactive Ssytem for Massively Parallel Probabilistic Programming

VIKASH MANSINGHKA vkm@mit.edu Massachusetts Institute of Technology

We demonstrate Monte, a massively parallel implementation of the probabilistic programming language Church. Using Monte, a probabilistic modeler or machine learning researcher can specify state-of-the-art nonparametric Bayesian models in just 10-20 lines of code, and automatically perform inference on them in the presence of arbitrary datasets. Monte also allows a probabilistic modeler to exploit Amazon's Elastic Compute Cloud (EC2) to perform massively parallel inference, significantly increasing the model complexity and dataset scale that Monte can handle over serial implementations of probabilistic programming languages. Church is a probabilistic programming language, based on Lisp, that enables compact specification of all computable probabilistic models and supports universal Bayesian inference via Markov chain Monte Carlo. Probabilistic programming languages have gained substantial attention in the NIPS community, serving as a representation language for probabilistic machine learning and the basis of a widely attended NIPS 2008 Workshop. However, very few implementations of probabilistic programming languages have been mature enough to support reliable interactive use. The Monte implementation of Church, developed at Navia Systems, Inc., is one of the first such systems. The best evidence of its robustness and interactivity lies in the compactness of Monte representations of state-of-the-art probabilistic models (see Figure 1) and in the simplicity of its use on a single processor workstation (see Figure 2). Users of our demonstration system will be able to implement state-of-the-art probabilistic models from the current NIPS conference and immediately perform inference on them in real time using Monte's engine, collecting and analyzing posterior samples of interest. Figure 2 shows one example of this use pattern in the context of a Dirichlet process Mixture Model with binary data.

D12 Nengo: Large-Scale Integrated Neural Modeling

CHRIS ELIASMITH celiasmith@uwaterloo.ca U of Waterloo

To develop large-scale neural models, we need software tools that not only support efficient simulation of hundreds of thousands of neurons, but also provide a structure for constructing and organizing these large-scale models. This demonstration showcases Nengo, our neural modelling system that integrates a wide variety of neural modelling It supports user-definable single cell and synaptic models, as well as techniques. providing several common default models. More importantly, it allows for the high-level design of complex neural circuitry by allowing users to specify representational properties of neural groups and the desired transformations that should be caused by synaptic connections between them. This is a unique feature of the software. Furthermore, the synaptic weights can then either be directly solved for using the Neural Engineering Framework (NEF) methods described in Eliasmith & Anderson (2003) or the weights can be generated using user-defined local learning rules. These features can be intuitively exploited through the GUI, or extensively customized through an interpreted command line and scripting. Nengo allows users to create large-scale models by focusing on this higher-level description. This method has been successfully used to construct models of: working memory using 25-dimensional neural integrators; motor

control using Kalman filters; escape and swimming control in zebrafish; the translational vestibular ocular reflex; symbolic decision-making and rule-following in the basal ganglia; and others. Uniquely, Nengo provides direct control over the efficiency and accuracy of the neural simulation by allowing for alternate neural models, switching from precise spiking models down to rate models and even to an abstract mode for directly computing the underlying transformations. Other efficiencies are implemented as well, including support for synaptic weight matricies of low rank and user-definable mathematics libraries. These aspects allow for fast simulation of large systems, making Nengo ideal for exploring interactive behaviour of complex neural models.

D13 Real-Time Adaptive Face Recognition

BRANISLAV KVETON Intel Corporation

branislav.kveton@intel.com

High-accuracy face recognition with a sufficiently high recall is a challenging problem. This demo shows how to achieve this efficiently with just a few labeled examples, streams of unlabeled data, and state-of-the-art machine learning techniques. In particular, we show an online semi-supervised learning algorithm that tracks the manifold of human faces in real time and based on it predicts whether the observed face belongs to one of the labeled individuals or not.

D14 Robust Speaker Recognition Using Approximate Bayesian Inference

CIIRA MAINA

cm527@drexel.edu

Drexel University

This is a demonstration of the application of variational Bayesian inference to the problem of joint speech enhancement and speaker identification. This approach to robust speaker identification illustrates the power of Bayesian inference techniques when applied to signal processing problems of interest. Speech enhancement and speaker identification are important signal processing problems which continue to attract research interest. These problems have traditionally been studied separately but we believe that there are a number of advantages to considering them jointly within a Bayesian framework. A number of authors have presented enhancement algorithms that employ speech and noise priors and use approximate Bayesian techniques for inference. However these prior speech models fail to capture speaker dependent information which could further improve performance at the cost of increased complexity. Also, current approaches to speaker identification mainly rely on directly modeling the speech feature vectors of the speakers to be identified and using clean speech to learn parameters of these models. This approach makes these methods sensitive to noise and as a result these systems do not perform well in real acoustic environments where noise is unavoidable. Furthermore, noise encountered in everyday acoustic environments may include unwanted speech sources with spectral characteristics similar to the desired signal. By considering speech enhancement and identification jointly within a Bayesian framework we can now take advantage of rich speaker dependent speech priors in the enhancement task and appropriately model noise in the identification task.

Thursday, December 10th

Oral Session (8:30–9:50am): "Bayesian Analysis"

8:30am Invited Talk: Bayesian Analysis of Markov Chains

PERSI DIACONIS diaconis@math.stanford.edu Stanford

Suppose we observe data and want to test if it comes from a Markov chain. If it does, we may want to estimate the transition operator. Working in a Bayesian way, we have to specify priors and compute posteriors. Interesting things happen if we want to put priors on reversible Markov chains. There are useful connections with reinforced random walk (work with Silke Rolles). On large-scale application to protein folding will be described. More generally, these problems arise in approximating a dynamical system by a Markov chain. For continuous state spaces, the usual conjugate prior analysis breaks down. Thesis work of Wai Liu (Stanford) gives useful families of priors where computations are "easy." These seem to work well in test problems and can be proved consistent.

Persi Diaconis is Professor of Mathematics and Statistics at Stanford University. He works in the mathematics and practice of Bayesian statistics (conjugacy, consistency properties of Dirichlet priors, deFinetti's theorem). In rates of convergence for Markov chains to stationality (how many times to shuffle cards, understanding the Metropolis algorithm, geometric theory of Markov chains). He also works in combinatorics, group theory, and random matrix theory and in the philosophy of probability.

9:30am Accelerating Bayesian Structural Inference for Non-Decomposable Gaussian Graphical Models

BABACK MOGHADDAMbaback@jpl.nasa.govCaltechbmarlin@cs.ubc.caBENJAMIN MARLINbmarlin@cs.ubc.caEMTIYAZ KHANemtiyaz@cs.ubc.caKEVIN MURPHYmurphyk@cs.ubc.caUniversity of British Columbia

In this paper we make several contributions towards accelerating approximate Bayesian structural inference for non-decomposable GGMs. Our first contribution is to show how to efficiently compute a BIC or Laplace approximation to the marginal likelihood of non-decomposable graphs using convex methods for precision matrix estimation. This optimization technique can be used as a fast scoring function inside standard Stochastic Local Search (SLS) for generating posterior samples. Our second contribution is a novel framework for efficiently generating large sets of high-quality graph topologies without performing local search. This graph proposal method, which we call "Neighborhood Fusion" (NF), samples candidate Markov blankets at each node using sparse regression techniques. Our final contribution is a hybrid method combining the complementary strengths of NF and SLS. Experimental results in structural recovery and prediction tasks demonstrate that NF and hybrid NF/SLS out-perform state-of-the-art local search

methods, on both synthetic and real-world datasets, when realistic computational limits are imposed.

Subject Area: Probabilistic Models and Methods

Oral Session (10:30–11:50am): "Neural Modeling and Imaging"

10:30am Invited Talk: Optogenetics: Development and Application

KARL DEISSEROTH Stanford University deissero@stanford.edu

In 1979, Francis Crick delineated the major challenges facing neuroscience and called for a technology by which all neurons of just one type could be controlled, ";leaving the others more or less unaltered";. A new set of technologies now called optogenetics, synthesizing microbial opsins and solid-state optics, has achieved the goal of millisecondprecision bidirectional control of defined cell types in freely behaving mammals. First, we found that neurons expressing Channel Rhodopsin-2 can fire light-triggered action potentials with millisecond precision. Second, we found that neurons targeted to express the light-activated chloride pump halorhodopsin can be hyperpolarized and inhibited from firing action potentials when exposed to vellow light in intact tissue. Third, we employed genomic strategies to discover and adapt for neuroscience a third major optogenetic tool, namely a cation channel with action spectrum significantly redshifted relative to ChR2, to allow tests of the combinatorial interaction of cell types in circuit computation or behavior. Fourth, we have created transgenic mice expressing microbial opsins in the brain. Finally, we have developed integrated fiberoptic and solid-state optical approaches to provide the complementary technology to allow specific cell types, even deep within the brain, to be controlled in freely behaving mammals. We also are now enhancing opsin function for long-term in vivo applications, and applying fast optical control and optical imaging to animal models of depression, Parkinson':s Disease, and altered social behavior relevant to autism, with insights into both normal circuit function and disease mechanisms beginning to emerge.

