NIPS 2016 Workshops

Workshop organizers make last-minute changes to their schedule. Download this document again to get the lastest changes, or use the NIPS mobile application.

Schedule Highlights

Dec. 9, 2016

AC Barcelona Hotel - Barcelona Room, **Practical Bayesian Nonparametrics** *Foti, Broderick, Campbell, Hughes, Miller, Schein, Williamson, Xu*

AC Barcelona, Sagrada Familia, **Interpretable Machine Learning for Complex Systems** *Wilson, Kim, Herlands*

Area 1, **Deep Reinforcement Learning** *Silver, Singh, Abbeel, Chen*

Area 2, **Learning in High Dimensions with Structure** *Rao, Jain, Yu, Yuan, Bach*

Area 3, **Adversarial Training** *Lopez-Paz, Bottou, Radford*

Area 5 + 6, **Nonconvex Optimization for Machine Learning: Theory and Practice** *Mobahi, Anandkumar, Liang, Jegelka, Choromanska*

Area 7 + 8, **Efficient Methods for Deep Neural Networks** *Rastegari, Courbariaux*

Hilton Diag. Mar, Blrm. A, **The Future of Interactive Machine Learning** *Mathewson, Subramanian, Ho, Loftin, Austerweil, Harutyunyan, Precup, El Asri, Gombolay, Zhu, Chernova, Isbell, Pilarski, Wong, Veloso, Shah, Taylor, Argall, Littman*

Hilton Diag. Mar, Blrm. B, **Cognitive Computation: Integrating Neural and Symbolic Approaches** *Besold, Bordes, Wayne, Garcez*

Hilton Diag. Mar, Blrm. C, **Intuitive Physics** *Lerer, Wu, Tenenbaum, Dupoux, Fergus*

Room 111, **Extreme Classification: Multi-class and Multi-label Learning in Extremely Large Label Spaces** *Cisse, Varma, Bengio*

Room 112, **Advances in Approximate Bayesian Inference** *Broderick, Mandt, McInerney, Tran, Blei, Murphy, Gelman, Jordan*

Room 113, **Reliable Machine Learning in the Wild** *Hadfield-Menell, Weller, Duvenaud, Steinhardt, Liang*

Room 114, **Representation Learning in Artificial and Biological Neural Networks** *Wehbe, Van Gerven, Grosse-Wentrup, Rish, Murphy, Langs, Cecchi, Nunez-Elizalde*

Room 115, **3D Deep Learning** *Yu, Lim, Fisher, Huang, Xiao*

Room 116, **Machine Learning for Health** *Shalit, Ghassemi, Fries, Ranganath, Karaletsos, Kale, Schulam, Fiterau*

Room 117, **Time Series Workshop** *Anava, Cuturi, Khaleghi, Kuznetsov, Rakhlin*

Room 120 + 121, **Crowdsourcing and Machine Learning** *Singla, Frongillo, Venanzi*

Room 122 + 123, **Adaptive Data Analysis** *Feldman, Ramdas, Roth, Smith*

Room 124 + 125, **Machine Learning for Intelligent Transportation Systems** *Li, Darrell*

Room 127 + 128, **Imperfect Decision Makers: Admitting Real-World Rationality** *Karny, Wolpert, Insua, Guy*

Room 129 + 130, **Challenges in Machine Learning: Gaming and Education** *Guyon, Viegas, Kégl, Hamner, Escalera*

Room 131 + 132, **Private Multi-Party Machine Learning** *Balle, Bellet, Evans, Gascón*

Room 133 + 134, **Learning, Inference and Control of Multi-Agent Systems** *Graepel, Lanctot, Leibo, Lever, Marecki, Oliehoek, Tuyls, Holgate*

Room 211, **Brains and Bits: Neuroscience meets Machine Learning** *Fletcher, Dyer, Sohl-Dickstein, Vogelstein, Koerding, Macke*

Room 212, **Machine Intelligence @ NIPS** *Mikolov, Marco, Joulin, Kruszewski, Lazaridou, Simonic*

VIP Room, **People and machines: Public views on machine learning, and what this means for machine learning researchers** *Odell, Donnelly, Montgomery, Hauert, Ghahramani, Gorman*

VIP Room, **Neurorobotics: A Chance for New Ideas, Algorithms and Approaches** *Rueckert, Riedmiller*

Dec. 10, 2016

Area 1, **Bayesian Deep Learning** *Gal, Louizos, Ghahramani, Murphy, Welling*

Area 2, **Optimizing the Optimizers** *Mahsereci, Davies, Hennig*

Area 3, **Deep Learning for Action and Interaction** *Finn, Hadsell, Held, Levine, Liang*

Area 5 + 6, **Learning with Tensors: Why Now and How?** *Anandkumar, Ge, Liu, Nickel, Yu*

Area 7 + 8, **Continual Learning and Deep Networks** *Pascanu, Ring, Schaul*

Hilton Diag. Mar, Blrm. A, **End-to-end Learning for Speech and Audio Processing** *Hershey, Brakel*

Hilton Diag. Mar, Blrm. B, **Machine Learning for Spatiotemporal Forecasting** *Popescu, Escalera, Baró, Ayache, Guyon*

Hilton Diag. Mar, Blrm. C, **Let's Discuss: Learning Methods for Dialogue** *Daume III, Mineiro, Stent, Weston*

Room 111, **Large Scale Computer Vision Systems** *Paluri, Torresani, Chechik, Garcia, Tran*

Room 112, **OPT 2016: Optimization for Machine Learning** *Sra, Bach, J. Reddi, He*

Room 113, **Neural Abstract Machines & Program Induction** *Bošnjak, de Freitas, Kulkarni, Neelakantan, Reed, Riedel, Rocktäschel*

Room 114, **Towards an Artificial Intelligence for Data Science** *Sutton, Geddes, Ghahramani, Smyth, Williams*

Room 115, **The Future of Gradient-Based Machine Learning Software** *Wiltschko, DeVito, Bastien, Lamblin*

Room 116, **Machine Learning Systems** *Lakshmiratan, Li, Sen, Bird, Mehanna*

Room 117, **Bayesian Optimization: Black-box Optimization and Beyond** *Calandra, Shahriari, Gonzalez, Hutter, Adams*

Room 120 + 121, **Adaptive and Scalable Nonparametric Methods in Machine Learning** *Ramdas, Sriperumbudur, Gretton, Liu, Lafferty, Kpotufe, Szabó*

Room 122 + 123, **Computing with Spikes** *Bohte, Nowotny, Savin, Zambrano*

Room 127 + 128, **Constructive Machine Learning** *Costa, Gärtner, Passerini, Pachet*

Room 129 + 130, **Machine Learning for Education** *Baraniuk, Ngiam, Studer, Grimaldi, Lan*

Room 131 + 132, **Connectomics II: Opportunities and Challenges for Machine Learning** *Jain, Turaga*

Room 133 + 134, **"What If?" Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems** *Silva, Shawe-Taylor, Swaminathan, Joachims*

Room 211, **Brains and Bits: Neuroscience meets Machine Learning (2nd day)**

Room 212, **Machine Learning in Computational Biology** *Quon, Mostafavi, Zou, Engelhardt, Stegle, Fusi*

VIP Room, **Neurorobotics: A Chance for New Ideas, Algorithms and Approaches (2nd day)**

Dec. 9, 2016

Practical Bayesian Nonparametrics

Nick Foti, Tamara Broderick, Trevor Campbell, Michael C. Hughes, Jeff Miller, Aaron Schein, Sinead A Williamson, Yanxun Xu

AC Barcelona Hotel - Barcelona Room, Fri Dec 09, 08:00 AM

In theory, Bayesian nonparametric (BNP) methods are well suited to the large data sets that arise in the sciences, technology, politics, and other applied fields. By making use of infinite-dimensional mathematical structures, BNP methods allow the complexity of a learned model to grow as the size of a data set grows, exhibiting desirable Bayesian regularization properties for small data sets and allowing the practitioner to learn ever more from larger data sets. These properties have resulted in the adoption of BNP methods across a diverse set of application areas---including, but not limited to, biology, neuroscience, the humanities, social sciences, economics, and finance. In practice, BNP methods present a number of computational and modeling challenges. Recent work has brought a wide range of models to bear on applied problems, going beyond the Dirichlet process and Gaussian process. Meanwhile, advances in accelerated inference are making these models tractable in big data problems. In this workshop, we will explore new BNP methods for diverse applied problems, including cutting-edge models being developed by application domain experts. We will also discuss the limitations of existing methods and discuss key problems that need to be solved. A major focus of the workshop will be to expose participants to practical software tools for performing Bayesian nonparametric analyses. In particular, we plan to host hands-on tutorials to introduce workshop participants to some of the software packages that can be used to easily perform posterior inference for BNP models, e.g. Stan, BNPy, and BNP.jl. We expect workshop participants to come from a variety of fields, including but not limited to machine learning, statistics, engineering, political science, and various biological sciences. The workshop will be relevant both to BNP experts as well as those interested in learning how to apply BNP models. There will be a special emphasis on work that makes BNP methods easy-to-use in practice and computationally efficient. Participants will leave the workshop with (i) exposure to recent advances in the field, (ii) hands-on experience with software implementing BNP methods, and (iii) an idea of the current challenges that need to be overcome in order to make BNP methods more widespread in practice. These goals will be accomplished through a series of invited and contributed talks, a poster session, and at least one hands-on tutorial session where participants can get their hands dirty with BNP methods. This workshop builds off of the “Bayesian Nonparametrics: The Next Generation” workshop held at NIPS in 2015. While that workshop had a broad remit, spanning theory, applications and computation, this year’s workshop shows a fresh focus on the practical aspects of BNP methods. During last year’s panel discussion, there were many questions about computational techniques and practical applications, suggesting that this direction will be of great interest to the many applied machine learning researchers who attend the conference.

Schedule

08:15 AM	**Welcome and Introductions**	
08:30 AM	**Tamara Broderick: Foundations Talk**	*Broderick*
09:00 AM	**Jennifer Hill: Invited Talk**	
09:30 AM	**Hyunjik Kim: Scaling up the Automatic Statistician: Scalable Structure Discovery in Regression using Gaussian Processes**	
09:45 AM	**Melanie F. Pradier: Sparse Three-parameter Restricted Indian Buffet Process for Understanding International Trade**	
10:00 AM	**Bailey Fosdick: Multiresolution Network Models**	
11:00 AM	**Poster Spotlights**	
11:15 AM	**Poster Session**	
12:15 PM	**Lunch Session Intro**	
12:45 PM	**Rob Trangucci: Stan Tutorial, with focus on Gaussian Processes**	
01:45 PM	**Mike Hughes: BNPy tutorial - Clustering with Dirichlet Processes and extensions in Python**	
03:30 PM	**Marc Deisenroth: Invited Talk**	
04:00 PM	**David Malmgren-Hansen: Analyzing Learned Convnet Features with Dirichlet Process Gaussian Mixture Models**	
04:15 PM	**Neil Dhir: Lions as Probabilistic Programs**	
04:30 PM	**Panel on Software Development**	
05:00 PM	**Maria DeYoreo: A Markovian Model for Nonstationary Time Series via Bayesian nonparametrics**	
05:30 PM	**Invited Panel on Models, Methods, and Applications**	

Abstracts (6):

Abstract 6: **Bailey Fosdick: Multiresolution Network Models in Practical Bayesian Nonparametrics**, 10:00 AM

Many existing statistical and machine learning tools for social network analysis focus on a single level of analysis. Methods designed for clustering optimize a global partition of the graph, whereas projection based approaches (e.g. the latent space model in the statistics literature) represent in rich detail the roles of individuals. Many pertinent questions in sociology and economics, however, span multiple scales of analysis. Further, many questions involve comparisons across disconnected graphs that will inevitably be of different sizes, either due to missing data or the inherent heterogeneity in real-world networks. We propose a class of network models that represent network structure on multiple scales and facilitate comparison across graphs with different numbers of individuals. These models differentially invest modeling effort within subgraphs of high density, often termed communities, while maintaining a parsimonious structure between said subgraphs. We show that our model class is projective, highlighting an ongoing discussion in the social network modeling literature on the dependence of inference paradigms on the size of the observed graph. We illustrate the utility of our method using data on household relations from Karnataka, India.