Professor Deisseroth received his bachelor's degree from Harvard in 1992, his PhD from Stanford in 1998, and his MD from Stanford in 2000. He completed medical internship and adult psychiatry residency at Stanford, and he was board-certified by the American Board of Psychiatry and Neurology in 2006. He joined the faculty at Stanford in Bioengineering and Psychiatry on January 1, 2005. As a bioengineer focused on neuroengineering, he has launched an effort to map neural circuit dynamics in psychiatric disease on the millisecond timescale, and his group at Stanford has developed optical neuroengineering technologies for noninvasive imaging and control of brain circuits, as they operate within living intact tissue. His work on optical control of neural circuits has launched a new field called ";optogenetics";. Professor Deisseroth has received many major awards including the NIH Director';s Pioneer Award, the Presidential Early Career Award for Science and Engineering (PECASE), the McKnight Foundation Technological Innovations in Neuroscience Award, the Larry Katz Prize in Neurobiology, the Schuetze Award in Neuroscience, the Whitehall Foundation Award, the Charles E. Culpeper Scholarship in Medical Science Award, the Klingenstein Fellowship Award and the Robert H. Ebert Clinical Scholar Award.

11:30am Time-rescaling methods for the estimation and assessment of non-Poisson neural encoding models

JONATHAN PILLOW pillow@mail.utexas.edu University of Texas at Austin

Recent work on the statistical modeling of neural responses has focused on modulated renewal processes in which the spike rate is a function of the stimulus and recent spiking history. Typically, these models incorporate spike-history dependencies via either: (A) a conditionally-Poisson process with rate dependent on a linear projection of the spike train history (e.g., generalized linear model); or (B) a modulated non-Poisson renewal process (e.g., inhomogeneous gamma process). Here we show that the two approaches can be combined, resulting in a *conditional renewal* (CR) model for neural spike trains. This model captures both real and rescaled-time effects, and can be fit by maximum likelihood using a simple application of the time-rescaling theorem [1]. We show that for any modulated renewal process model, the log-likelihood is concave in the linear filter parameters only under certain restrictive conditions on the renewal density (ruling out many popular choices, e.g. gamma with $\kappa \neq 1$), suggesting that real-time history effects are easier to estimate than non-Poisson renewal properties. Moreover, we show that goodnessof-fit tests based on the time-rescaling theorem [1] quantify relative-time effects, but do not reliably assess accuracy in spike prediction or stimulus-response modeling. We illustrate the CR model with applications to both real and simulated neural data. Subject Area: Neuroscience

Mini Symposia (1:30–4:30pm)

MS1 Assistive Machine Learning for People with Disabilities

Fernando Perez-Cruz	fernandop@ieee.org
University Carlos III in Madrid	
Emilio Parrado-Hernandez	emipar@tsc.uc3m.es
University Carlos III de Madrid	
David Hardoon	drhardoon@i2r.a-star.edu.sg
Institute for Infocomm Research	
Jaisiel Madrid-Sanchez	jmadrid@technosite.es
Fundosa Technosite	

Nowadays, there are massive amounts of heterogeneous electronic information available on the Web. People with disabilities, among other groups potentially influenced by the digital gap, face great barriers when trying to access information. Sometimes their disability makes their interaction the ICT environment (eg., computers, mobile phones, multimedia players and other hardware devices) more difficult. Furthermore, the contents are delivered in such formats that cannot be accessed by people with disability and the elderly. The challenge for their complete integration in information society has to be analyzed from different technology approaches. Recent developments in Machine Learning are improving the way people with disabilities access to digital information resources. From the hardware perspective, Machine Learning can be a core part for the correct design of accessible interaction systems of such users with computers (such as BCI). From the contents perspective, Machine Learning can provide tools to adapt contents (for instance changing the modality in which it is accessed) to users with special needs. From the users' perspective, Machine Learning can help constructing a good user modeling, as well as the particular context in which the information is accessed.

Location: Regency B

MS2 Causality and Time Series Analysis

FLORIN POPESCU	florin.popescu@first.fraunhofer.de
Fraunhofer FIRST	
ISABELLE GUYON	guyon@clopinet.com
Clopinet	
Guido Nolte	guido.nolte@first.fraunhofer.de

This symposium addresses a topic that has spurred vigorous scientific debate of late in the fields of neuroscience and machine learning: causality in time-series data. In neuroscience, causal inference in brain signal activity (EEG, MEG, fMRI, etc.) is challenged by relatively rough prior knowledge of brain connectivity and by sensor limitations (mixing of sources). On the machine learning side, as the Causality workshop last year's NIPS conference has evidenced for static (non-time series) data, there are issues of whether or not graphical models (directed acyclic graphs) pioneered by Judea Pearl, Peter Spirtes, and others can reliably provide a cornerstone of causal inference, whereas in neuroscience there are issues of whether Granger type causality inference is appropriate given the source mixing problem, traditionally addressed by ICA methods. Further topics, yet to be fully explored, are non-linearity, non-Gaussianity and full causal graph inference in high-dimensional time series data. Many ideas in causality research have been developed by and are of direct interest and relevance to researchers from fields beyond ML and neuroscience: economics (i.e. the Nobel Prize winning work of the late Clive Granger, which we will pay tribute to), process and controls engineering, sociology, etc. Despite the long-standing challenges of time-series causality, both theoretical and computational, the recent emergence of cornerstone developments and efficient computational learning methods all point to the likely growth of activity in this seminal topic. Along with the stimulating discussion of recent research on time-series causality, we will present and highlight time-series datasets added to the Causality Workbench, which have grown out of last year's Causality challenge and NIPS workshop, some of which are neuroscience related.

Location: Regency A

MS3 Machine Learning for Sustainability

J. Zico Kolter	kolter@cs.stanford.edu
Stanford University	
THOMAS DIETTERICH	tgd@cs.orst.edu
Oregon State University	
Andrew Ng	ang@cs.stanford.edu
Stanford University	

The world has a sustainability problem. Humans currently consume an average of 16TW of power (and rising), more than 86% of which comes from (unsustainable) fossil fuels. There is a range of estimates as to when this supply will run out, but this is a scenario that may well happen within our lifetimes. Even more pressing is the effect that such fuels have on our climate: given no attempts to reduce the world's fossil fuel usage, even the most conservative climate models predict that the world temperature will increase by over five degrees (Fahrenheit) in the next 90 years, an increase that could cause ecological disasters on a global scale. Building a sustainable infrastructure for energy and ecosystems is shaping up to be one of the grand scientific and political challenges of the 21st century. Furthermore, there is a growing consensus that many aspects of sustainability are fundamentally information systems problems, tasks where machine learning can play a significant role. This mini-symposium will bring together leading researchers with both machine learning backgrounds and energy/sustainability backgrounds to address the question: how can machine learning help address the world's sustainability problem? The mini-symposium will also seek to answer: What is the current state of work directed at sustainability, energy, and ecology in the machine learning, operations research, and optimization communities? What are the primary scientific and technical challenges in information processing for sustainability? And finally, what are (and what aren't) areas where machine learning can make a genuine impact on the science of sustainability? Because this is an emerging field of research, the talks at this symposium will aimed at the general NIPS audience. There is a growing number of researchers working in sustainability, but even more broadly, we think that such problems have the potential to advance basic machine learning in a manner similar to other important applications, such as computer vision, natural language processing, and computational biology. Sustainability problems offer an equally rich set of domains, and solutions to these problems will have a genuine impact on the world.

Location: Regency D

MS4 Machine Learning in Computational Biology

Yanjun Qi	yanjun@nec-labs.com
NEC Labs America	
Gal Chechik	gal@ai.stanford.edu
Google	
Alexander Zien	Alexander.Zien@lifebiosystems.com
LIFE Biosystems GmbH	
Tomer Hertz	hertz@microsoft.com
Fred Hutchnison Cancer Research	Center
William Noble	noble@gs.washington.edu
University of Washington	

The field of computational biology has seen dramatic growth over the past fewyears, both in terms of new available data, new scientific questions, and newchallenges for learning and inference. In particular, biological data are often relationally structured and highly diverse, well-suited to approaches that combine multiple weak evidence from heterogeneous sources. These data may include sequenced genomes of a variety of organisms, gene expression data from multiple technologies, protein expression data, protein sequence and 3D structural data, protein interactions, gene ontology and pathway databases, genetic variation data (such as SNPs), and an enormous amount of textual datain the biological and medical literature. New types of scientific and clinical problems require the development of novel supervised and unsupervised learningmethods that can use these growing resources. Furthermore, next generationsequencing technologies are yielding terabyte scale data sets that requirenevel algorithmic solutions. The goal of this min-symposium is to presentemerging problems and machine learning techniques in computational biology.

Location: Regency C

MS5 Partially Observable Reinforcement Learning

MARCUS HUTTERmarcus@hutter1.netAustralian National Universitywillian.uther@nicta.com.auWILL UTHERwillian.uther@nicta.com.auNICTA Neville Road Research Labppoupart@cs.uwaterloo.caPASCAL POUPARTppoupart@cs.uwaterloo.caUniversity of Waterlooppoupart@cs.uwaterloo.ca

For many years, the reinforcement learning community primarily focused on sequential decision making in fully observable but unknown domains while the planning under uncertainty community focused on known but partially observable domains. Since most problems are both partially observable and (at least partially) unknown, recent years have seen a surge of interest in combining the related, but often different, algorithmic machineries developed in the two communities. The time thus seems ripe for a symposium that brings these two communities together and reviews recent advances in this convergence. A reinforcement learning agent for a partially observable environment is often broken into two parts: 1) the inference of an environment model from data; and 2) the solution of the associated control/planning problem. There has been significant progress on both these fronts in recent years. Both linear and non-linear models of various forms can now be learned from history data. Modern POMDP solvers can also now handle some models with millions of states. This symposium brings together five active researchers in PORL research to present some state-of-the-art developments.

Location: Regency E/F

Poster and Demonstrations Floor Plans



Wednesday Demonstrations



Monday Posters





Monday Posters by Subject





Tuesday Posters





Tuesday Posters by Subject





Wednesday Posters





Wednesday Posters by Subject





POSTER AND DEMO FLOOR PLANS
Reviewers

PIETER ABBEEL NAOKI ABE RYAN ADAMS Felix Agakov ALEKH AGARWAL DEEPAK AGARWAL Edo Airoldi Shotaro Akaho KARTEEK ALAHARI YASEMIN ALTUN Chandrasekaran Anand CHARLES ANDERSON GALEN ANDREW ANDRAS ANTOS CEDRIC ARCHAMBEAU ANDREAS ARGYRIOU ARTIN ARMAGAN HIDEKI ASOH CHRIS ATKESON JEAN-YVES AUDIBERT Peter Auer Meg Aycinena Lippow FRANCIS BACH TIMOTHY BAILEY SIVARAMAN BALAKRISHNAN ARINDAM BANERJEE ZIV BAR-JOSEPH RICHARD BARANIUK DAVID BARBER EVGENIY BART ANDREW BARTO REGINA BARZILAY SUMIT BASU AARON BATISTA CHRISTIAN BECKMANN Ulrik Beierholm GILL BEJERANO MIKHAIL BELKIN ASA BEN HUR SAMY BENGIO KRISTIN BENNETT MATTHIAS BETHGE BRETT BETHKE Jinbo Bi STEFFEN BICKEL DANNY BICKSON HORST BISCHOF ANDREW BLAKE GILLES BLANCHARD

MATHEW BLASCHKO DAVID BLEI KWABENA BOAHEN KARSTEN BORGWARDT VIVEK BORKAR MATTHEW BOTVINICK GUILLAUME BOUCHARD RICHARD BOWDEN MICHAEL BOWLING JORDAN BOYD-GRABER ULF BREFELD TABA BRIAN SEBASTIEN BUBECK JOACHIM BUHMANN WOLFRAM BURGARD CHRIS BURGES Keith Bush KATHERINE CAMERON STEPHANE CANU **OLIVIER CAPPE** LARRY CARIN FRANCOIS CARON MIGUEL A. CARREIRA-PERPINAN XAVIER CARRERAS CARLOS CARVALHO Rui Castro GAVIN CAWLEY LAWRENCE CAYTON TAYLAN CEMGIL NICOLO CESA-BIANCHI Soumen Chakrabarti JONATHAN CHANG Kamalika Chaudhuri Gal Chechik LI CHENG Sonia Chernova DAVID CHIANG SILVIA CHIAPPA HAI LEONG CHIEU Yejin Choi SUMIT CHOPRA ANDREAS CHRISTMANN KEN CHURCH ANDRZEJ CICHOCKI Stephen Clark MARK COATES MICHAEL COLLINS RONAN COLLOBERT FRANCESCO CORONA

REVIEWERS

GREG CORRADO TIMOTHEE COUR **AARON COURVILLE** KOBY CRAMMER Antonio Criminisi JAMES CUSSENS MARCO CUTURI FLORENCE D'ALCHE-BUC KIMBERLEE D'ARDENNE Arnak Dalalyan DAVID DANKS ANIRBAN DASGUPTA SANJOY DASGUPTA DENVER DASH HAL DAUME NATHANIEL DAW NANDO DE FREITAS Rob de Ruyter VIRGINIA DE SA DENNIS DECOSTE OFER DEKEL OLIVIER DELALLEAU INDERJIT DHILLON THOMAS DIETTERICH CHRIS DING CABLOS DIUK WASSER Adrian Dobra Eizaburo Doi BRENT DOIRON PIOTR DOLLAR FINALE DOSHI Arnaud Doucet Petros Drineas JOHN DUCHI MIROSLAV DUDIK DAVID DUNSON URI EDEN JAMES ELDER GAL ELIDAN CHARLES ELKAN DANIEL ELLIS YAAKOV ENGEL BARBARA ENGELHARDT Dumitru Erhan ELEAZAR ESKIN JASON FARQUHAR JACOB FELDMAN Rob Fergus ALAN FERN VITTORIO FERRARI ILA FIETE

Mario Figueiredo JENNY FINKEL Jozsef Fiser JOHN FISHER ANDREW FITZGIBBON Peter Foldiak FOPE FOLOWOSELE DAVID FORSYTH CHARLESS FOWLKES EMILY FOX MICHAEL FRANK PAOLO FRASCONI Peter Frazier Desobry Frederic BILL FREEMAN NIR FRIEDMAN KARL FRISTON Kenji Fukumizu KUZMAN GANCHEV GILLES GASSO Peter Gehler CLAUDIO GENTILE CAUWENBERGHS GERT SEBASTIAN GERWINN Mohammad Ghavamzadeh DEBASHIS GHOSH Indiveri Giacomo ANDREW GILPIN MARK GIROLAMI AMIR GLOBERSON JOZIEN GOENSE JACOB GOLDBERGER SHARON GOLDWATER NOAH GOODMAN Geoff Gordon DILAN GORUR VIVEK GOYAL JOAO GRACA YVES GRANDVALET DAVID GRANGIER KRISTEN GRAUMAN RUSS GREINER ARTHUR GRETTON Remi Gribonval JIM GRIFFIN THOMAS GRIFFITHS MORITZ GROSSE-WENTRUP GREG GRUDIC YUHONG GUO TODD GURECKIS MICHAEL GUTMANN

PATRICK HAFFNER Adrian Haith TARA HAMILTON JIHUN HAMM BARBARA HAMMER FIRAS HAMZE EDWIN HANCOCK ZAID HARCHAOUI DAVID HARDOON STEFAN HARMELING JOHN HARRIS MATT HARRISON RICHARD HARTLEY ANDREA HASENSTAUB MILOS HAUSKRECHT ELAD HAZAN XUMING HE MARTIAL HEBERT MATTHIAS HEIN KATHERINE HELLER RALF HERBRICH MARK HERBSTER Tom Heskes JEREMY HILL Geoffrey Hinton SEPP HOCHREITER Jesse Hoey HEIKO HOFFMANN JAKE HOFMAN THOMAS HOFMANN STEVEN CHU-HONG HOI DEREK HOIEM ANTTI HONKELA ERIC HORVITZ MATTHEW HOWARD PATRIK HOYER JONATHAN HUANG MARCUS HUTTER QUENTIN HUYS AAPO HYVARINEN CHRISTIAN IGEL ALEXANDER IHLER Shiro Ikeda MASATO INOUE MICHAEL ISARD Tommi Jaakkola LAURENT JACOB ROBERT JACOBS HERBERT JAEGER FRANK JAEKEL MICHAEL JAMES