Abstract 13: **David Malmgren-Hansen: Analyzing Learned Convnet Features with Dirichlet Process Gaussian Mixture Models in Practical Bayesian Nonparametrics**, 04:00 PM

Contributed Talk

Abstract 14: **Neil Dhir: Lions as Probabilistic Programs in Practical Bayesian Nonparametrics**, 04:15 PM

Contributed Talk

Abstract 15: **Panel on Software Development in Practical Bayesian Nonparametrics**, 04:30 PM

Dustin Tran, Columbia University
Lead developer of Edward

Aki Vehtari, Aalto University
Stan contributor and Lead developer of GPstuff

Martin Trapp, Austrian Research Institute for Artificial Intelligence
Lead developer of BNP.jl (Julia implementation of BNP methods)

Mike Hughes, Harvard University
Lead developer of BNPy

Abstract 16: **Maria DeYoreo: A Markovian Model for Nonstationary Time Series via Bayesian nonparametrics in Practical Bayesian Nonparametrics**, 05:00 PM

Stationary time series models built from parametric distributions are, in general, limited in scope due to the assumptions imposed on the residual distribution and autoregression relationship. We present a modeling approach for univariate time series data, which makes no assumptions of stationarity, and can accommodate complex dynamics and capture non-standard distributions. The model for the transition density arises from the conditional distribution implied by a Bayesian nonparametric mixture of bivariate normals. This results in a flexible autoregressive form for the conditional transition density, defining a time-homogeneous, non-stationary Markovian model for real-valued data indexed in discrete time. To obtain a computationally tractable algorithm for posterior inference, we utilize a square-root-free Cholesky decomposition of the mixture kernel covariance matrix. Results from simulated data suggest that the model is able to recover challenging transition densities and non-linear dynamic relationships. We also illustrate the model on time intervals between eruptions of the Old Faithful geyser. Extensions and open questions about accommodating higher order structure and developing state-space models are also discussed.

Abstract 17: **Invited Panel on Models, Methods, and Applications in Practical Bayesian Nonparametrics**, 05:30 PM

Invited Panel:
Bailey Fosdick, Colorado State University
Maria DeYoreo, Duke University
Suchi Saria, Johns Hopkins University
Jim Griffin, University of Kent
Marc Deisenroth, Imperial College London

Interpretable Machine Learning for Complex Systems

Andrew G Wilson, Been Kim, William Herlands

AC Barcelona, Sagrada Familia, Fri Dec 09, 08:00 AM

Complex machine learning models, such as deep neural networks, have recently achieved great predictive successes for visual object recognition, speech perception, language modelling, and information retrieval. These predictive successes are enabled by automatically learning expressive features from the data. Typically, these learned features are a priori unknown, difficult to engineer by hand, and hard to interpret. This workshop is about interpreting the structure and predictions of these complex models.

Interpreting the learned features and the outputs of complex systems allows us to more fundamentally understand our data and predictions, and to build more effective models. For example, we may build a complex model to predict long range crime activity. But by interpreting the learned structure of the model, we can gain new insights into the processing driving crime events, enabling us to develop more effective public policy. Moreover, if we learn, for example, that the model is making good predictions by discovering how the geometry of clusters of crime events affect future activity, we can use this knowledge to design even more successful predictive models.

This 1 day workshop is focused on interpretable methods for machine learning, with an emphasis on the ability to learn structure which provides new fundamental insights into the data, in addition to accurate predictions. We will consider a wide range of topics, including deep learning, kernel methods, tensor methods, generalized additive models, rule based models, symbolic regression, visual analytics, and causality. A poster session, coffee breaks, and a panel guided discussion will encourage interaction between attendees. We wish to carefully review and enumerate modern approaches to the challenges of interpretability, share insights into the underlying properties of popular machine learning algorithms, and discuss future directions.

Schedule

08:45 AM	**Opening Remarks**
09:00 AM	**Honglak Lee**

09:30 AM	**Why Interpretability: A Taxonomy of Interpretability and Implications for Principled Evaluation (Finale Doshi-Velez)**
10:00 AM	**Best paper award talks**
11:00 AM	**Intelligible Machine Learning for HealthCare (Rich Caruana)**
02:30 PM	**Maya Gupta**
02:30 PM	**The Power of Monotonicity ■for Practical■ Machine Learning (Maya Gupta)**
03:30 PM	**Finding interpretable sparse structure in fMRI data with dependent relevance determination priors (Jonathan Pillow)**
04:00 PM	**Poster session**
04:30 PM	**Better Machine Learning Through Data (Saleema Amershi)**
05:00 PM	**Future Directions in Interpretable Machine Learning**

Abstracts (6):

Abstract 3: **Why Interpretability: A Taxonomy of Interpretability and Implications for Principled Evaluation (Finale Doshi-Velez) in Interpretable Machine Learning for Complex Systems**, 09:30 AM

With a growing interest in interpretability, there is an increasing need to characterize what exactly we mean by it and how to sensibly compare the interpretability of different approaches. In this talk, I suggest that our current desire for "interpretability" is as vague as asking for "good predictions" -- a desire that. while entirely reasonable, must be formalized into concrete needs such as high average test performance (perhaps held-out likelihood is a good metric) or some kind of robust performance (perhaps sensitivity or specificity are more appropriate metrics). This objective of this talk is to start a conversation to do the same for interpretability: I will describe distinct, concrete objectives that all fall under the umbrella term of interpretability and how each objective suggests natural evaluation procedures. I will also describe highlight important open questions in the evaluation of interpretable models.

Joint work with Been Kim, and the product of discussions with countless collaborators and colleagues.

Abstract 4: **Best paper award talks in Interpretable Machine Learning for Complex Systems**, 10:00 AM

Title: An unexpected unity among methods for interpreting model predictions
Scott Lundberg and Su-In Lee

Title: Feature Importance Measure for Non-linear Learning Algorithms
Marina M.-C. Vidovic, Nico Görnitz, Klaus-Robert Müller, and Marius Kloft

Abstract 5: **Intelligible Machine Learning for HealthCare (Rich Caruana) in Interpretable Machine Learning for Complex Systems**, 11:00 AM

In machine learning often a tradeoff must be made between accuracy and intelligibility: the most accurate models usually are not very intelligible (e.g., deep neural nets, boosted trees, and random forests), and the most intelligible models usually are less accurate (e.g., linear/logistic regression). This tradeoff often limits the accuracy of models that can be applied in mission-critical applications such as healthcare where being able to understand, validate, edit, and ultimately trust a learned model is important. We have developed a learning method based on generalized additive models (GAMs) that is often as accurate as full complexity models, but remains as intelligible as linear/logistic regression models. In the talk I'll present two case studies where these high-performance generalized additive models (GA2Ms) are applied to healthcare problems yielding intelligible models with state-of-the-art accuracy. In the pneumonia risk prediction case study, the intelligible model uncovers surprising patterns in the data that previously prevented complex learned models from being deployed, but because it is intelligible and modular allows these patterns to easily be recognized and removed. In the 30-day hospital readmission case study, we show that the same methods scale to large datasets containing hundreds of thousands of patients and thousands of attributes while remaining intelligible and providing accuracy comparable to the best (unintelligible) machine learning methods.

Abstract 7: **The Power of Monotonicity ■for Practical■ Machine Learning (Maya Gupta) in Interpretable Machine Learning for Complex Systems**, 02:30 PM

■What prior knowledge do humans have about machine learning problems that we can take advantage of as regularizers? One common intuition is that certain inputs should have a positive (only) effect on the output, for example, the price of a house should only increase as its size goes up, if all else is the same. Incorporating such monotonic priors into our machine learning algorithms can dramatically increase their interpretability and debuggability. We'll discuss state-of-the-art algorithms to learn flexible monotonic functions, and share some stories about why monotonicity is such an important regularizer for practical problems where train and test samples are not IID, especially when learning from clicks.

Abstract 8: **Finding interpretable sparse structure in fMRI data with dependent relevance determination priors (Jonathan Pillow) in Interpretable Machine Learning for Complex Systems**, 03:30 PM

In many problem settings, parameters are not merely sparse, but dependent in such a way that non-zero coefficients tend to cluster together. We refer to this form of dependency as region sparsity". Classical sparse regression methods, such as the lasso and automatic relevance determination (ARD), which models parameters as independent a priori, and therefore do not exploit such dependencies. Here we introduce a hierarchical model for smooth, region-sparse weight vectors and tensors in a linear regression setting. Our approach represents a hierarchical extension of the relevance determination framework, where we add a transformed Gaussian process to model the dependencies between the prior variances of regression weights. We combine this with a structured model of the prior variances of Fourier coefficients, which eliminates unnecessary high frequencies. The resulting prior encourages weights to be region-sparse in two different bases simultaneously. We develop Laplace approximation and Monte Carlo Markov Chain (MCMC) sampling to provide efficient inference for the posterior, and show substantial improvements over existing methods for both simulated and real fMRI datasets.

Abstract 10: **Better Machine Learning Through Data (Saleema Amershi) in Interpretable Machine Learning for Complex Systems**, 04:30 PM

Machine learning is the product of both an algorithm and data. While machine learning research tends to focus on algorithmic advances, taking the data as given, machine learning practice is quite the opposite. Most of the influence practitioners have in using machine learning to build predictive models comes through interacting with data, including crafting the data used for training and examining results on new data to inform future iterations. In this talk, I will present tools and techniques we have been developing in the Machine Teaching Group at Microsoft Research to support the model building process. I will then discuss some of the open challenges and opportunities in improving the practice of machine learning.

Deep Reinforcement Learning

David Silver, Satinder Singh, Pieter Abbeel, Xi Chen

Area 1, Fri Dec 09, 08:00 AM

Although the theory of reinforcement learning addresses an extremely general class of learning problems with a common mathematical formulation, its power has been limited by the need to develop task-specific feature representations. A paradigm shift is occurring as researchers figure out how to use deep neural networks as function approximators in reinforcement learning algorithms; this line of work has yielded remarkable empirical results in recent years. This workshop will bring together researchers working at the intersection of deep learning and reinforcement learning, and it will help researchers with expertise in one of these fields to learn about the other.

Schedule

09:00 AM	**Rich Sutton**	*Sutton*
09:30 AM	**Contributed Talks - Session 1**	
10:00 AM	**John Schulman**	*Schulman*
11:00 AM	**Raia Hadsell**	*Hadsell*
11:30 AM	**Contributed Talks - Session 2**	
12:00 PM	**Chelsea Finn**	*Finn*
12:30 PM	**Lunch**	
01:30 PM	**Nando De Freitas**	*de Freitas*
02:00 PM	**Contributed Talks - Session 3**	
02:30 PM	**Posters - Session 1**	
03:00 PM	**Coffee Break**	
03:30 PM	**Late Breaking Talk**	
03:45 PM	**Junhyuk Oh**	*Oh*
04:15 PM	**Josh Tenenbaum**	*Tenenbaum*
04:45 PM	**Panel Discussion**	
05:30 PM	**Posters - Session 2**	

Learning in High Dimensions with Structure

Nikhil Rao, Prateek Jain, Hsiang-Fu Yu, Ming Yuan, Francis Bach

Area 2, Fri Dec 09, 08:00 AM

Several applications necessitate learning a very large number of parameters from small amounts of data, which can lead to overfitting, statistically unreliable answers, and large training/prediction costs. A common and effective method to avoid the above mentioned issues is to restrict the parameter-space using specific structural constraints such as sparsity or low rank. However, such simple constraints do not fully exploit the richer structure which is available in several applications and is present in the form of correlations, side information or higher order structure. Designing new structural constraints requires close collaboration between domain experts and machine learning practitioners. Similarly, developing efficient and principled algorithms to learn with such constraints requires further collaborations between experts in diverse areas such as statistics, optimization, approximation algorithms etc. This interplay has given rise to a vibrant area of "learning with structure in high dimensions". The goal of this workshop is to bring together the aforementioned diverse set of people who have worked in these areas and encourage discussions with an aim to help define the current frontiers for the area and initiate a discussion about meaningful and challenging problems that require attention.

Schedule

08:30 AM	**Richard Samworth**
09:10 AM	**Po-Ling Loh**
09:50 AM	**Sahand Negahban**
11:00 AM	**Mark Schmidt**
11:40 AM	**Kai-Wei Chang**
12:20 PM	**Poster Spotlights**
02:00 PM	**Allen Yang**
02:40 PM	**Chinmay Hegde**
03:30 PM	**Rene Vidal**
04:10 PM	**Guillaume Obozinski**
04:50 PM	**Lorenzo Rosasco**

Adversarial Training

David Lopez-Paz, Alec Radford, Leon Bottou

Area 3, Fri Dec 09, 08:00 AM

In adversarial training, a set of machines learn together by pursuing competing goals. For instance, in Generative Adversarial Networks (GANs, Goodfellow et al., 2014) a generator function learns to synthesize samples that best resemble some dataset, while a discriminator function learns to distinguish between samples drawn from the dataset and samples synthesized by the generator. GANs have emerged as a promising framework for unsupervised learning: GAN generators are able to produce images of unprecedented visual quality, while GAN discriminators learn features with rich semantics that lead to state-of-the-art semi-supervised learning (Radford et al., 2016). From a conceptual perspective, adversarial training is fascinating because it bypasses the need of loss functions in learning, and opens the door to new ways of regularizing (as well as fooling or attacking) learning machines. In this one-day workshop, we invite scientists and practitioners interested in adversarial training to gather, discuss, and establish new research collaborations. The workshop will feature invited talks, a hands-on demo, a panel discussion, and contributed spotlights and posters.

Among the research topics to be addressed by the workshop are

* Novel theoretical insights on adversarial training
* New methods and stability improvements for adversarial optimization
* Adversarial training as a proxy to unsupervised learning of representations
* Regularization and attack schemes based on adversarial perturbations
* Adversarial model evaluation
* Adversarial inference models
* Novel applications of adversarial training

Want to learn more? Get started by generating your own MNIST digits using a GAN in 100 lines of Torch: https://goo.gl/Z2leZF

Schedule

09:00 AM	**Set up posters**	
09:15 AM	**Welcome**	*Lopez-Paz, Radford, Bottou*
09:30 AM	**Introduction to Generative Adversarial Networks**	*Goodfellow*
10:00 AM	**How to train a GAN?**	*Chintala*
11:00 AM	**Learning features to distinguish distributions**	*Gretton*
11:00 AM	**Learning features to compare distributions**	
11:30 AM	**Training Generative Neural Samplers using Variational Divergence**	*Nowozin*
12:00 PM	**Lunch break**	
02:00 PM	**Adversarially Learned Inference (ALI) and BiGANs**	*Courville*
02:30 PM	**Energy-Based Adversarial Training and Video Prediction**	*LeCun*
03:00 PM	**Discussion panel**	*Goodfellow, Chintala, Gretton, Nowozin, Courville, LeCun, Denton*
04:00 PM	**Coffee break**	
04:30 PM	**Spotlight presentations**	
06:00 PM	**Poster session**	
06:00 PM	**Additional poster and open discussions**	

Abstracts (9):

Abstract 2: **Welcome in Adversarial Training**, *Lopez-Paz, Radford, Bottou* 09:15 AM

Just a quick introduction to the first NIPS workshop on Adversarial Training.

Abstract 3: **Introduction to Generative Adversarial Networks in Adversarial Training**, *Goodfellow* 09:30 AM

Generative adversarial networks are deep models that learn to generate samples drawn from the same distribution as the training data. As with many deep generative models, the log-likelihood for a GAN is intractable. Unlike most other models, GANs do not require Monte Carlo or variational methods to overcome this intractability. Instead, GANs are trained by seeking a Nash equilibrium in a game played between a discriminator network that attempts to distinguish real data from model samples and a generator network that attempts to fool the discriminator. Stable algorithms for finding Nash equilibria remain an important research direction. Like many other models, GANs can also be applied to semi-supervised learning.

Abstract 6: **Learning features to compare distributions in Adversarial Training**, 11:00 AM

An important component of GANs is the discriminator, which tells apart samples from the generator and samples from a reference set. Discriminators implement empirical approximations to various divergence measures between probability densities (originally Jensen-Shannon, and more recently other f-divergences and integral probability metrics). If we think about this problem in the setting of hypothesis testing, a good discriminator can tell generator samples from reference samples with high probability: in other words, it maximizes the test power. A reasonable goal then becomes to learn a discriminator to directly maxmize test power (we will briefly look at relations between test power and classifier performance).

I will demonstrate ways of training a discriminator with maximum test power using two divergence measures: the maximum mean discrepancy (MMD), and differences of learned smooth features (the ME test, NIPS 2016). In both cases, the key point is that variance matters: it is not enough to have a large empirical divergence; we also need to have high confidence in the value of our divergence. Using an optimized MMD discriminator, we can detect subtle differences in the distribution of GAN outputs and real hand-written digits which humans are unable to find (for instance, small imbalances in the proportions of certain digits, or minor distortions that are implausible in normal handwriting).

Abstract 7: **Training Generative Neural Samplers using Variational Divergence in Adversarial Training**, *Nowozin* 11:30 AM

Generative neural samplers are probabilistic models that implement sampling using feedforward neural networks: they take a random input vector and produce a sample from a probability distribution defined by the network weights. These models are expressive and allow efficient computation of samples and derivatives, but cannot be used for computing likelihoods or for marginalization. The generative-adversarial training method allows to train such models through the use of an auxiliary discriminative neural network. We show that the generative-adversarial approach is a special case of an existing more general variational divergence estimation approach. We show that any f-divergence can be used for training generative neural samplers. We discuss the benefits of various choices of divergence functions on training complexity and the quality of the obtained generative models.

Abstract 9: **Adversarially Learned Inference (ALI) and BiGANs in Adversarial Training**, *Courville* 02:00 PM

We introduce the adversarially learned inference (ALI) model, which jointly learns a generation network and an inference network using an adversarial process. The generation network maps samples from stochastic latent variables to the data space while the inference network maps training examples in data space to the space of latent variables. An adversarial game is cast between these two networks and a discriminative network that is trained to distinguish between joint latent/data-space samples from the generative network and joint samples from the inference network. We illustrate the ability of the model to learn mutually coherent inference and generation networks through the inspections of model samples and reconstructions and confirm the usefulness of the learned representations by obtaining a performance competitive with other recent approaches on the semi-supervised SVHN task.

Abstract 11: **Discussion panel in Adversarial Training**, *Goodfellow, Chintala, Gretton, Nowozin, Courville, LeCun, Denton* 03:00 PM

Submit your questions to

https://www.reddit.com/r/MachineLearning/comments/5fm66i/d_nips_2016_ask_a_w

Abstract 13: **Spotlight presentations in Adversarial Training**, 04:30 PM

David Pfau and Oriol Vinyals. Connecting Generative Adversarial Networks and Actor-Critic Methods

Shakir Mohamed and Balaji Lakshminarayanan. Learning in Implicit Generative Models

Guim Perarnau, Joost Van De Weijer, Bogdan Raducanu and Jose M. Álvarez. Invertible Conditional GANs for image editing

Augustus Odena, Christopher Olah and Jonathon Shlens. Conditional Image Synthesis with Auxiliary Classifier GANs

Luke Metz, Ben Poole, David Pfau and Jascha Sohl-Dickstein. Unrolled Generative Adversarial Networks

Chelsea Finn, Paul Christiano, Pieter Abbeel and Sergey Levine. A Connection Between Generative Adversarial Networks, Inverse Reinforcement Learning, and Energy-Based Models

Pauline Luc, Camille Couprie, Soumith Chintala and Jakob Verbeek. Semantic Segmentation using Adversarial Networks

Tarik Arici and Asli Celikyilmaz. Associative Adversarial Networks

Nina Narodytska and Shiva Kasiviswanathan. Simple Black-Box Adversarial Perturbations for Deep Networks

Pedro Tabacof, Julia Tavares and Eduardo Valle. Adversarial Images for Variational Autoencoders

Yuhuai Wu, Yuri Burda, Ruslan Salakhutdinov and Roger Grosse. On the Quantitative Analysis of Decoder-Based Generative Models

Takeru Miyato, Andrew Dai and Ian Goodfellow. Adversarial Training Methods for Semi-Supervised Text Classification

Abstract 14: **Poster session in Adversarial Training**, 06:00 PM

The posters will be up and running from the beginning of the day, and accessible during all breaks. However, from this point in time we leave the room for their dedicated exposition and discussion.

Browse the list of papers at https://sites.google.com/site/nips2016adversarial/home/accepted-papers

Abstract 15: **Additional poster and open discussions in Adversarial Training**, 06:00 PM

The posters will be up and running from the beginning of the day, and accessible during all breaks.

However, from this point in time we leave the room for their dedicated exposition and discussion.

Browse the list of papers at https://sites.google.com/site/nips2016adversarial/home/accepted-papers

Nonconvex Optimization for Machine Learning: Theory and Practice

Hossein Mobahi, Anima Anandkumar, Percy S Liang, Stefanie Jegelka, Anna E Choromanska

Area 5 + 6, Fri Dec 09, 08:00 AM

A large body of machine learning problems require solving nonconvex optimization. This includes deep learning, Bayesian inference, clustering, and so on. The objective functions in all these instances are highly non-convex, and it is an open question if there are provable, polynomial time algorithms for these problems under realistic assumptions. A diverse set of approaches have been devised to solve nonconvex problems in a variety of approaches. They range from simple local search approaches such as gradient descent and alternating minimization to more involved frameworks such as simulated annealing, continuation method, convex hierarchies, Bayesian optimization, branch and bound, and so on. Moreover, for solving special class of nonconvex problems there are efficient methods such as quasi convex optimization, star convex optimization, submodular optimization, and matrix/tensor decomposition. There has been a burst of recent research activity in all these areas. This workshop brings researchers from these vastly different domains and hopes to create a dialogue among them. In addition to the theoretical frameworks, the workshop will also feature practitioners, especially in the area of deep learning who are developing new methodologies for training large scale neural networks. The result will be a cross fertilization of ideas from diverse areas and schools of thought.

Schedule

08:15 AM	**Opening Remarks**	
08:30 AM	**Learning To Optimize**	*de Freitas*
09:00 AM	**Morning Poster Spotlight**	
09:30 AM	**Morning Poster Session**	
10:30 AM	**Coffee Break**	
11:00 AM	**The moment-LP and moment-SOS approaches in optimization and some related applications**	*Lasserre*
11:30 AM	**Non-convexity in the error landscape and the expressive capacity of deep neural networks**	*Ganguli*
12:00 PM	**Leveraging Structure in Bayesian Optimization**	*Adams*
12:30 PM	**Lunch Break**	
01:30 PM	**Submodular Optimization and Nonconvexity**	*Jegelka*
02:00 PM	**Submodular Functions: from Discrete to Continuous Domains**	*Bach*
02:30 PM	**Taming non-convexity via geometry**	*Sra*
03:00 PM	**Break**	
03:30 PM	**Discussion Panel**	
04:30 PM	**Afternoon Poster Spotlight**	
05:00 PM	**Afternoon Poster Session**	

Abstracts (7):

Abstract 2: **Learning To Optimize in Nonconvex Optimization for Machine Learning: Theory and Practice**, *de Freitas* 08:30 AM

The move from hand-designed features to learned features in machine learning has been wildly successful. In spite of this, optimization algorithms are still designed by hand. In this talk I describe how the design of an optimization algorithm can be cast as a learning problem, allowing the algorithm to learn to exploit structure in the problems of interest in an automatic way. The learned algorithms, implemented by LSTMs, outperform generic, hand-designed competitors on the tasks for which they are trained, and also generalize well to new tasks with similar structure.

Abstract 6: **The moment-LP and moment-SOS approaches in optimization and some related applications in Nonconvex Optimization for Machine Learning: Theory and Practice**, *Lasserre* 11:00 AM

In a first part we provide an introduction to the basics of the moment-LP and moment-SOS approaches to global polynomial optimization. In particular, we describe the hierarchy of LP and semidefinite programs to approximate the optimal value of such problems. In fact, the same methodology also applies to solve (or approximate) Generalized Moment Problems (GMP) whose data are described by basic semi-algebraic sets and polynomials (or even semi-algebraic functions). Indeed, Polynomial optimization is a particular (and even the simplest) instance of the GMP.

In a second part, we describe how to use this methodology for solving some problems (outside optimization) viewed as particular instances of the GMP. This includes:
- Approximating compact basic semi-algebraic sets defined by quantifiers.
- Computing convex polynomials underestimators of polynomials on a box.
- Bounds on measures satisfying some moment conditions.
- Approximating the volume of compact basic semi-algebraic sets.
- Approximating the Gaussian measure of non-compact basic semi-algebraic sets.
- Approximating the Lebesgue decomposition of a measure μ w.r.t. another measure ν, based only on the moments of μ and ν.

Abstract 7: **Non-convexity in the error landscape and the expressive capacity of deep neural networks in Nonconvex Optimization for Machine Learning: Theory and Practice**, *Ganguli* 11:30 AM

A variety of recent work has studied saddle points in the error landscape of deep neural networks. A clearer understanding of these saddle points is likely to arise from an understanding of the geometry of deep functions. In particular, what do the generic functions computed by a deep network “look like?” How can we quantify and understand their geometry, and what implications does this geometry have for reducing generalization error as well as training error? We combine Riemannian geometry with the mean field theory of high dimensional chaos to study the nature of generic deep functions. Our results reveal an order-to-chaos expressivity phase transition, with networks in the chaotic phase computing nonlinear functions whose global curvature grows exponentially with depth but not width. Moreover, we formalize and quantitatively demonstrate the long conjectured idea that deep networks can disentangle highly curved manifolds in input space into flat manifolds in hidden space. Our theoretical analysis of the expressive power of deep networks broadly applies to arbitrary nonlinearities, and provides intuition for why initializations at the edge of chaos can lead to both better optimization as well as superior generalization capabilities.

Abstract 8: **Leveraging Structure in Bayesian Optimization in Nonconvex Optimization for Machine Learning: Theory and Practice**, *Adams* 12:00 PM

Bayesian optimization is an approach to non-convex optimization that leverages a probabilistic model to make decisions about candidate points to evaluate. The primary advantage of this approach is the ability to incorporate prior knowledge about the objective function in an explicit way. While such prior information has typically been information about the smoothness of the function, many machine learning problems have additional structure that can be leveraged. I will talk about how such prior information can be found across tasks, within inner-loop optimizations, and in constraints.