Dominik Janzing Mehrdad Jazayeri TONY JEBARA Stefanie Jegelka SHANE JENSEN Rong Jin THORSTEN JOACHIMS MARK JOHNSON NEBOJSA JOJIC MICHAEL JORDAN FREDERIC JURIE Balázs Kégl MICHAEL KALISH TAKAFUMI KANAMORI TAKAFUMI KANAMORI HYUN MIN KANG TOMMY KAPLAN Bert Kappen YAN KARKLIN Meier Karlheinz HISASHI KASHIMA SAMUEL KASKI SATHIYA KEERTHI CHARLES KEMP KRISTIAN KERSTING Joseph Keshet **Kee-Eung Kim** SERGEY KIRSHNER Jyrki Kivinen STEFAN KLANKE ROBERT KLEINBERG JENS KOBER KILIAN KOEPSELL KONRAD KOERDING PUSHMEET KOHLI ERIC KOLACZYK MLADEN KOLAR DAPHNE KOLLER VLADIMIR KOLMOGOROV ZICO KOLTER **RISI KONDOR** ARYEH KONTOROVICH TERRY KOO JANA KOSECKA URS KOSTER RICHARD KRAUZLIS BALAJI KRISHNAPURAM EYAL KRUPKA BRIAN KULIS M. PAWAN KUMAR SANJIV KUMAR

REVIEWERS

TAKIO KURITA JAMES KWOK JOHN LAFFERTY CHRISTOPH LAMPERT NIELS LANDWEHR HUGO LAROCHELLE JAN LARSEN PAVEL LASKOV Francois Laviolette NEIL LAWRENCE Alessandro Lazaric SVETLANA LAZEBNIK NICOLAS LE ROUX GUY LEBANON DANIEL LEE TAI SING LEE WEE SUN LEE ROBERT LEGENSTEIN MATE LENGYEL VINCENT LEPETIT CHRISTINA LESLIE ANAT LEVIN ROGER LEVY LIHONG LI Ping Li YISHENG LI PERCY LIANG HSUAN-TIEN LIN CE LIU TIE-YAN LIU Yan Liu DANIEL LIZOTTE YONATAN LOEWENSTEIN MICHAEL LONDON Phil Long Tomas Lozano-Perez Elliot Ludvig JUSTIN MA Wei Ji Ma JAKOB MACKE Sridhar Mahadevan Satyaki Mahalanabis JULIEN MAIRAL Jesus Malo GIDEON MANN Shie Mannor VIKASH MANSINGHKA JIRI MATAS JON MCAULIFFE CHRIS MEEK Ron Meir

Francisco Melo ROLAND MEMISEVIC VIVIENNE MING EINAT MINKOV MICHAEL MISTRY SRINJOY MITRA DAICHI MOCHIHASHI BABACK MOGHADDAM SHAKIR MOHAMED CLAIRE MONTELEONI JORIS MOOIJ Greg Mori QUAID MORRIS MICHAEL MOZER SAYAN MUKHERJEE KLAUS-ROBERT MULLER Remi Munos Noboru Murata KEVIN MURPHY IAIN MURRAY BOAZ NADLER RAMESH NALLAPATI RAMA NATARAJAN DANIEL NAVARRO Gerhard Neumann DAVID NEWMAN XUANLONG NGUYEN DUY NGUYEN-TUONG HANNES NICKISCH ALEX NICULESCU-MIZIL Yael Niv WILLIAM NOBLE ROB NOWAK SEBASTIAN NOWOZIN KLAUS OBERMAYER UWE OHLER AUDE OLIVA BRUNO OLSHAUSEN CHENG SOON ONG Sylvie Ong TAKASHI ONODA MANFRED OPPER FRANCESCO ORABONA GERGO ORBAN Peter Orbanz MILES OSBORNE CHRISTIAN OSENDORFER SIMON OSINDERO DAVID PAGE CHRIS PAL Sylvain Paris

HYEYOUNG PARK Ronald Parr NAAMA PARUSH ANDREA PASSERINI MIROSLAV PAWLAK Itsik Pe'er BARAK PEARLMUTTER DMITRY PECHYONY KRISTIAAN PELCKMANS Marcello Pelillo JAAKKO PELTONEN Fernando Pereira Francisco Pereira Fernando Perez-Cruz JAN PETERS SLAV PETROV JONATHAN PILLOW MARK PITT CHRISTIAN PLAGEMANN JOHN PLATT JEAN PONCE Massi Pontil BRIAN POTETZ PASCAL POUPART DOINA PRECUP HUBERT PREISSL PATRICK PREZ YUAN QI Ariadna Quattoni JOAQUIN QUINONERO-CANDELA FILIP RADLINSKI MAXIM RAGINSKY Ali Rahimi RAJAT RAINA UMESH RAJASHEKAR Alain Rakotomamonjy LIVA RALAIVOLA Deva Ramanan MARC'AURELIO RANZATO CARL RASMUSSEN NATHAN RATLIFF MAGNUS RATTRAY PRADEEP RAVIKUMAR MARK REID STEFAN RIEZLER Philippe Rigollet Helge Ritter ABEL RODRIGUEZ TIMOTHY ROGERS JUSTIN ROMBERG LORENZO ROSASCO

DAVID ROSENBERG DAVID ROSS MICHAEL ROSS STEPHANE ROSS AFSHIN ROSTAMIZADEH STEFAN ROTH VOLKER ROTH CARSTEN ROTHER DANIEL ROY Cynthia Rudin THOMAS RUECKSTIESS ALEX RUSSEL BRYAN RUSSELL MANEESH SAHANI PAUL SAJDA RUSLAN SALAKHUTDINOV Adam Sanborn MARK SANDLER TED SANDLER GUIDO SANGUINETTI SCOTT SANNER BEN SAPP CRAIG SAUNDERS STEFAN SCHAAL TOBIAS SCHEFFER Bernt Schiele CORDELIA SCHMID SCOTT SCHMIDLER Mikkel Schmidt JEFF SCHNEIDER BERNHARD SCHOELKOPF PAUL SCHRATER TANJA SCHULTZ CLAYTON SCOTT MICHELE SEBAG MATTHIAS SEEGER FRANK SEHNKE WALTER SENN Peggy Series THOMAS SERRE Rocco Servedio Fei Sha PATRICK SHAFTO GREGORY SHAKHNAROVICH Ohad Shamir TATYANA SHARPEE Amnon Shashua JOHN SHAWE-TAYLOR CHRISTIAN SHELTON XIAOTONG SHEN PRADEEP SHENOY

REVIEWERS

BERTRAM SHI LIU SHIH-CHII SHOHEI SHIMIZU JONATHON SHLENS LAVI SHPIGELMAN ILYA SHPITSER Oren Shriki IVO SHTEREV LEONID SIGAL RICARDO SILVA DAVID SILVER EERO SIMONCELLI OZGUR SIMSEK VIKAS SINDHWANI AMIT SINGER YORAM SINGER AARTI SINGH Tomas Singliar FABIAN SINZ Josef Sivic NOAM SLONIM CRISTIAN SMINCHISESCU Stelios Smirnakis NOAH SMITH ALEX SMOLA ED SNELSON Stefano Soatto Peter Sollich FRITZ SOMMER LE SONG Alessandro Sperduti SUVRIT SRA NATHAN SREBRO KARTHIK SRIDHARAN BHARATH SRIPERUMBUDUR DANIEL STEFANKOVIC **OLIVER STEGLE** JOCHEN STEIL FLORIAN STEINKE INGO STEINWART MARK STEYVERS SUSANNE STILL ALAN STOCKER Amos Storkey JONATHAN STROUD NATHAN STURTEVANT Amarnag Subramanya Erik Sudderth Ilya Sutskever CHARLES SUTTON RICH SUTTON

UMAR SYED CSABA SZEPESVARI PRASAD TADEPALLI Partha Talukdar ERIK TALVITE Ameet Talwalkar Amos Tanay TONG BOON TANG MARSHALL TAPPEN BEN TASKAR Sekhar Tatikonda GRAHAM TAYLOR MATT TAYLOR RUSS TEDRAKE YEE WHYE TEH JOSHUA TENENBAUM CHOON HUI TEO Ambuj Tewari **OLIVIER TEYTAUD** Evangelos Theodorou **ROMAIN THIBAUX** MICHAEL THON Robert Tibshirani Robert Tillman JO-ANNE TING MICHALIS TITSIAS MICHAEL TODD ANDREA TOLIAS Antonio Torralba MARC TOUSSAINT KRISTINA TOUTANOVA Volker Tresp JOCHEN TRIESCH BILL TRIGGS Julia Trommershauser JOEL TROPP IVOR WAI-HUNG TSANG IOANNIS TSOCHANTARIDIS Koji Tsuda Zhuowen Tu TINNE TUYTELAARS EIJI UCHIBE NAONORI UEDA Lyle Ungar RAQUEL URTASUN WILL UTHER GIORGIO VALENTINI ANTAL VAN DEN BOSCH PATRICK VAN DER SMAGT MARTIJN VAN OTTERLO MARK VAN ROSSUM