Abstract 10: **Submodular Optimization and Nonconvexity in Nonconvex Optimization for Machine Learning: Theory and Practice**, *Jegelka* 01:30 PM

Despite analogies of submodularity and convexity, submodular optimization is closely connected with certain "nice" non-convex optimization problems for which theoretical guarantees are still possible. In this talk, I will review some of these connections and make them specific at the example of a challenging robust influence maximization problem, for which we obtain new, tractable formulations and algorithms.

Abstract 11: **Submodular Functions: from Discrete to Continuous Domains in Nonconvex Optimization for Machine Learning: Theory and Practice**, *Bach* 02:00 PM

Submodular set-functions have many applications in combinatorial optimization, as they can be minimized and approximately maximized in polynomial time. A key element in many of the algorithms and analyses is the possibility of extending the submodular set-function to a convex function, which opens up tools from convex optimization. Submodularity goes beyond set-functions and has naturally been considered for problems with multiple labels or for functions defined on continuous domains, where it corresponds essentially to cross second-derivatives being nonpositive. In this paper, we show that most results relating submodularity and convexity for set-functions can be extended to all submodular functions. In particular, (a) we naturally define a continuous extension in a set of probability measures, (b) show that the extension is convex if and only if the original function is submodular, (c) prove that the problem of minimizing a submodular function is equivalent to a typically non-smooth convex optimization problem, and (d) propose another convex optimization problem with better computational properties (e.g., a smooth dual problem). Most of these extensions from the set-function situation are obtained by drawing links with the theory of multi-marginal optimal transport, which provides also a new interpretation of existing results for set-functions. We then provide practical algorithms to minimize generic submodular functions on discrete domains, with associated convergence rates.

Abstract 12: **Taming non-convexity via geometry in Nonconvex Optimization for Machine Learning: Theory and Practice**, *Sra* 02:30 PM

In this talk, I will highlight some aspects of geometry and its role in optimization. In particular, I will talk about optimization problems whose parameters are constrained to lie on a manifold or in a specific metric space. These geometric constraints often make the problems numerically challenging, but they can also unravel properties that ensure tractable attainment of global optimality for certain otherwise non-convex problems.

We'll make our foray into geometric optimization via geodesic convexity, a concept that generalizes the usual notion of convexity to nonlinear metric spaces such as Riemannian manifolds. I will outline some of our results that contribute to g-convex analysis as well as to the theory of first-order g-convex optimization. I will mention several very interesting optimization problems where g-convexity proves remarkably useful. In closing, I will mention extensions to large-scale non-convex geometric optimization as well as key open problems.

Efficient Methods for Deep Neural Networks

Mohammad Rastegari, Matthieu Courbariaux

Area 7 + 8, Fri Dec 09, 08:00 AM

Deep Neural Networks have been revolutionizing several application domains in artificial intelligence: Computer Vision, Speech Recognition and Natural Language Processing. Concurrent to the recent progress in deep learning, significant progress has been happening in virtual reality, augmented reality, and smart wearable devices. These advances create unprecedented opportunities for researchers to tackle fundamental challenges in deploying deep learning systems to portable devices with limited resources (e.g. Memory, CPU, Energy, Bandwidth). Efficient methods in deep learning can have crucial impacts in using distributed systems, embedded devices, and FPGA for several AI tasks. Achieving these goals calls for ground-breaking innovations on many fronts: learning, optimization, computer architecture, data compression, indexing, and hardware design.

This workshop is sponsored by Allen Institute for Artificial Intelligence (AI2). We offer partial travel grant and registration for limited number of people participating in the workshop.

The goal of this workshop is providing a venue for researchers interested in developing efficient techniques for deep neural networks to present new work, exchange ideas, and build connections. The workshop will feature keynotes and invited talks from prominent researchers as well as a poster session that fosters in depth discussion. Further, in a discussion panel the experts discuss about the possible approaches (hardware, software, algorithm, ...) toward designing efficient methods in deep learning.

We invite submissions of short papers and extended abstracts related to the following topics in the context of efficient methods in deep learning:

-Network compression
-Quantized neural networks (e.g. Binary neural networks)
-Hardware accelerator for neural networks
-Training and inference with low-precision operations.
-Real-time applications in deep neural networks (e.g. Object detection, Image segmentation, Online language translation, ...)
-Distributed training/inference of deep neural networks
-Fast optimization methods for neural networks

Schedule

09:00 AM	**Mohammad Rastegari: Introductory remarks**	*Rastegari*
09:15 AM	**William Dally: Efficient Methods and Hardware for Deep Neural Networks**	*Dally*

09:45 AM	**Amir Khosrowshahi: Processor architectures for deep learning**	*Khosrowshahi*
11:00 AM	**Ali Farhadi: Deep Learning on Resource Constraint Devices**	*Farhadi*
11:30 AM	**Oral Presentations (Session A)**	
12:00 PM	**Lunch (on your own)**	
01:30 PM	**Vivienne Sze: Joint Design of Algorithms and Hardware for Energy-efficient DNNs**	*Sze*
02:00 PM	**Yoshua Bengio: From Training Low Precision Neural Nets to Training Analog Continuous-Time Machines**	*Bengio*
02:30 PM	**Poster presentations and Coffee break**	
03:30 PM	**Kurt Keutzer: High-Performance Deep Learning**	*Keutzer*
04:00 PM	**Oral Presentations (Session B)**	
04:30 PM	**Mohammad Rastegari: Closing remarks**	

The Future of Interactive Machine Learning

Kory Mathewson, Kaushik Subramanian, Mark K Ho, Robert Loftin, Joe L Austerweil, Anna Harutyunyan, Doina Precup, Layla El Asri, Matthew Gombolay, Jerry Zhu, Sonia Chernova, Charles L Isbell, Patrick M Pilarski, Weng-Keen Wong, Manuela Veloso, Julie A Shah, Matthew Taylor, Brenna Argall, Michael Littman

Hilton Diag. Mar, Blrm. A, Fri Dec 09, 08:00 AM

Interactive machine learning (IML) explores how intelligent agents solve a task together, often focusing on adaptable collaboration over the course of sequential decision making tasks. Past research in the field of IML has investigated how autonomous agents can learn to solve problems more effectively by making use of interactions with humans. Designing and engineering fully autonomous agents is a difficult and sometimes intractable challenge. As such, there is a compelling need for IML algorithms that enable artificial and human agents to collaborate and solve independent or shared goals. The range of real-world examples of IML spans from web applications such as search engines, recommendation systems and social media personalization, to dialog systems and embodied systems such as industrial robots and household robotic assistants, and to medical robotics (e.g. bionic limbs, assistive devices, and exoskeletons). As intelligent systems become more common in industry and in everyday life, the need for these systems to interact with and learn from the people around them will also increase.

This workshop seeks to brings together experts in the fields of IML, reinforcement learning (RL), human-computer interaction (HCI), robotics, cognitive psychology and the social sciences to share recent advances and explore the future of IML. Some questions of particular interest for this workshop include: How can recent advancements in machine learning allow interactive learning to be deployed in current real world applications? How do we address the challenging problem of seamless communication between autonomous agents and humans? How can we improve the ability to collaborate safely and successfully across a diverse set of users?

We hope that this workshop will produce several outcomes:
- A review of current algorithms and techniques for IML, and a focused perspective on what is lacking;
- A formalization of the main challenges for deploying modern interactive learning algorithms in the real world; and
- A forum for interdisciplinary researchers to discuss open problems and challenges, present new ideas on IML, and plan for future collaborations.

Topics relevant to this workshop include:
Human-robot interaction
Collaborative and/or shared control
Semi-supervised learning with human intervention
Learning from demonstration, interaction and/or observation
Reinforcement learning with human-in-the-loop
Active learning, Preference learning
Transfer learning (human-to-machine, machine-to-machine)
Natural language processing for dialog systems
Computer vision for human interaction with autonomous systems
Transparency and feedback in machine learning
Computational models of human teaching
Intelligent personal assistants and dialog systems
Adaptive user interfaces
Brain-computer interfaces (e.g. human-semi-autonomous system interfaces)
Intelligent medical robots (e.g. smart wheelchairs, prosthetics, exoskeletons)

Schedule

08:20 AM	**Opening Remarks, Invited Talk: Michael C. Mozer**	*Mozer*
09:10 AM	**A Human-in-the-loop Approach for Troubleshooting Machine Learning Systems, Besmira Nushi, Ece Kamar, Donald Kossmann and Eric Horvitz**	
09:30 AM	**Efficient Exploration in Monte Carlo Tree Search using Human Action Abstractions, Kaushik Subramanian, Jonathan Scholz, Charles Isbell and Andrea Thomaz**	
09:50 AM	**Invited Talk: Mattew E. Taylor**	*Taylor*
10:30 AM	**Coffee Break 1**	

11:00 AM	**Invited Talk: Olivier Pietquin**	*Pietquin*
11:40 AM	**Poster Spotlight Talks 1**	
12:10 PM	**Invited Talk: Todd Gureckis**	*Gureckis*
12:50 PM	**Lunch Break**	
02:00 PM	**Poster Spotlight Talks 2**	
02:30 PM	**Invited Talk: Aude Billard**	*Billard*
03:10 PM	**Coffee Break 2**	
03:30 PM	**Poster Session**	
04:30 PM	**Enabling Robots to Communicate Reward Functions, Sandy Huang, David Held, Pieter Abbeel and Anca Dragan**	
04:50 PM	**Hierarchical Multi-Agent Reinforcement Learning through Communicative Actions for Human-Robot Collaboration, Elena Corina Grigore and Brian Scassellati**	
05:10 PM	**Invited Talk: Emma Brunskill**	*Brunskill*
05:50 PM	**Panel Discussion, Closing Remarks**	

Abstracts (6):

Abstract 2: **A Human-in-the-loop Approach for Troubleshooting Machine Learning Systems, Besmira Nushi, Ece Kamar, Donald Kossmann and Eric Horvitz in The Future of Interactive Machine Learning**, 09:10 AM

We study the problem of troubleshooting machine learning systems that rely on analytical pipelines of distinct components. Understanding and fixing errors that arise in such integrative systems is difficult as failures can occur at multiple points in the execution workflow. Moreover, errors can propagate, become amplified or be suppressed, making blame assignment difficult. We propose a human-in-the-loop methodology which leverages human intellect for troubleshooting system failures. The approach simulates potential component fixes through human computation tasks and measures the expected improvements in the holistic behavior of the system. The method provides guidance to designers about how they can best improve the system. We demonstrate the effectiveness of the approach on an automated image captioning system that has been pressed into real-world use.

Abstract 3: **Efficient Exploration in Monte Carlo Tree Search using Human Action Abstractions, Kaushik Subramanian, Jonathan Scholz, Charles Isbell and Andrea Thomaz in The Future of Interactive Machine Learning**, 09:30 AM

Monte Carlo Tree Search (MCTS) is a family of methods for planning in large domains. It focuses on finding a good action for a particular state, making its complexity independent of the size of the state space. However such methods are exponential with respect to the branching factor. Effective application of MCTS requires good heuristics to arbitrate action selection during learning. In this paper we present a policy-guided approach that utilizes action abstractions, derived from human input, with MCTS to facilitate efficient exploration. We draw from existing work in hierarchical reinforcement learning, interactive machine learning and show how multi-step actions, represented as stochastic policies, can serve as good action selection heuristics. We demonstrate the efficacy of our approach in the PacMan domain and highlight its advantages over traditional MCTS.

Abstract 7: **Poster Spotlight Talks 1 in The Future of Interactive Machine Learning**, 11:40 AM

SPARC: an efficient way to combine reinforcement learning and supervised autonomy, Emmanuel Senft, Paul Baxter, Séverin Lemaignan and Tony Belpaeme

Near-optimal Bayesian Active Learning with Correlated and Noisy Tests, Yuxin Chen, Hamed Hassani and Andreas Krause

A Multimodal Human-Robot Interaction Dataset, Pablo Azagra, Yoan Mollard, Florian Golemo, Ana Cristina Murillo, Manuel Lopes and Javier Civera

Cross-Entropy as a Criterion for Robust Interactive Learning of Latent Properties, Johannes Kulick, Robert Lieck and Marc Toussaint

Ensemble Co-Training of Image and EEG-based RSVP Classifiers for Improved Image Triage, Steven Gutstein, Vernon Lawhern and Brent Lance

Active Reinforcement Learning: Observing Rewards at a Cost, David Krueger, Owain Evans, Jan Leike and John Salvatier

ReVACNN: Steering Convolutional Neural Network via Real-Time Visual Analytics, Sunghyo Chung, Cheonbok Park, Sangho Suh, Kyeongpil Kang, Jaegul Choo and Bum Chul Kwon

Analysis of a Design Pattern for Teaching with Features and Labels, Christopher Meek, Patrice Simard and Jerry Zhu

Agent-Agnostic Human-in-the-Loop Reinforcement Learning, David Abel, Owain Evans, John Salvatier and Andreas Stuhlmüller

Abstract 10: **Poster Spotlight Talks 2 in The Future of Interactive Machine Learning**, 02:00 PM

Probabilistic Expert Knowledge Elicitation of Feature Relevances in Sparse Linear Regression, Pedram Daee, Tomi Peltola, Marta Soare and Samuel Kaski

Socratic Learning, Paroma Varma, Rose Yu, Dan Iter, Chris De Sa and Christopher Re

Probabilistic Active Learning for Active Class Selection, Daniel Kottke, Georg Krempl, Marianne Stecklina, Cornelius Styp von Rekowski, Tim Sabsch, Tuan Pham Minh, Matthias Deliano, Myra Spiliopoulou and Bernhard Sick

Regression Analysis in Small-n-Large-p Using Interactive Prior Elicitation of Pairwise Similarities, Homayun Afrabandpey, Tomi Peltola and Samuel Kaski

Scalable batch mode Optimal Experimental Design for Deep Networks, Mélanie Ducoffe, Geoffrey Portelli and Frederic Precioso

Interactive Preference Learning of Utility Functions for Multi-Objective Optimization, Ian Dewancker, Michael Mccourt and Samuel Ainsworth

Improving Online Learning of Visual Categories by Deep Features, Lydia Fischer, Stephan Hasler, Sebastian Schrom and Heiko Wersing

Interactive user intent modeling for eliciting priors of a normal linear model, Iiris Sundin, Luana Micallef, Pekka Marttinen, Muhammad Ammad-Ud-Din, Tomi Peltola, Marta Soare, Giulio Jacucci and Samuel Kaski

Training an Interactive Humanoid Robot Using Multimodal Deep Reinforcement Learning, Heriberto Cuayahuitl, Guillaume Couly and Clement Olalainty

Abstract 14: **Enabling Robots to Communicate Reward Functions, Sandy Huang, David Held, Pieter Abbeel and Anca Dragan in The Future of Interactive Machine Learning**, 04:30 PM

Understanding a robot's reward function is key to anticipating how the robot will act in a new situation. Our goal is to generate a set of robot behaviors that best illustrates a robot's reward function. We build on prior work modeling inference of the reward function from example behavior via Inverse Reinforcement Learning (IRL). Prior work using IRL has focused on people teaching machines and assumes exact inference. Our insight is that when teaching people, they will not perform exact inference. We show that while leveraging models of noisy inference can be beneficial, it is also important to achieve coverage in the space of possible strategies the robot can use. We introduce a hybrid algorithm that targets informative examples via both a noisy inference model and coverage.

Abstract 15: **Hierarchical Multi-Agent Reinforcement Learning through Communicative Actions for Human-Robot Collaboration, Elena Corina Grigore and Brian Scassellati in The Future of Interactive Machine Learning**, 04:50 PM

As we expect robots to start moving from working in isolated industry settings into human populated environments, our need to develop suitable learning algorithms for the latter increases. Human-robot collaboration is a particular area that has tremendous gains from endowing a robot with such learning capabilities, focusing on robots that can work side-by-side with a human and provide supportive behaviors throughout a task executed by the human worker. In this paper, we propose a framework based on hierarchical multi-agent reinforcement learning that considers the human as an ``expert'' agent in the system—an agent whose actions we cannot control but whose actions, jointly with the robot's actions, impact the state of the task. Our framework aims to provide the learner (the robot) with a way of learning how to provide supportive behaviors to the expert agent (the person) during a complex task. The robot employs communicative actions to interactively learn from the expert agent at key points during the task. We use a hierarchical approach in order to integrate the communicative actions in the multi-agent reinforcement learning framework and allow for simultaneously learning the quality of performing different supportive behaviors for particular combinations of task states and expert agent actions. In this paper, we present our proposed framework, detail the motion capture system data collection we performed in order to learn about the task states and characterize the expert agent's actions, and discuss how we can apply the framework to our human-robot collaboration scenario.

Cognitive Computation: Integrating Neural and Symbolic Approaches

Tarek R. Besold, Antoine Bordes, Greg Wayne, Artur Garcez

Hilton Diag. Mar, Blrm. B, Fri Dec 09, 08:00 AM

While early work on knowledge representation and inference was primarily symbolic, the corresponding approaches subsequently fell out of favor, and were largely supplanted by connectionist methods. In this workshop, we will work to close the gap between the two paradigms, and aim to formulate a new unified approach that is inspired by our current understanding of human cognitive processing. This is important to help improve our understanding of Neural Information Processing and build better Machine Learning systems, including the integration of learning and reasoning in dynamic knowledge-bases, and reuse of knowledge learned in one application domain in analogous domains.

The workshop brings together established leaders and promising young scientists in the fields of neural computation, logic and artificial intelligence, knowledge representation, natural language understanding, machine learning, cognitive science and computational neuroscience. Invited lectures by senior researchers will be complemented with presentations based on contributed papers reporting recent work (following an open call for papers) and a poster session, giving ample opportunity for participants to interact and discuss the complementary perspectives and emerging approaches.

The workshop targets a single broad theme of general interest to the vast majority of the NIPS community, namely translations between connectionist models and symbolic knowledge representation and reasoning for the purpose of achieving an effective integration of neural learning and cognitive reasoning, called neural-symbolic computing. The study of neural-symbolic computing is now an established topic of wider interest to NIPS with topics that are relevant to almost everyone studying neural information processing. In the 2016 edition of the workshop, special emphasis will be put on language-related aspects and applications of neural-symbolic integration and relevant cognitive computation paradigms.

Keywords: neural-symbolic computing; language processing and reasoning; cognitive agents; multimodal learning; deep networks; knowledge extraction; symbol manipulation; variable binding; memory-based networks; dynamic knowledge-bases.

Schedule

08:45 AM	**Welcome/Opening**	*Besold, Bordes, Wayne, Garcez*

09:00 AM	**"Neuro-symbolic EDA-based Optimisation using ILP-enhanced DBNs" (Sarmimala Saikia, Lovekesh Vig, Ashwin Srinivasan, Gautam Shroff, Puneet Agarwal, Rawat Richa)**	
09:30 AM	**Invited talk Barbara Hammer (Bielefeld University, Germany)**	*Hammer*
10:00 AM	**Invited talk Risto Miikkulainen (University of Texas at Austin & Sentient Technologies Inc., USA)**	*Miikkulainen*
11:00 AM	**Invited talk Kristina Toutanova (Microsoft Research Redmond, USA)**	*Toutanova*
11:30 AM	**Poster Pitches**	
11:50 AM	**Poster Presentations**	
12:45 PM	**Lunch break**	
02:00 PM	**"Variable binding through assemblies in spiking neural networks" (Robert Legenstein, Christos Papadimitriou, Santosh Vempala, Wolfgang Maass)**	
02:20 PM	**"Pre-Wiring and Pre-Training: What does a neural network need to learn truly general identity rules?" (Raquel Alhama, Willem Zuidema)**	
02:40 PM	**"ReasoNet: Learning to Stop Reading in Machine Comprehension" (Yelong Shen, Po-Sen Huang, Jianfeng Gao, Weizhu Chen)**	
03:00 PM	**Coffee break**	
03:30 PM	**Invited talk Dan Roth (University of Illinois at Urbana-Chambaign, USA)**	*Roth*
04:00 PM	**Panel on "Explainable AI" (Yoshua Bengio, Alessio Lomuscio, Gary Marcus, Stephen Muggleton, Michael Witbrock)**	*Bengio, Lomuscio, Marcus, Muggleton, Witbrock*
05:25 PM	**Summary/Goodbye**	*Besold, Garcez, Bordes, Wayne*

Abstracts (2):

Abstract 6: **Poster Pitches in Cognitive Computation: Integrating Neural and Symbolic Approaches**, 11:30 AM

1) "Analogy-based Reasoning With Memory Networks for Future Prediction" (Daniel Andrade, Bing Bai, Ramkumar Rajendran, Yotaro Watanabe)
2) "Multiresolution Recurrent Neural Networks: An Application to Dialogue Response Generation" (Iulian Vlad Serban, Tim Klinger, Gerald Tesauro, Kartik Talamadupula, Bowen Zhou, Yoshua Bengio, Aaron Courville)
3) "Crossmodal language grounding, learning, and teaching" (Stefan Heinrich, Cornelius Weber, Stefan Wermter, Ruobing Xie, Yankai Lin, Zhiyuan Liu)
4) "Diagnostic classifiers: revealing how neural networks process hierarchical structure" (Sara Veldhoen, Dieuwke Hupkes, Willem Zuidema)
5) "Neuro-symbolic EDA-based Optimisation using ILP-enhanced DBNs" (Sarmimala Saikia, Lovekesh Vig, Ashwin Srinivasan, Gautam Shroff, Puneet Agarwal, Rawat Richa)
"6) Top-Down and Bottom-Up Interactions between Low-Level Reactive Control and Symbolic Rule Learning in Embodied Agents" (Clement Moulin-Frier, Xerxes Arsiwalla, Jordi-Ysard Puigbo, Marti Sanchez-Fibla, Armin Duff, Paul Verschure)
7) "Accuracy and Interpretability Trade-offs in Machine Learning Applied to Safer Gambling" (Sanjoy Sankar, Tillman Weyde, Artur D'Avila Garcez, Gregory Slabaugh, Simo Dragicevic, Chris Percy)
8) "A Simple but Tough-to-Beat Baseline for Sentence Embeddings" (Sanjeev Arora, Yingyu Liang, Tengyu Ma)
9) "MS MARCO: A Human-Generated MAchine Reading COmprehension Dataset" (Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, Li Deng)

Abstract 7: **Poster Presentations in Cognitive Computation: Integrating Neural and Symbolic Approaches**, 11:50 AM

1) "Analogy-based Reasoning With Memory Networks for Future Prediction" (Daniel Andrade, Bing Bai, Ramkumar Rajendran, Yotaro Watanabe)
2) "Multiresolution Recurrent Neural Networks: An Application to Dialogue Response Generation" (Iulian Vlad Serban, Tim Klinger, Gerald Tesauro, Kartik Talamadupula, Bowen Zhou, Yoshua Bengio, Aaron Courville)
3) "Crossmodal language grounding, learning, and teaching" (Stefan Heinrich, Cornelius Weber, Stefan Wermter, Ruobing Xie, Yankai Lin, Zhiyuan Liu)
4) "Diagnostic classifiers: revealing how neural networks process hierarchical structure" (Sara Veldhoen, Dieuwke Hupkes, Willem Zuidema)
5) "Neuro-symbolic EDA-based Optimisation using ILP-enhanced DBNs" (Sarmimala Saikia, Lovekesh Vig, Ashwin Srinivasan, Gautam Shroff, Puneet Agarwal, Rawat Richa)
"6) Top-Down and Bottom-Up Interactions between Low-Level Reactive Control and Symbolic Rule Learning in Embodied Agents" (Clement Moulin-Frier, Xerxes Arsiwalla, Jordi-Ysard Puigbo, Marti Sanchez-Fibla, Armin Duff, Paul Verschure)
7) "Accuracy and Interpretability Trade-offs in Machine Learning Applied to Safer Gambling" (Sanjoy Sankar, Tillman Weyde, Artur D'Avila Garcez, Gregory Slabaugh, Simo Dragicevic, Chris Percy)
8) "A Simple but Tough-to-Beat Baseline for Sentence Embeddings" (Sanjeev Arora, Yingyu Liang, Tengyu Ma)
9) "MS MARCO: A Human-Generated MAchine Reading

COmprehension Dataset" (Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, Li Deng)

Intuitive Physics

Jiajun Wu, Adam Lerer, Josh Tenenbaum, Emmanuel Dupoux, Rob Fergus

Hilton Diag. Mar, Blrm. C, Fri Dec 09, 08:00 AM

Despite recent progress, AI is still far away from achieving common sense reasoning. One area that is gathering a lot of interest is that of intuitive or naive physics. It concerns the ability that humans and, to a certain extent, infants and animals have to predict outcomes of physical interactions involving macroscopic objects. There is extensive experimental evidence that infants can predict the outcome of events based on physical concepts such as gravity, solidity, object permanence and conservation of shape and number, at an early stage of development, although there is also evidence that this capacity develops through time and experience. Recent work has attempted to build neural models that can make predictions about stability, collisions, forces and velocities from images or videos, or interactions with an environment. Such models could be both used to understand the cognitive and neural underpinning of naive physics in humans, but also to provide with AI applications more better inference and reasoning abilities.

This workshop will bring together researchers in machine learning, computer vision, robotics, computational neuroscience, and cognitive development to discuss artificial systems that capture or model intuitive physics by learning from footage of, or interactions with a real or simulated environment. There will be invited talks from world leaders in the fields, presentations and poster sessions based on contributed papers, and a panel discussion.