JARNO VANHATALO NUNO VASCONCELOS GAURAV VEDA ANDREA VEDALDI AKI VEHTARI JAKOB VERBEEK Alessandro Verri JEAN-PHILIPPE VERT Sethu Vijayakumar S.V.N. VISHWANATHAN R. Jacob Vogelstein ULRIKE VON LUXBURG SLOBODAN VUCETIC Ed Vul MARTIN WAINWRIGHT CHRISTIAN WALDER HANNA WALLACH Lei Wang ZHUANG WANG LARRY WASSERMAN Chu Wei KILIAN WEINBERGER DAVID WEISS YAIR WEISS MAX WELLING WIM WIEGERINCK ERIC WIEWIORA Rebecca Willett CHRIS WILLIAMS ROBERT WILSON DAVID WINGATE John Winn **OLE WINTHER** PATRICK WOLFE DANIEL WOLPERT JENNIFER WORTMAN VAUGHAN STEVE WRIGHT Mingrui Wu

Wei Wu YING-NIAN WU ERIC XING Linli Xu KATSU YAMANE MING-HSUAN YANG QIANG YANG Zhijun Yang CHEN YANOVER JIEPING YE Scott Wen-tau Yih YIMING YING ELAD YOM-TOV Byron Yu CHUN-NAM YU Kai Yu Shipeng Yu Stella Yu YISONG YUE MIKHAIL ZASLAVSKIY Assaf Zeevi LUKE ZETTLEMOYER Hongyuan Zha Dell Zhang Kun Zhang ALICE ZHENG LU ZHENGDONG DING XUAN ZHOU Shuheng Zhou Long Zhu Xiaojin Zhu TODD ZICKLER ANDREAS ZIEHE MARTIN ZINKEVICH ANDREW ZISSERMAN Onno Zoeter BARBARA ZWICKNAGL

REVIEWERS

224

Subject Index

Applications, 78, 95, 125, 128, 156, 161, 179, 182, 183, 193 Bioinformatics, 60, 93, 109, 125, 126, 146, 198 Collaborative Filtering, 77, 114 Graphics, 158, 170 Information Retrieval, 75, 77, 89, 106, 112, 162, 182, 189 Natural Language Processing, 55, 75, 110, 126, 162, 190, 194, 195 Systems Biology, 125 Time Series Prediction, 55, 90, 98, 174, 205 Web Applications, 72, 106, 112, 125, 132, 168 Cognitive Science, 79, 106–108, 122–124, 133, 158, 162, 173, 177, 178, 199 Attention, 108, 124 Development, 162, 199 Inference & Reasoning, 78, 79, 140, 162, 177, 178, 199 Knowledge Representation & Acquisition, 79, 106, 122, 158, 159, 162, 177, 178, 198, 199 Language, 107, 123 Learning, 79, 106, 107, 122, 123, 158, 159, 177, 178, 198 Memory, 60, 79, 108, 124, 159, 178 Model Comparison Methods, 79, 124 Perception, 79, 108, 124, 140, 162, 173, 176, 178Reinforcement Learning, 154, 163 Response Time Modeling, 64 Control and Reinforcement Learning, 97, 149, 179 Bayesian RL, 147, 148 Control, 95, 105, 110, 147, 148, 179 Markov Decision Processes, 105, 110, 147, 148 Multi-Agent Systems and Game Theory, 82, 97, 180 Planning and Decision Making, 105, 147, 179 Policy Search, 96 POMDPs, 95, 147, 148, 180 Reinforcement Learning, 56, 71, 95–97, 105, 110, 146, 147, 156, 161, 179 Hardware, 123

- Neuroscience, 55, 66–69, 137, 154, 159, 163, 174, 175, 205
 - Brain Imaging, 66, 67, 69, 89, 105, 106, 139, 141, 142, 159, 175, 176, 191

- Brain-computer Interfaces & Neural Prostheses, 65, 71, 108, 110, 137, 141, 154, 163
- Computational Neural Models, 65, 84, 106, 109, 135, 138–140, 154, 159, 162, 163, 174–176, 205
- Cortex, 66, 67
- Motor Control, 154, 163
- Neural Decoding, 65, 66, 68, 69, 89, 140, 154, 174, 175
- Neural Populations, 66, 68, 106, 135, 138, 140, 162, 174, 176
- Plasticity, 109, 140, 154, 163, 175
- Spiking Neurons, 55, 65–67, 96, 109, 140, 162, 176
- Optimization, 55, 73, 88, 128, 130, 155, 158, 168, 170, 189
 - Combinatorial Optimization, 76, 107, 109, 128, 129
 - Constrained Optimization, 55, 88, 91, 130
 - Convergence Analysis, 106, 127, 130, 131, 142, 155, 168
 - Convex Optimization, 61, 70, 80, 90, 93, 106, 112, 130, 131, 142, 144, 146, 155, 159, 162, 168, 183, 185, 188
 - Gradient Methods, 67, 84, 106, 114, 117, 130, 131, 142, 155, 156, 162, 168, 179, 183, 185
 - Non-convex Optimization, 55, 67, 88, 130, 158, 170

Stochastic Methods, 80, 84, 128, 131, 162, 183

- Probabilistic Models and Methods, 56, 59, 62, 64, 72, 74, 102, 110, 116–118, 121, 157–159, 168, 191, 193, 195, 198
 - Bayesian Methods, 55, 56, 58–63, 75, 86, 102, 105, 106, 109–113, 116, 119–121, 124, 132, 133, 141, 154, 160, 162, 167, 173, 175, 191–194, 196, 197
 - Belief Propagation, 106, 119, 149, 195, 196
 - Causal Inference, 122
 - Density Estimation, 98, 170, 174, 192, 197, 205
 - Exact and Approximate Inference, 56, 58–60, 62, 64, 88, 109, 110, 116, 118, 121, 136, 157, 192, 193, 195, 196
 - Gaussian Processes, 59, 66, 81, 105, 111, 120, 159, 193
 - Graphical Models, 56, 58–60, 62, 63, 72, 78, 85, 94, 120, 123, 125, 162, 188, 190–193,

195, 196, 204

- Missing Data, 89, 102, 114, 116
- Mixture Models, 61, 191, 192, 194
- Model Selection & Structure Learning, 55, 58, 60, 62, 86, 90, 95, 109, 120, 122, 126, 131, 196, 204
- Monte Carlo Methods, 55, 58, 61, 62, 72, 86, 106, 110, 112, 113, 120, 123, 126, 160, 194, 195
- Structured and Relational Data, 56, 63, 87, 106, 112, 117, 122, 139, 195
- Temporal Models and Sequence Data, 60, 71, 98, 109, 118, 121, 159, 160, 184, 187, 193, 194
- Variational Methods, 58, 59, 123, 141, 159, 191, 193, 195

Robotics

Exploration & Map Building, 149

- Speech and Signal Processing, 134, 138
 - Signal Processing, 55, 67, 88, 117, 135, 161, 167
 - Source Separation, 57
 - Speaker Identification, 134
 - Speech Recognition, 56, 57, 164
- $\begin{array}{l} \mbox{Supervised Learning, 82, 88, 99, 106, 110, 126, 141,} \\ \mbox{142, 145, 161, 165, 167, 182, 198} \end{array}$
 - Classification, 56, 81, 84, 87–90, 92, 93, 105, 106, 109, 139, 142, 145, 146, 157–159, 161, 163–165, 169, 172, 174, 181, 183, 185, 187, 197, 198
 - Ensemble Methods and Boosting, 55, 88, 92, 93, 106, 142, 145, 158, 163, 181
 - Kernel Methods, 56, 80, 81, 84, 91, 95, 98, 106, 107, 120, 127, 129, 142, 144, 146, 157, 161, 164–166, 171, 172, 183
 - Large Margin Methods, 146, 158, 165
 - Learning with Structured Data, 74, 80, 83, 84, 87, 89, 94, 99, 106, 110, 117, 118, 126, 142, 146, 161, 164, 165, 169, 181
 - Model Selection, 55, 83, 90, 127, 144, 161, 167
 - Neural Networks, 77, 84, 109, 143, 145, 166, 169
 - Regression, 55, 56, 59, 80, 83, 84, 88, 90, 94, 100, 105, 111, 120, 127, 133, 143–145, 147, 161, 166, 167, 194, 198
 - Sparsity and Feature Selection, 55, 80, 83, 87, 90, 99, 105, 109, 120, 126, 127, 129, 133, 135, 139, 144–146, 161, 162, 167, 184–186
- Theory, 56, 81, 82, 84, 94, 99, 102, 105, 107, 109,