Topics of discussion will include:
- Learning models of Newtonian physics (deep networks, structured probabilistic generative models, physics engines)
- How to combine model-based and bottom-up approaches to intuitive physics
- Application of intuitive physics models to higher-level tasks such as navigation, video prediction, robotics, etc.
- How cognitive science and computational neuroscience literature may inform the design of artificial systems for physical prediction
- Methodology for comparing models of infant learning with clinical studies
- Development of new datasets or platforms for intuitive physics and visual commonsense

Schedule

08:40 AM	**Opening Remarks**	*Tenenbaum*
09:00 AM	**Naive Physics 101: A Tutorial**	*Dupoux, Tenenbaum*
09:30 AM	**Poster Spotlights**	
10:00 AM	**Ali Farhadi**	*Farhadi*
10:30 AM	**Coffee Break / Posters 1**	
11:00 AM	**Peter Battaglia**	*battaglia*
11:30 AM	**Jitendra Malik and Pulkit Agrawal**	*Malik, Agrawal*
12:00 PM	**Lunch Break**	
02:00 PM	**Abhinav Gupta**	*Gupta*
02:30 PM	**Bill Freeman**	*Freeman*
03:00 PM	**Coffee Break / Posters 2**	
03:30 PM	**Imagination-Based Decision Making with Physical Models in Deep Neural Networks**	*Hamrick*
03:50 PM	**Visual Stability Prediction and Its Application to Manipulation**	*Li*
04:10 PM	**Deep Visual Foresight for Planning Robot Motion**	*Finn*
04:30 PM	**Datasets, Methodology, and Challenges in Intuitive Physics**	*Dupoux, Tenenbaum*
05:30 PM	**Panel Discussion**	

Extreme Classification: Multi-class and Multi-label Learning in Extremely Large Label Spaces

Moustapha Cisse, Manik Varma, Samy Bengio

Room 111, Fri Dec 09, 08:00 AM

Extreme classification, where one needs to deal with multi-class and multi-label problems involving a very large number of labels, has opened up a new research frontier in machine learning. Many challenging applications, such as photo or video annotation, web page categorization, gene function prediction, language modeling can benefit from being formulated as supervised learning tasks with millions, or even billions, of labels. Extreme classification can also give a fresh perspective on core learning problems such as ranking and recommendation by reformulating them as multi-class/label tasks where each item to be ranked or recommended is a separate label.

Extreme classification raises a number of interesting research questions including those related to:

* Large scale learning and distributed and parallel training
* Log-time and log-space prediction and prediction on a test-time budget
* Label embedding and tree-based approaches
* Crowd sourcing, preference elicitation and other data gathering techniques
* Bandits, semi-supervised learning and other approaches for dealing with training set biases and label noise
* Bandits with an extremely large number of arms
* Fine-grained classification
* Zero shot learning and extensible output spaces
* Tackling label polysemy, synonymy and correlations
* Structured output prediction and multi-task learning
* Learning from highly imbalanced data

* Dealing with tail labels and learning from very few data points per label
* PU learning and learning from missing and incorrect labels
* Feature extraction, feature sharing, lazy feature evaluation, etc.
* Performance evaluation
* Statistical analysis and generalization bounds
* Applications to ranking, recommendation, knowledge graph construction and other domains

The workshop aims to bring together researchers interested in these areas to encourage discussion and improve upon the state-of-the-art in extreme classification. In particular, we aim to bring together researchers from the natural language processing, computer vision and core machine learning communities to foster interaction and collaboration. Several leading researchers will present invited talks detailing the latest advances in the area. We also seek extended abstracts presenting work in progress which will be reviewed for acceptance as spotlight+poster or a talk. The workshop should be of interest to researchers in core supervised learning as well as application domains such as recommender systems, computer vision, computational advertising, information retrieval and natural language processing. We expect a healthy participation from both industry and academia.

http://www.manikvarma.org/events/XC16/schedule.html

Schedule

09:00 AM	**Opening Remarks by Manik, Moustapha & Samy**
09:05 AM	**Label Ranking with Biased Partial Feedback**
09:35 AM	**Distributed Optimization of Multi-Class SVMs**
09:50 AM	**DiSMEC - Distributed Sparse Machines for Extreme Multi-label Classification**
10:05 AM	**A Primal and Dual Sparse Approach to Extreme Classification**
11:00 AM	**Extreme Multi-label Loss Functions for Tagging, Ranking & Recommendation**
11:30 AM	**Log-time and Log-space Extreme Classification**
11:45 AM	**Extreme Classification with Label Features**
12:00 PM	**Dual Decomposed Learning with Factorwise Oracles for Structural SVMs of Large Output Domain**
12:15 PM	**Lunch**
01:30 PM	**Semi-supervised dimension reduction for large numbers of classes**
02:00 PM	**A Theoretical Framework for Structured Prediction using Factor Graph Complexity**
02:15 PM	**Deep Schatten Networks**
02:45 PM	**Regret Bounds for Non-decomposable Metrics with Missing Labels**
03:30 PM	**Coffee Break**
03:30 PM	**Training neural networks in time independent of output layer size**
04:00 PM	**Efficient softmax approximation for GPUs**
04:15 PM	**Pointer Sentinel Mixture Models**
04:30 PM	**A Simple but Tough-to-Beat Baseline for Sentence Embeddings**
04:45 PM	**Break**
05:00 PM	**iCaRL: incremental classifier and representation learning**
05:30 PM	**Is a picture worth a thousand words? a Deep Multi Modal Product Classification Architecture for e-commerce**
05:45 PM	**Learning to Solve Vision without Annotating Millions of Images**

Advances in Approximate Bayesian Inference

Tamara Broderick, Stephan Mandt, James McInerney, Dustin Tran, David Blei, Kevin P Murphy, Andrew Gelman, Michael I Jordan

Room 112, Fri Dec 09, 08:00 AM

Bayesian analysis has seen a resurgence in machine learning, expanding its scope beyond traditional applications. Increasingly complex models have been trained with large and streaming data sets, and they have been applied to a diverse range of domains. Key to this resurgence has been advances in approximate Bayesian inference. Variational and Monte Carlo methods are currently the mainstay techniques, where recent insights have improved their approximation quality, provided black box strategies for fitting many models, and enabled scalable computation.

In this year's workshop, we would like to continue the theme of approximate Bayesian inference with additional emphases. In particular, we encourage submissions not only advancing approximate inference but also regarding (1) unconventional inference techniques, with the aim to bring together diverse communities; (2) software tools for both the applied and methodological researcher; and (3) challenges in applications, both in non-traditional domains and when applying these techniques to advance current domains.

Schedule

08:30 AM	**Introduction**
08:35 AM	**Invited talk 1**
09:00 AM	**Contributed talk 1**
09:15 AM	**Invited talk 2**
09:40 AM	**Panel on Advances in Software for Approximate Bayesian Inference**
11:00 AM	**Contributed talk 2**
11:15 AM	**Poster spotlights**
11:35 AM	**Poster session**
02:10 PM	**Invited talk 3**
03:30 PM	**Contributed talk 3**
03:45 PM	**Invited talk 4**
04:10 PM	**Contributed talk 4**
04:25 PM	**Panel On the Foundations and Future of Approximate Inference**

Abstracts (2):

Abstract 5: **Panel on Advances in Software for Approximate Bayesian Inference in Advances in Approximate Bayesian Inference**, 09:40 AM

Noah Goodman (WebPPL; Stanford University)
Dustin Tran (Edward; Columbia University)
Michael Hughes (BNPy; Harvard University)
TBA (TensorFlow, BayesFlow; Google)
TBA (Stan)

Abstract 13: **Panel On the Foundations and Future of Approximate Inference in Advances in Approximate Bayesian Inference**, 04:25 PM

Ryan Adams, Barbara Engelhardt, Philipp Hennig, Richard Turner, Neil Lawrence

Reliable Machine Learning in the Wild

Dylan Hadfield-Menell, Adrian Weller, David Duvenaud, Jacob Steinhardt, Percy S Liang

Room 113, Fri Dec 09, 08:00 AM

When will a system that has performed well in the past continue to do so in the future? How do we design such systems in the presence of novel and potentially adversarial input distributions? What techniques will let us safely build and deploy autonomous systems on a scale where human monitoring becomes difficult or infeasible? Answering these questions is critical to guaranteeing the safety of emerging high stakes applications of AI, such as self-driving cars and automated surgical assistants. This workshop will bring together researchers in areas such as human-robot interaction, security, causal inference, and multi-agent systems in order to strengthen the field of reliability engineering for machine learning systems. We are interested in approaches that have the potential to provide assurances of reliability, especially as systems scale in autonomy and complexity. We will focus on four aspects — robustness (to adversaries, distributional shift, model mis-specification, corrupted data); awareness (of when a change has occurred, when the model might be mis-calibrated, etc.); adaptation (to new situations or objectives); and monitoring (allowing humans to meaningfully track the state of the system). Together, these will aid us in designing and deploying reliable machine learning systems.

Schedule

08:40 AM	**Opening Remarks**	*Steinhardt*
09:00 AM	**Rules for Reliable Machine Learning**	*Zinkevich*
09:30 AM	**What's your ML Test Score? A rubric for ML production systems**	*Sculley*
09:45 AM	**Poster Spotlights I**	
10:30 AM	**Robust Learning and Inference**	*Mansour*
11:00 AM	**Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition**	*Hill*
11:30 AM	**Robust Covariate Shift Classification Using Multiple Feature Views**	*Liu*
11:45 AM	**Poster Spotlights II**	
01:15 PM	**Doug Tygar**	*Tygar*
01:45 PM	**Adversarial Examples and Adversarial Training**	*Goodfellow*
02:15 PM	**Summoning Demons: The Pursuit of Exploitable Bugs in Machine Learning**	*Suciu*
02:30 PM	**Poster Spotlights III**	
02:45 PM	**Poster Session**	
03:30 PM	**Learning Reliable Objectives**	*Dragan*
04:00 PM	**Building and Validating the AI behind the Next-Generation Aircraft Collision Avoidance System**	*Kochenderfer*

04:30 PM	**Online Prediction with Selfish Experts**	*Schrijvers*
04:45 PM	**TensorFlow Debugger: Debugging Dataflow Graphs for Machine Learning**	*Sculley*
05:00 PM	**What are the challenges to making machine learning reliable in practice?**	

Abstracts (1):

Abstract 5: **Robust Learning and Inference in Reliable Machine Learning in the Wild**, *Mansour* 10:30 AM

Robust inference is an extension of probabilistic inference,
where some of the observations may be adversarially corrupted.
We limit the adversarial corruption to a finite set of modification rules.
We model robust inference as a zero-sum game between an adversary, who selects a modification rule, and a predictor, who wants to accurately predict the state of nature.

There are two variants of the model, one where the adversary needs to pick
the modification rule in advance and one where the adversary can select the modification rule after observing the realized uncorrupted input.
For both settings we derive efficient near optimal policy runs in polynomial time.
Our efficient algorithms are based on methodologies for developing local computation algorithms.

We also consider a learning setting where the predictor receives a set of uncorrupted inputs and their classification.
The predictor needs to select a hypothesis, from a known set of hypotheses, and is tested on inputs which the adversary corrupts. We show how to utilize an ERM oracle to derive a near optimal predictor strategy, namely, picking a hypothesis that minimizes the error on the corrupted test inputs.

Based on joint works with
Uriel Feige, Aviad Rubinstein, Robert Schapira, Moshe Tennenholtz, Shai Vardi.

Representation Learning in Artificial and Biological Neural Networks

Leila Wehbe, Anwar O Nunez-Elizalde, Marcel Van Gerven, Moritz Grosse-Wentrup, Irina Rish, Brian Murphy, Georg Langs, Guillermo Cecchi

Room 114, Fri Dec 09, 08:00 AM

This workshop explores the interface between cognitive neuroscience and recent advances in AI fields that aim to reproduce human performance such as natural language processing and computer vision, and specifically deep learning approaches to such problems.

When studying the cognitive capabilities of the brain, scientists follow a system identification approach in which they present different stimuli to the subjects and try to model the response that different brain areas have of that stimulus. The goal is to understand the brain by trying to find the function that expresses the activity of brain areas in terms of different properties of the stimulus. Experimental stimuli are becoming increasingly complex with more and more people being interested in studying real life phenomena such as the perception of natural images or natural sentences. There is therefore a need for a rich and adequate vector representation of the properties of the stimulus, that we can obtain using advances in NLP, computer vision or other relevant ML disciplines.

In parallel, new ML approaches, many of which in deep learning, are inspired to a certain extent by human behavior or biological principles. Neural networks for example were originally inspired by biological neurons. More recently, processes such as attention are being used which have are inspired by human behavior. However, the large bulk of these methods are independent of findings about brain function, and it is unclear whether it is at all beneficial for machine learning to try to emulate brain function in order to achieve the same tasks that the brain achieves.

In order to shed some light on this difficult but exciting question, we bring together many experts from these converging fields to discuss these questions, in a new highly interactive format focused on short lectures from experts in both fields, followed by a guided discussion.

This workshop is a continuation of a successful workshop series: Machine Learning and Interpretation in Neuroimaging (MLINI). MLINI has already had 5 iterations in which methods for analyzing and interpreting neuroimaging data were discussed in depth. In keeping with previous tradition in the workshop, we also visit the blossoming field of machine learning applied to neuroimaging data, and specifically the recent trend of utilizing neural network models to analyze brain data, which is evolving on a seemingly orthogonal plane to the use of these algorithms to represent the information content in the brain. This way we will complete the loop of studying the advances of neural networks in neuroscience both as a source of models for brain representations, and as a tool for brain image analysis.

Schedule

08:30 AM	**Introductory remarks**	
08:45 AM	**Jessica Thompson - How can deep learning advance computational modeling of sensory information processing?**	*Thompson*
09:00 AM	**Matthias Bethge - How much understanding have we gained about deep CNN features for computer vision since AlexNet?**	*Bethge*
09:30 AM	**Sven Eberhardt - More Feedback, Less Depth: Approximating Human Vision with Deep Networks.**	*Eberhardt*
10:00 AM	**Panel discussion I**	
10:30 AM	**Coffee Break I**	

11:00 AM	**Rajesh Rao - Modeling human decision making using POMDPs**	*Rao*
11:30 AM	**Tal Yarkoni - What does it mean to 'understand' what a neural network is doing?**	*Yarkoni*
12:00 PM	**Panel discussion II**	
12:30 PM	**Lunch Break**	
02:00 PM	**Spotlight Talks**	
03:00 PM	**Coffee Break II**	
03:30 PM	**Poster Session**	
04:30 PM	**Richard Socher - Tackling the Limits of Deep Learning for NLP**	*Socher*
05:00 PM	**Alex Huth - Using Natural Language for Studying the Human Cortex**	*Huth*
05:30 PM	**Panel discussion III**	
06:00 PM	**General Discussion**	

Abstracts (4):

Abstract 2: **Jessica Thompson - How can deep learning advance computational modeling of sensory information processing? in Representation Learning in Artificial and Biological Neural Networks**, *Thompson* 08:45 AM

Deep learning, computational neuroscience, and cognitive science have overlapping goals related to understanding intelligence such that perception and behaviour can be simulated in computational sys- tems. In neuroimaging, machine learning methods have been used to test computational models of sensory information processing. Recently, these model comparison techniques have been used to evaluate deep neural networks (DNNs) as models of sensory information processing. However, the interpretation of such model evaluations is muddied by imprecise statistical conclusions. Here, we make explicit the types of conclusions that can be drawn from these existing model comparison techniques and how these conclusions change when the model in question is a DNN. We discuss how DNNs are amenable to new model comparison techniques that allow for stronger conclusions to be made about the computational mechanisms underlying sensory information processing.

Abstract 4: **Sven Eberhardt - More Feedback, Less Depth: Approximating Human Vision with Deep Networks. in Representation Learning in Artificial and Biological Neural Networks**, *Eberhardt* 09:30 AM

Recent advances in Deep Convolutional Networks (DCNs) supporting increasingly deep architectures have demonstrated significant gains in object recognition accuracy when trained on large labeled image databases. While a growing body of work indicates this surge in DCN performance carries concomitant improvement in fitting both neural data in higher areas of the primate visual cortex and human psychophysical data during object recognition, key differences remain. To investigate these differences, we assess the correlation between computational models and human behavioral responses on a rapid animal vs. non-animal categorization task. We find that DCN recognition accuracy increases with higher stages of visual processing (higher level stages indeed outperforming human participants on the same task) but that human decisions agree best with predictions from intermediate stages. These results suggest that while DCNs properly model visual features of intermediate complexity as used by the human visual system, more advanced visual processing relies on mechanisms not captured by these models. What kind of features do humans and DCNs base object decisions off of? To test this, we introduce a competitive web-based game for discovering features that humans use for object recognition: One participant from a pair sequentially reveals parts of an object in an image until the other correctly identifies its category. Scoring image regions according to their proximity to correct recognition yields maps of visual feature importance for individual images. We find that these ``realization'' maps exhibit only weak correlation with relevance maps derived from DCNs or image salience algorithms. Cueing DCNs to attend to features emphasized by these maps improves their object recognition accuracy. Our results thus suggest that realization maps identify visual features that humans deem important for object recognition but are not adequately captured by DCNs. Finally, we suggest a novel DCN training approach in which we base our representation on object and surface structure, rather than picture class labels, to build a more human-like visual representation.

Abstract 8: **Tal Yarkoni - What does it mean to 'understand' what a neural network is doing? in Representation Learning in Artificial and Biological Neural Networks**, *Yarkoni* 11:30 AM

In recent years, researchers have drawn strong parallels between the information-processing architectures and learned representations found in the human brain and in deep neural networks (DNNs). There is increasing interest in trying to use insights gained from either neuroscience or deep learning to reciprocally inform work in the other field. A common claim by practitioners in both fields is that we still do not understand very much about the representations learned by neural networks--whether biological or artificial. In this talk, I argue that this "mysterian" view is both surprising and troubling. It is surprising in that it is often expressed by people who demonstrably do understand an enormous amount about the systems they are studying. And it is troubling in that, if the claim is taken to be true, it does not lend itself to optimism about our future ability to understand what exactly neural networks are learning. I argue that the most productive avenues of research in both neuroscience and deep learning may be those that largely sidestep questions about information content and focus instead on architectural and algorithmic considerations.

Abstract 14: **Richard Socher - Tackling the Limits of Deep Learning for NLP in Representation Learning in Artificial and Biological Neural Networks**, *Socher* 04:30 PM

Deep learning has made great progress in a variety of language and vision tasks.
However, there are still many practical and theoretical problems and limitations.
In this talk I will introduce solutions to the following questions:
How to have a single input and output encoding for words.
How to predict previously unseen words during test time encounters.
How to grow a single deep learning model for many increasingly complex language tasks.

Can an end-to-end trainable architecture solve both visual and textual question answering?

3D Deep Learning

Fisher Yu, Joseph J Lim, Matt D Fisher, Qixing Huang, Jianxiong Xiao

Room 115, Fri Dec 09, 08:00 AM

Deep learning is proven to be a powerful tool to build models for language (one-dimensional) and image (two-dimensional) understanding. Tremendous efforts have been devoted into these areas, however, it is still at the early stage to apply deep learning to 3D data, despite their great research values and broad real-world applications. In particular, existing methods poorly serve the three-dimensional data that drives a broad range of critical applications such as augmented reality, autonomous driving, graphics, robotics, medical imaging, neuroscience, and scientific simulations. These problems have drawn attention of researchers in different fields such as neuroscience, computer vision and graphics.

Different from text or images that can be naturally represented as 1D or 2D arrays, 3D data have multiple representation candidates, such as volumes, polygonal meshes, multi-views renderings, depth maps, and point clouds. Coupled with these representations are the myriad 3D learning problems, such as object recognition, scene layout estimation, compositional structure parsing, novel view synthesis, model completion and hallucination, etc. 3D data opens new and vast research space, which naturally calls for interdisciplinary expertise ranging from Computer Vision, Computer Graphics, to Machine Learning.

The goal of this workshop is to foster interdisciplinary communication of researchers working on 3D data (Computer Vision and Computer Graphics), so that more attention of broader community can be drawn to 3D deep learning problems. Through those studies, new ideas and discoveries are expected to emerge, which can inspire advances in related fields.

This workshop is composed of invited talks, oral presentations of outstanding submissions and a poster session to showcase the state-of-the-art results in the topic. In particular, a panel discussion among leading researchers in the field is planned, so as to provide a common playground for inspiring discussions and stimulating debates.

We aim to build a venue for publishing original research results in 3D deep learning, as well as exhibiting the latest trends and ideas. To be specific, we are interested in the following topics using 3D deep learning methods:

3D object detection from depth images and videos
3D scene understanding
3D spatial understanding from 2D images
3D shape classification and segmentation
3D mapping and reconstruction
Learning 3D geometrical properties and representations
Analysis of 3D medical and biological imaging data

We accept two tracks of submissions to the workshop on those topics: paper (6 - 9 pages) and extended abstract (4 pages). We are inviting researchers of related fields to join the workshop program committee to review the submissions. All the submissions will follow NIPS main conference paper style. The paper will be reviewed in double-blind form from three researchers in the workshop program committee. High quality papers will be selected for oral presentation. The abstracts will be reviewed by the workshop committee in single-blind fashion. Accepted submissions will either be presented as posters or talks at the workshop. We encourage submissions of works that has been previously published or is to be presented in the main conference.

Schedule

08:30 AM	**Oral Presentation**	
08:45 AM	**Learning 3D representations, disparity estimation, and structure from motion**	*Brox*
10:30 AM	**Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling**	
11:00 AM	**FusionNet: 3D Object Classification Using Multiple Data Representations**	
02:30 PM	**Invited Talk by Abhinav Gupta**	
03:00 PM	**Invited Talk by Michael Bronstein**	
04:00 PM	**Invited Talk**	*Funkhouser*
04:30 PM	**Generative and Discriminative Voxel Modeling with Convolutional Neural Networks**	
05:00 PM	**Sparse 3D Convolutional Neural Networks for Large-Scale Shape Retrieval**	

Machine Learning for Health

Uri Shalit, Marzyeh Ghassemi, Jason Fries, Rajesh Ranganath, Theofanis Karaletsos, David Kale, Peter Schulam, Madalina Fiterau

Room 116, Fri Dec 09, 08:00 AM

The last decade has seen unprecedented growth in the availability and size of digital health data, including electronic health records, genetics, and wearable sensors. These rich data sources present opportunities to develop and apply machine learning methods to enable precision medicine. The aim of this workshop is to engender discussion between machine learning and clinical researchers about how statistical learning can enhance both the science and the practice of medicine.

Of particular interest to this year's workshop is a phrase recently coined by the British Medical Journal, "Big Health Data", where the focus is on modeling and improving health outcomes across large numbers of patients with diverse genetic, phenotypic, and environmental characteristics. The majority of clinical informatics research has focused on narrow populations representing, for example, patients from a single institution or sharing a common disease, and on modeling clinical factors, such as lab test results and treatments. Big health considers large and diverse cohorts, often reaching over 100 million patients in size, as well as environmental factors that are known to impact health outcomes, including socioeconomic status, health care delivery and utilization, and pollution. Big Health Data problems pose a variety of challenges for standard statistical learning, many of them nontraditional. Including a patient's race and income in statistical analysis, for example, evokes concerns about patient privacy. Novel approaches to differential privacy may help alleviate such concerns. Other examples include modeling biased measurements and non-random missingness and causal inference in the presence of latent confounders.

In this workshop we will bring together clinicians, health data experts, and machine learning researchers working on healthcare solutions. The goal is to have a discussion to understand clinical needs and the technical challenges resulting from those needs including the development of interpretable techniques which can adapt to noisy, dynamic environments and the handling of biases inherent in the data due to being generated during routine care.

Part of our workshop includes a clinician pitch, a five-minute presentation of open clinical problems that need data-driven solutions. These presentations will be followed by a discussion between invited clinicians and attending ML researchers to understand how machine learning can play a role in solving the problem presented. Finally, the pitch plays a secondary role of enabling new collaborations between machine learning researchers and clinicians: an important step for machine learning to have a meaningful role in healthcare. A general call for clinician pitches will be disseminated to clinical researchers and major physician organizations, including clinician social networks such as Doximity.

We will invite submission of two page abstracts (not including references) for poster contributions and short oral presentations describing innovative machine learning research on relevant clinical problems and data. Topics of interest include but are not limited to models for diseases and clinical data, temporal models, Markov decision processes for clinical decision support, multiscale data-integration, modeling with missing or biased data, learning with non-stationary data, uncertainty and uncertainty propagation, non i.i.d. structure in the data, critique of models, causality, model biases, transfer learning, and incorporation of non-clinical (e.g., socioeconomic) factors.

We are seeking sponsorship to help cover the travel and registration costs for students that are
presenting posters or short contributed talks, and for clinicians participating as speakers or presenting problem pitches. Workshop organizers have already discussed sponsorship with
the NSF, and also plan to approach industry leaders.