111, 126, 128, 129, 132, 133, 155, 157, 180, 181, 183

- Consistency, 56, 82, 83, 88, 95, 100, 117, 133, 157, 158, 162, 181, 183, 184, 186
- Convergence Analysis, 56, 64, 84, 97, 100, 127, 133, 162, 184, 186
- Game Theory & Computational Economics, 82, 90, 128, 155, 181
- Information Theory, 66, 82, 155, 162, 181, 183, 184
- Online Learning, 55, 76, 80–82, 84, 88, 102, 107, 116, 127, 128, 130–132, 155, 158, 161, 165, 168, 181, 183, 184
- Statistical Learning Theory, 56, 83, 84, 88, 89, 94, 100, 114, 123, 127, 131, 133, 143, 157, 162, 165, 182–185
- Unsupervised and Semi-supervised Learning, 55, 70, 72–74, 76, 78, 101, 106, 109, 112, 114– 116, 134, 145, 158, 159, 162, 166, 173, 182, 186–188, 190, 191
 - Active Learning, 71, 78, 116, 141, 185
 - Anomaly Detection, 98, 158, 186
 - Clustering, 55, 63, 70, 73, 74, 76, 79, 92, 106, 112, 113, 125, 159, 160, 171, 188, 189, 192, 194, 198
 - Embeddings and Manifold Learning, 55, 71, 73, 83, 102, 111, 115, 159, 162, 186, 188, 189, 193
 - ICA, PCA, CCA & Other Linear Models, 62, 70, 74, 86, 106, 112, 186, 187, 197
 - Matrix Factorization, 63, 77, 113, 125, 161, 172
 - Similarity and Distance Learning, 75, 102, 106, 111–113, 133, 144, 159, 185, 188
 - Spectral Methods, 101, 114, 134, 158, 171, 186
 - Topic Models, 55, 59, 60, 72, 75, 87, 101, 106, 112, 125, 134, 137, 160, 184, 188
 - Transduction Learning, 73, 92, 115, 116, 159, 188
- Vision, 78, 85, 93, 101, 133, 134, 158, 161, 168, 170–172

3D Reconstruction, 133, 169

- Biological Vision, 109, 135, 137, 139, 166, 170
- Image Coding, 68, 102, 111, 115
- Image Segmentation, 158, 169, 171, 172
- Machine Vision Applications, 75, 149, 169
- Motion and Tracking, 55, 88, 133, 160, 171, 194
- Natural Scene Statistics, 85, 86, 102, 109, 116, 139, 143, 158, 170, 193

 $\begin{array}{c} \text{Object Recognition, 61, 80, 86, 101, 109, 133,} \\ 134, 136, 137, 145, 158, 161, 163, 168, 169, \\ 171{-}174 \end{array}$

 $Visual Features, 66, 86, 101, 102, 116, 134, 169 \\ Visual Perception, 86, 109, 133, 135, 139, 162, \\ 168, 173$

Visual Processing, 86, 109, 135, 139

SUBJECT INDEX

Author Index

Pirsiavash, Hamed, 173

Abe, Naoki, 145 Agarwal, Alekh, 183 Ailon, Nir, 76 Allen, Martin, 180 Alvarez, George, 108 Amini, Massih, 182 Anati, Roy, 149 Arleo, Angelo, 65 Arlot, Sylvain, 127 Arora, Raman, 88 Avres Pereira, Jose, 61 Bach, Francis, 50, 82, 127 Bai, Bing, 76 Baker, Chris, 199 Bard, Nolan, 179 Barnard, Kobus, 86 Bartels, Andreas, 69 Bartlett, Peter, 183 Barto, Andrew, 179 Beck, Diane, 138, 176 Bejan, Cosmin, 194 Belkin, Mikhail, 75 Ben-Tal, Aharon, 91 Bengio, Samy, 74, 115 Bengio, Yoshua, 102, 166, 196 Berens, Philipp, 137, 174 Bergsma, Jacob, 78 Bergstra, James, 166 Berkes, Pietro, 137 Bethge, Matthias, 66, 137, 170, 174 Bhatnagar, Shalabh, 97, 178 Bhattacharyya, Chiranjib, 91 Bian, Wei, 186 Bilmes, Jeff, 57, 115, 128 Black, Michael, 100 Blair, Alan, 156 Blankertz, Benjamin, 141 Blaschko, Mathew, 69 Blei, David, 59, 112, 160 Blum, Avrim, 182 Bo, Liefeng, 93, 171 Boben, Marko, 169 Borgwardt, Karsten, 161 Bottou, Leon, 102 Bouchard, Guillaume, 82 Bouchard-Côté, Alexandre, 126

Boutsidis, Christos, 70 Bouvrie, Jake, 83 Bowling, Michael, 81, 179 Boyd-Graber, Jordan, 160 Brückner, Michael, 90 Bradski, Gary, 100 Brasselet, Romain, 65 Brefeld, Ulf, 145 Briggman, Kevin, 169 Bush, Keith, 71 Cadambi, Hari, 199 Caetano, Tiberio, 117, 194 Cai, Chenghui, 148 Campbell, Colin, 142, 185 Campbell, William, 134 Canny, John, 124 Caputo, Barbara, 173 Carbonetto, Peter, 58 Carin, Lawrence, 101, 148, 191 Caron, Francois, 192 Castellani, Umberto, 93 Cavagnaro, Daniel, 124 Cayton, Lawrence, 112 Cecchi, Guillermo, 105 Cepeda, Nicholas, 177 Cevher, Volkan, 94 Chai, Barry, 176 Chai, Kian Ming, 80 Chang, Jonathan, 160 Chang, Kai-min, 177 Chang, Yu-Ming, 84 Chase, Steven, 153 Chater, Nick, 198 Chaudhuri, Kamalika, 184 Chechik, Gal, 74, 207 Chen, Haojun, 101 Chen, Jianhui, 135 Chen, Kewei, 175 Chen, Wei, 89 Chen, Xi, 143 Chen, Ye, 124 Chen, Yuanahao, 85 Chieu, Hai Leong, 118 Chin, Tat-Jun, 171 Chklovskii, Dmitri, 68 Cho, Youngmin, 165 Choi, Arthur, 196 Choi, Seungjin, 125

Christmann, Andreas, 84 Clémençon, Stéphan, 87 Coen-Cagli, Ruben, 139 Collins, Michael, 164 Collobert, Ronan, 51, 76 Conroy, Bryan, 67 Coquelin, Pierre-Arnaud, 117 Cortes, Corinna, 76, 144 Coti, Camille, 111 Courville, Aaron, 196 Crammer, Koby, 165 Cristani, Marco, 93 Cseke, Botond, 191 Culpepper, Benjamin, 193 Cuturi, Marco, 98 D'Aspremont, Alexandre, 98 Dai, Guang, 74 Dalalyan, Arnak, 133 Dallmeier, Valentin, 192 Daniilidis, Kostas, 149 Danoczy, Marton, 141 Darrell, Trevor, 100, 136, 189 Darwiche, Adnan, 196 Daume, Hal, 197 Davis, Randall, 77 Dayan, Peter, 139, 154 de Freitas, Nando, 52 De la Torre, Fernando, 187 Deguest, Romain, 117 Deisseroth, Karl, 204 Dekel, Ofer, 94 Denk, Winfried, 169 Depecker, Marine, 87, 151 Desai, Vijay, 148 Dhillon, Inderjit, 77 Diaconis, Persi, 203 Diao, Dongcui, 178 Dietterich, Thomas, 206 Dietterich, Tom, 102 Dietz, Laura, 192 Doshi-Velez, Finale, 62, 147 Doucet, Arnaud, 52, 192 Doya, Kenji, 96 Dredze, Mark, 165 Drineas, Petros, 70 Du, Lan, 191 Duchi, John, 155 Dunson, David, 191 Eck, Douglas, 196

Ecker, Alexander, 174

Efros, Alyosha, 133, 172 Eliasmith, Chris, 201 Evans, Owain, 199 Fan, Wei, 92 Farias, Vivek, 148, 168 Fazli, Siamac, 141 Fei-Fei, Li, 138, 176 Fergus, Rob, 101, 170 Ferrari, Vittorio, 173 Fidler, Sanja, 169 Figueiras-Vidal, Anibal, 119 Finkel, Leif, 140 Fiser, Jozsef, 137 Fleisher, Adam, 175 Fletcher, Alyson, 63, 167 Fowlkes, Charless, 173 Fox, Emily, 160 Frank, Andrew, 195 Frank, Jordan, 150 Frank, Michael, 108 Freeman, Bill, 85, 172 Freund, Yoav, 184 Fritz, Mario, 100 Fromer, Menachem, 157 Fujiwara, Yusuke, 68 Fukumizu, Kenji, 95, 119, 156 G. Dinesh, 91 Gamez de Leon, Jorge, 150 Ganchev, Kuzman, 190 Ganesh, Arvind, 70 Gao, Jing, 92 Gao, Tianshi, 171 Garcia, Eric, 166 Germain, Pascal, 164 Gerrish, Sean, 160 Gershman, Samuel, 59, 178 Gerstner, Wulfram, 96 Gerwinn, Sebastian, 66, 137, 174 Ghahramani, Zoubin, 62, 103 Ghebreab, Sennay, 85 Gibson, Bryan, 122 Girolami, Mark, 142 Globerson, Amir, 157 Goldberger, Jacob, 118 Goldstein, Rita, 58 Golovin, Daniel, 127 Gong, Yihong, 114 Goodfellow, Ian, 143 Goodman, Noah, 199 Gorur, Dilan, 119

Gould, Stephen, 171 Goutte, Cyril, 182 Goval, Vivek, 63 Grünewälder, Steffen, 67 Graca, Joao, 190 Grangier, David, 76 Gray, Alexander, 72, 129, 197 Gretton, Arthur, 95, 122, 156 Griffiths, Thomas, 63, 107, 175 Grozea, Cristian, 141 Grzegorczyk, Marco, 60 Guestrin, Carlos, 121 Guillory, Andrew, 115 Gupta, Ankur, 200 Gupta, Maya, 166 Guyon, Isabelle, 151, 206 Haffari, Gholamreza, 74 Hamze, Firas, 58 Han, Jiawei, 92 Harabagiu, Sanda, 194 Harchaoui, Zaid, 95 Hardoon, David, 205 Hartmann, Mitra, 153 Hasson, Uri, 142 Haxby, James, 67 Hazan, Elad, 82, 155 Hebert, Martial, 136 Hein, Matthias, 83, 114 Heller, Katherine, 198 Helmstaedter, Moritz, 169 Hemmer, Pernille, 79 Henao, Ricardo, 120 Hennequin, Guillaume, 96 Henniges, Marc, 190 Hertz, Tomer, 207 Heskes, Tom, 191 Hickl, Andrew, 194 Hinton, Geoffrey, 89, 104, 144, 188 Hoi, Steven Chu-Hong, 81, 113 Honorio, Jean, 58 Horvitz, Eric, 71 Hsu, Anne, 107 Hsu. Chun-Nan. 84 Hsu, Daniel, 99, 184 Hu, Bao-gang, 86 Hu, Chonghai, 130 Hu, Tao, 68 Huang, Hanshen, 84 Huang, Jonathan, 121 Huang, Kaizhu, 185 Huang, Shuai, 175

Husmeier, Dirk, 60 Hutter, Marcus, 180, 208 Ihler, Alexander, 195 Ilin, Alexander, 193 Isbell, Charles, 97 Iwata, Tomoharu, 71 Jagabathula, Srikanth, 168 Jagarlapudi, Saketha Nath, 91 Jain, Prateek, 77 Jaiswal, Ragesh, 76 Jern, Alan, 78, 122, 177 Jin, Rong, 81, 88, 113, 127, 187 Johansson, Roland, 65 Jojic, Nebojsa, 93 Jones, Matt, 64 Jordan, Michael, 63, 82, 160 Jung, Kyomin, 116 Jylänki, Pasi, 59 K.R., Ramakrishnan, 91 Kakade, Sham, 99 Kalai, Adam, 181 Kale, Satyen, 82, 155 Kamitani, Yukiyasu, 68 Kanade, Varun, 181 Kao, Yi-hao, 143 Kapoor, Ashish, 71 Kapralov, Michael, 124 Karam, Zahi, 134 Karasuyama, Masayuki, 90 Karayev, Sergey, 100 Kaschube, Matthias, 66 Kawahara, Yoshinobu, 128 Keefer, Edward, 65 Kemp, Charles, 78, 122, 176, 177 Keriven, Renaud, 133 Keshavan, Raghunandan, 114 Khan, Emtivaz, 203 Kim, Gunhee, 168 Kim, Jong Kyoung, 125 Kim, Junae, 92 Kim, Kwang In, 114 Kim, Seyoung, 197 King, Irwin, 73, 187 King, Matthew, 58 Klampfl, Stefan, 66 Klein, Dan, 126 Kloft, Marius, 145 Knowles, David, 62 Koerding, Konrad, 172

Kohli, Pushmeet, 116 Kolar, Mladen, 90, 125 Koller, Daphne, 61, 102, 171 Kolter, J. Zico, 206 Konidaris, George, 179 Kpotufe, Samory, 146 Krause, Andreas, 127 Krishnan, Dilip, 170 Kulesza, Alex, 165 Kulis, Brian, 189 Kumar, M. Pawan, 61 Kumar, Sanjiv, 164 Kveton, Branislav, 202 Kwok, James, 130 Lücke, Jörg, 190 Lacasse, Alexandre, 164 Lall, Ashwin, 60 Lamme, Victor, 85 Lan, Yanyan, 89 Lanckriet, Gert, 129, 156 Lanctot, Marc, 81 Langford, John, 99, 130 Largman, Yan, 138 Laskov, Pavel, 145 Laviolette, Francois, 164 Lazaro-Gredilla, Miguel, 119 Lazebnik, Svetlana, 102 Le, Quoc, 143 LeCun, Yann, 102 Lee, Dongryeol, 72, 129 Lee, Honglak, 138 Lee, Michael, 79 Lee, Wee Sun, 118 Lee, Yuh-Jye, 84 Legenstein, Robert, 153 Lengyel, Mate, 154 Leonardis, Ales, 169 Leonardo, Anthony, 68 Leordeanu, Marius, 136 Leshem, Amir, 118 Li, Hang, 89, 181 Li, Jing, 175 Li, Shuo-Yen Robert, 188 Li, Wu-Jun, 111 Li, Yuxi, 128 Li, Zhenguo, 188 Liang, Feng, 92 Liang, Percy, 82 Liao, Xuejun, 148 Lim, John, 194 Lin, Zhouchen, 108

Lindsey, Robert, 177 Littman, Michael, 49 Liu, Han, 143 Liu, Jun, 135, 175 Liu, Tie-Yan, 89, 181 Liu, Wentai, 65 Lozano, Aurelie, 145 Lu, Hongjing, 198 Luo, Jie, 173 Luttinen, Jaakko, 193 Lyu, Michael, 187 Müller, Klaus-Robert, 145 Ma, Yi, 70 Ma, Zhi-Ming, 89 Maass, Wolfgang, 66, 103, 140, 153 Mac Dermed, Liam, 97 Macindoe, Owen, 199 Macke, Jakob, 66 Madrid-Sanchez, Jaisiel, 205 Maei, Hamid, 97 Mahoney, Michael, 70 Maillard, Odalric, 83 Maina, Ciira, 202 Malisiewicz, Tomasz, 133 Mann, Gideon, 163 Mansinghka, Vikash, 151, 201 Mao, Jianchang, 88 March, William, 129 Marchand, Mario, 164 Margaritis, Dimitris, 87 Marlin, Benjamin, 203 Martelli, Catherine, 105 Martinot, Jean-Luc, 105 Martinsson, Gunnar, 47 McAuley, Julian, 117 McCallum, Andrew, 75, 146, 195 McDonald, Ryan, 163 Meek, Christopher, 95 Meka, Raghu, 77 Meng, Yicong, 135 Miller, Brent, 79 Miller, Kurt, 63 Mimno, David, 75 Mishra, Nina, 132 Mitchell, Tom, 89 Miyawaki, Yoichi, 68 Moallemi, Ciamac, 148 Moghaddam, Baback, 203 Mohamed, Shakir, 62 Mohri, Mehrvar, 76, 144, 163, 164 Montanari, Andrea, 61, 114

Monteleoni, Claire, 76 Mori, Greg, 74 Morimura, Tetsuro, 96 Movellan, Javier, 78 Mozer, Michael, 64, 177 Muller, Klaus-Robert, 141 Munos, Remi, 83, 117 Murino, Vittorio, 93 Murphy, Kevin, 203 Myung, Jay, 124 Nadler, Boaz, 185 Nagano, Kiyohito, 128 Nagarajan, Srikantan, 167 Nair, Vinod, 144 Negahban, Sahand, 100 Nessler, Bernhard, 140 Neven, Hartmut, 150 Ng, Andrew, 138, 143, 206 Niemasik, Jamie, 152 Noble, William, 207 Nolte, Guido, 206 Norman, Kenneth, 59 Nowak, Rob, 131 Obermayer, Klaus, 67 Oh, Sewoong, 114 Oja, Erkki, 73 Olshausen, Bruno, 193 Onken, Arno, 67 Oostenveld, Robert, 191 Orbanz, Peter, 132 Ortiz, Luis, 58 Ouyang, Hua, 72 Ouyang, Tom, 77 OZAKIN, ARKADAS, 197 Paillere-Martinot, Marie-Laure, 105 Paisley, John, 101 Palatucci, Mark, 89 Pan, Weike, 130 Paragios, Nikos, 58 Park, Minwoo, 200 Parrado-Hernandez, Emilio, 205 Pashler, Harold, 177 Pavlov, Dmitry, 124 Pelillo, Marcello, 189 Peng, Jian, 93 Peng, Yigang, 70 Pereira, Fernando, 102, 115, 190 Perez-Cruz, Fernando, 205 Perina, Alessandro, 93

Perkins, Theodore, 121 Perotte, Adler, 59 Petrik, Marek, 147 Petrov, Slav, 126 Petterson, James, 72, 117 Pfeiffer, Michael, 140 Pfister, Jean-Pascal, 154 Pham, Peter, 138 Pillow, Jonathan, 205 Pineau, Joelle, 71 Pitt, Mark, 124 Plaze, Marion, 105 Poggio, Tomaso, 83 Poline, Jean-Baptiste, 105 Pomerleau, Dean, 89 Popescu, Florin, 141, 206 Poupart, Pascal, 208 Precup, Doina, 97 Qi, Yanjun, 76, 207 Qi, Yuan, 123 Quadrianto, Novi, 72, 194 Raginsky, Maxim, 102 Rai, Piyush, 197 Raj, Bhiksha, 57 Ram, Parikshit, 72, 129 Ramadge, Peter, 67, 142 Ramanan, Deva, 173 Rangan, Sundeep, 63, 167 Rao, Shankar, 70 Rao, Vinayak, 61 Raskutti, Garvesh, 183 Ravikumar, Pradeep, 100, 183 Reiman, Eric, 175 Ren, Lu, 101, 191 Rish, Irina, 105 Rogers, Timothy, 122 Rohanimanesh, Khashayar, 146 Rosasco, Lorenzo, 83 Rostamizadeh, Afshin, 144 Rota Bulò, Samuel, 189 Russell, Bryan, 172 Ruvolo, Paul, 78 S, Raman, 91 Sadamasa, Kunihiko, 76 Saenko, Kate, 136 Sahani, Maneesh, 190 Salakhutdinov, Ruslan, 62, 120, 188 Saligrama, Venkatesh, 186 Samaras, Dimitris, 58

Sanborn, Adam, 198 Sapiro, Guillermo, 101 Saul, Lawrence, 102, 165 Saxe, Andrew, 143 Scheffer, Tobias, 90, 192 Schlecht, Joseph, 86 Schmidt, Mikkel, 113 Schoelkopf, Bernhard, 156 Scholte, Steven, 85 Schultz, Karl, 195 Schuurmans, Dale, 128, 194 Schwartz, Andrew, 153 Schwartz, Odelia, 139 Sederberg, Per. 59 Seeger, Matthias, 141 Sejnowski, Terry, 103 Sen, Subhabrata, 182 Seung, Sebastian, 169 Shah, Devavrat, 116, 168 Shalit, Uri, 74 Shani, Guy, 95 Shanian, Sara, 164 Sharma, Varun, 74 Shashanka, Madhusudana, 57 Shelton, Jacquelyn, 69 Shen, Chunhua, 92 Shervashidze, Nino, 161 Shi, Bertram, 135 Shi, Lei, 175 Silberman, Nathan, 163 Silver, David, 97, 156 Simoncelli, Eero, 170 Singer, Ben, 67 Singer, Yoram, 115, 155 Singh, Sameer, 146, 195 Singh-Miller, Natasha, 164 Sinha, Kaushik, 75 Sinz, Fabian, 170 Sivic, Josef, 172 Slivkins, Aleksandrs, 132 Smaragdis, Paris, 57 Smeulders, Arnold, 85 Sminchisescu, Cristian, 171 Smola, Alex, 72, 130 Smyth, Padhraic, 195 So, Anthony Man-Cho, 188 Socher, Richard, 59 Sollich, Peter, 111 Song, Dawn, 182 Song, Le, 90, 125 Sonnenburg, Soeren, 145

Spatscheck, Oliver, 182 Spirtes, Peter, 122 Sprekeler, Henning, 96 Srebro, Nathan, 185 Sriperumbudur, Bharath, 95, 129, 156 Steinke, Florian, 114 Steinwart, Ingo, 84 Stevenson, Ian, 172 Steyvers, Mark, 79 Streeter, Matthew, 127 Strelow, Dennis, 115 Sturim, Douglas, 134 Subramanya, Amarnag, 57 Sudderth, Erik, 160 Sukthankar, Rahul, 136 Sun, Liang, 135, 175 Sun, Yizhou, 92 Suter, David, 171 Sutskever, Ilva, 62 Sutton, Rich, 97, 178 Swirszcz, Grzegorz, 145 Syed, Umar, 132 Szepesvari, Csaba, 97, 128, 178 Takeuchi, Ichiro, 90 Talwalkar, Ameet, 164 Tao, Dacheng, 186 Taskar, Ben, 190 Teh, Yee Whye, 61, 119 Tenenbaum, Joshua, 62, 108, 178, 199 Thirion. Bertrand, 105 Thyreau, Benjamin, 105 Tillman, Robert, 122 Titsworth, Matthew, 194 Todorov, Emanuel, 149 Torralba, Antonio, 48, 85, 101, 168 Tsuda, Koji, 128 Turaga, Srinivas, 169 Turner, Richard, 190 Uchibe, Eiji, 96 Ueda, Naonori, 71 Ullman, Tomer, 199 Urry, Matthew, 111 Usunier, Nicolas, 182 Uther, Will, 208 Uther, William, 156

Valizadegan, Hamed, 88 van den Hengel, Anton, 92 Van Durme, Benjamin, 60 Van Gerven, Marcel, 191 Van Roy, Benjamin, 143 Vanhatalo, Jarno, 59 Vanpaemel, Wolf, 79 Vayatis, Nicolas, 87 Vedaldi, Andrea, 80 Vehtari, Aki, 59 Veness, Joel, 156 Venkataraman, Shobha, 182 Verdu, Sergio, 99 Vert, Jean-Philippe, 98 Vul, Ed, 108, 177, 178 Wainwright, Martin, 100, 183 Walker, Dan, 163 Wallach, Hanna, 75 Walther, Dirk, 138, 176 Wang, Chong, 59, 112, 160 Wang, Hanzi, 171 Wang, Huixia Judy, 192 Wang, Lei, 92 Wang, Liwei, 185 Wang, Shaojun, 74 Wang, Shijun, 127 Wang, Yang, 74 Wang, Yongmei Michelle, 192 Watanabe, Yusuke, 119 Waugh, Kevin, 81, 179 Weiden, Matthew, 198 Weiss, Yair, 85, 101, 157 Weston, Jason, 51, 76 White, Ben, 137 White, Leonard, 66 Whitehill, Jacob, 78 Wick, Michael, 146 Wilder, Matthew, 64 Willsky, Alan, 160 Wilson, Robert, 140 Winther, Ole, 120 Wipf, David, 167 Wright, John, 70 Wu, Dan, 118 Wu, Lei, 113 Wu, Teresa, 175 Wu, Ting-fan, 78 WU, Xiao-Ming, 188 Xi, Yongxin, 142 Xia, Fen, 181 Xiang, Zhen, 142 Xiao, Lin, 80 Xing, Eric, 90, 125, 197 Xu, Fei, 78

Xu, Jinbo, 93 Xu, Ningyi, 123 Xu, Zenglin, 73, 187 Yamada, Takeshi, 71 Yan, Feng, 123 Yan, Xiang, 143 Yang, Shuang-Hong, 86 Yang, Xiaolin, 197 Yang, Zhi, 65 Yang, Zhirong, 73, 187 Yao, Bangpeng, 138 Yao, Hengshuai, 178 Ye, Jieping, 135, 175 Ye, Nan, 118 Yeung, Dit-Yan, 111 Ying, Yiming, 142, 185 Yoshimoto, Junichiro, 96 Yu, Bin, 100, 183 Yu, Jin, 117 Yu, Kai, 114 Yu, Nenghai, 113 Yu, Yao-Liang, 128 Yuille, Alan, 198 Zeller, Andreas, 192 Zha, Hongyuan, 86 Zhang, Ruofei, 88 Zhang, Tong, 99, 114 Zhang, Zhihua, 74, 111 Zhao, Manqi, 186 Zhao, Peilin, 81 Zhao, Qi, 65 Zheng, Wenming, 108 Zhou, Chunxiao, 192 Zhou, Feng, 187 Zhou, Mingyuan, 101 Zhou, Shuheng, 131 Zhou, Xueyuan, 185 Zhou, Yang, 127 Zhu, Jianke, 113, 187 Zhu, Long, 85 Zhu, Xiaojin, 122 Zien, Alexander, 145, 207 Zilberstein, Shlomo, 147, 180 Zinkevich, Martin, 81, 130 Zisserman, Andrew, 80, 172 Zoran, Daniel, 85

AUTHOR INDEX

236

NOTES

Notes

238