Schedule

08:15 AM	**Introduction**	
08:25 AM	**Opening Keynote by Leo Anthony Celi: Data-Driven Healthcare**	*Celi*
09:10 AM	**Eric Xing**	*Xing*
09:40 AM	**Contributed spotlights I**	
10:30 AM	**Coffee break and poster session**	
11:00 AM	**Award session I: a word from the sponsors, followed by student talks**	
11:30 AM	**Clinician pitches & discussion I**	
01:45 PM	**Keynote by Neil Lawrence: Challenges for Delivering Machine Learning in Health**	
02:30 PM	**Award session II: a word from the sponsors, followed by student talks**	
03:00 PM	**Coffee Break**	
03:30 PM	**Niels Peek: Opportunities and Challenges of Learning Health Systems**	
04:00 PM	**Sendhil Mullainathan: Misuses of Machine Learning in Health Policy**	
04:30 PM	**Clinician pitches & discussion II**	
05:00 PM	**Poster session**	
05:30 PM	**Jenna Wiens**	

Abstracts (4):

Abstract 2: **Opening Keynote by Leo Anthony Celi: Data-Driven Healthcare in Machine Learning for Health**, *Celi* 08:25 AM

The widespread adoption of electronic medical records has created new opportunities for clinical investigation using big data techniques. The potential for nuanced investigation across a full range of clinical questions is tremendous, contingent on the investment hospitals and health systems can make in big data infrastructure. Secondary analysis of electronic health records will enable the use of real patient data to assist clinical decision-making, with the goal of eventually providing near-real time support for bedside encounters. Clinicians and patients will derive value from data-driven decision making, while hospitals and health systems may see returns in quality, patient safety, and satisfaction. For big data analytics to achieve their potential in clinical medicine, issues of data structure, analytics staffing, funding, and data security will have to be addressed, but the future is bright and fertile for the application of big data to medical care.

Abstract 8: **Keynote by Neil Lawrence: Challenges for Delivering Machine Learning in Health in Machine Learning for Health**, 01:45 PM

The wealth of data availability presents new opportunities in health but also challenges. In this talk we will focus on challenges for machine learning in health: 1. Paradoxes of the Data Society, 2. Quantifying the Value of Data, 3. Privacy, loss of control, marginalization.

Each of these challenges has particular implications for machine learning. The paradoxes relate to our evolving relationship with data and our changing expectations. Quantifying value is vital for accounting for the influence of data in our new digital economies and issues of privacy and loss of control are fundamental to how our pre-existing rights evolve as the digital world encroaches more closely on the physical.

One of the goals of research community should be to provide the technological tooling to address these challenges ensure that we are empowered to avoid the pitfalls of the data driven society, allowing us to reap the benefits of machine learning in applications from personalized health to health in the developing world.

Abstract 11: **Niels Peek: Opportunities and Challenges of Learning Health Systems in Machine Learning for Health**, 03:30 PM

Health systems worldwide are under pressure to deliver better care for more people from fewer resources. The global economic crisis has shrunk the resources available for healthcare but the growth in demand for care services continues unabated. "Learning Health Systems" is a novel health informatics paradigm that blends quality improvement methods with data science. The goal is to create an integrated health system which harnesses routinely-collected health data to learn from every patient, and feed the knowledge of “what works best” back to clinicians, public health professionals, patients, and other stakeholders to create cycles of continuous improvement. In this talk we dissect the new paradigm and explore its opportunities and challenges for data scientists.

Abstract 12: **Sendhil Mullainathan: Misuses of Machine Learning in Health Policy in Machine Learning for Health**, 04:00 PM

We highlight some common (and costly) reasons for misuse of machine learning in health, illustrated using the potential outcomes framework from econometric work on causal inference. First, the failure to specify the decision which will be influenced by the prediction: the same prediction can lead to valid inferences for certain decisions but highly suspect ones for other decisions. Second, the selective labels problem: the data used to form the prediction is endogenously generated. Third, the conflation of averages with margins. We illustrate these points with two predictors that are commonly misused: readmissions and mortality. We argue that on the one hand, ignoring these problems can lead to highly misleading applications; on the other hand, judicious choice of applications and methods can allow one to circumvent these problems.

Time Series Workshop

Oren Anava, Marco Cuturi, Azadeh Khaleghi, Vitaly Kuznetsov, Sasha Rakhlin

Room 117, Fri Dec 09, 08:00 AM

Data, in the form of time-dependent sequential observations emerge in many key real-world problems, ranging from biological data, financial markets, weather forecasting to audio/video processing. However, despite the ubiquity of such data, most mainstream machine learning algorithms have been primarily developed for settings in which sample points are drawn i.i.d. from some (usually unknown) fixed distribution. While there exist algorithms designed to handle non-i.i.d. data, these typically assume specific parametric form for the data-generating distribution. Such assumptions may undermine the complex nature of modern data which can possess long-range dependency patterns, and for which we now have the computing power to discern. On the other extreme lie on-line learning algorithms that consider a more general framework without any distributional assumptions. However, by being purely-agnostic, common on-line algorithms may not fully exploit the stochastic aspect of time-series data.

Our workshop will build on the success of the first NIPS Time Series Workshop that was held at NIPS 2015. The goal of this workshop is to bring together theoretical and applied researchers interested in the analysis of time series and development of new algorithms to process sequential data. This includes algorithms for time series prediction, classification, clustering, anomaly and change point detection, correlation discovery, dimensionality reduction as well as a general theory for learning and comparing stochastic processes. We invite researchers from the related areas of batch and online learning, reinforcement learning, data analysis and statistics, econometrics, and many others to contribute to this workshop.

We also hope that this workshop will serve as an excellent companion to a tutorial on "Theory and Algorithms for Forecasting Non-Stationary Time Series" which is going to be presented at NIPS this year.

This year selected proceedings will be published in the JMLR special issue on "Time Series Analysis".

Schedule

08:45 AM	**Opening remarks: Azadeh Khaleghi**
09:05 AM	**Mehryar Mohri**
09:50 AM	**Yan Liu**
10:35 AM	**Morning coffee break**
11:00 AM	**Panel Discussion**
11:45 AM	**Poster session**
12:30 PM	**Lunch Break (and poster viewing)**
02:30 PM	**Contributed talk #1**
02:45 PM	**Andrew Nobel**
03:30 PM	**Afternoon cofee break**
04:00 PM	**Contributed talk #2**
04:15 PM	**Inderjit Dhillon**
05:00 PM	**Contributed talk #3**
05:15 PM	**Stephen Roberts**
06:00 PM	**Contributed talk #4**
06:15 PM	**Closing remarks**

Crowdsourcing and Machine Learning

Adish Singla, Matteo Venanzi, Rafael Frongillo

Room 120 + 121, Fri Dec 09, 08:00 AM

Building systems that seamlessly integrate machine learning (ML) and human intelligence can greatly push the frontier of our ability to solve challenging real-world problems. While ML research usually focuses on developing more efficient learning algorithms, it is often the quality and amount of training data that predominantly govern the performance of real-world systems. This is only amplified by the recent popularity of large scale and complex learning methodologies such as Deep Learning, which can require millions to billions of training instances to perform well. The recent rise of human computation and crowdsourcing approaches, made popular by task-solving platforms like Amazon Mechanical Turk and CrowdFlower, enable us to systematically collect and organize human intelligence. Crowdsourcing research itself is interdisciplinary, combining economics, game theory, cognitive science, and human-computer interaction, to create robust and effective mechanisms and tools. The goal of this workshop is to bring crowdsourcing and ML experts together to explore how crowdsourcing can contribute to ML and vice versa. Specifically, we will focus on the design of mechanisms for data collection and ML competitions, and conversely, applications of ML to complex crowdsourcing platforms.

CROWDSOURCING FOR DATA COLLECTION

Crowdsourcing is one of the most popular approaches to data collection for ML, and therefore one of the biggest avenues through which crowdsourcing can advance the state of the art in ML. We seek cost-efficient and fast data collection methods based on crowdsourcing, and ask how design decisions in these methods could impact subsequent stages of ML system. Topics of interest include:
- Basic annotation: What is the best way to collect and aggregate labels for unlabeled data from the crowd? How can we increase fidelity by flagging labels as uncertain given the crowd feedback? How can we do the above in the most cost-efficient manner?
- Beyond simple annotation tasks: What is the most effective way to collect probabilistic data from the crowd? How can we collect data requiring global knowledge of the domain such as building Bayes net structure via crowdsourcing?
- Time-sensitive and complex tasks: How can we design crowdsourcing systems to handle real-time or time-sensitive tasks, or those requiring more complicated work dependencies? Can we encourage collaboration on complex tasks?
- Data collection for specific domains: How can ML researchers apply the crowdsourcing principles to specific domains (e.g., healthcare) where privacy and other concerns are at play?

ML RESEARCH VIA COMPETITIONS

Through the Netflix challenge and now platforms like Kaggle, we are seeing the crowdsourcing of ML research itself. Yet the mechanisms underlying these competitions are extremely simple. Here our focus is on the design of such competitions; topics of interest include:
- What is the most effective way to incentivize the crowd to participate in the ML competitions? What is the most efficient method; rather than the typically winner-takes-all, can we design a mechanism which makes better use of the net research-hours devoted to the competition?
- Competitions as recruiting: how would we design a competition differently if (as is often the case) the result is not a winning algorithm but instead a job offer?
- Privacy issues with data sharing are one of the key barriers to holding such competitions. How can we design privacy-aware mechanisms which allow enough access to enable a meaningful competition?
- Challenges arising from the sequential and interactive nature of competitions, e.g., how can we maintain unbiased leaderboards without allowing for overfitting?

ML FOR CROWDSOURCING SYSTEMS

General crowdsourcing systems such as Duolingo, FoldIt, and Galaxy Zoo confront challenges of reliability, efficiency, and scalability, for which ML can provide powerful solutions. Many ML approaches have already been applied to output aggregation, quality control, work flow management and incentive design, but there is much more that could be done, either through novel ML methods, major redesigns of workflow or mechanisms, or on new crowdsourcing problems. Topics here include:
- Dealing with sparse, noisy and large number of label classes, for example, in tagging image collection for Deep Learning based computer vision algorithms.
- Optimal budget allocation and active learning in crowdsourcing.
- Open theoretical questions in crowdsourcing that can be addressed by statistics and learning theory, for instance, analyzing label aggregation algorithms such as EM, or budget allocation strategies.
- Applications of ML to emerging crowd-powered marketplaces (e.g., Uber, AirBnb). How can ML improve the efficiency of these markets?

Schedule

08:30 AM	**Poster Setup by Authors**	
09:00 AM	**Opening Remarks**	
09:05 AM	**Jennifer Wortman Vaughan: "The Communication Network Within the Crowd"**	*Wortman Vaughan*
09:55 AM	**Edoardo Manino: "Efficiency of Active Learning for the Allocation of Workers on Crowdsourced Classification Tasks"**	
10:05 AM	**Yao-Xiang Ding: "Crowdsourcing with Unsure Option"**	
10:15 AM	**Yang Liu: "Doubly Active Learning: When Active Learning meets Active Crowdsourcing"**	
10:30 AM	**Coffee + Posters**	
11:00 AM	**Sewoong Oh: "The Minimax Rate for Adaptive Crowdsourcing"**	*Oh*
11:45 AM	**Matteo Venanzi: "Time-Sensitive Bayesian Information Aggregation for Crowdsourcing Systems"**	

11:55 AM	**Miles E. Lopes: "A Sharp Bound on the Computation-Accuracy Tradeoff for Majority Voting Ensembles"**	
12:10 PM	**Ashish Kapoor: "Identifying and Accounting for Task-Dependent Bias in Crowdsourcing"**	
12:30 PM	**Lunch**	
02:00 PM	**Boi Faltings: "Incentives for Effort in Crowdsourcing Using the Peer Truth Serum"**	
02:15 PM	**David Parkes: "Peer Prediction with Heterogeneous Tasks"**	
02:30 PM	**Jens Witkowski: "Proper Proxy Scoring Rules"**	
02:45 PM	**Jordan Suchow: "Rethinking Experiment Design as Algorithm Design"**	
03:00 PM	**Afternoon Coffee + Posters**	
03:30 PM	**Ben Hamner (Kaggle): "Kaggle Competitions and The Future of Reproducible Machine Learning"**	*Hamner*
04:30 PM	**Poster Presentation by Authors**	

Abstracts (3):

Abstract 3: **Jennifer Wortman Vaughan: "The Communication Network Within the Crowd" in Crowdsourcing and Machine Learning**, *Wortman Vaughan* 09:05 AM

Since its inception, crowdsourcing has been considered a black-box approach to solicit labor from a crowd of workers. Furthermore, the "crowd" has been viewed as a group of independent workers. Recent studies based on in-person interviews have opened up the black box and shown that the crowd is not a collection of independent workers, but instead that workers communicate and collaborate with each other. In this talk, I will describe our attempt to quantify this discovery by mapping the entire communication network of workers on Amazon Mechanical Turk, a leading crowdsourcing platform. We executed a task in which over 10,000 workers from across the globe self-reported their communication links to other workers, thereby mapping the communication network among workers. Our results suggest that while a large percentage of workers indeed appear to be independent, there is a rich network topology over the rest of the population. That is, there is a substantial communication network within the crowd. We further examined how online forum usage relates to network topology, how workers communicate with each other via this network, how workers' experience levels relate to their network positions, and how U.S. workers differ from international workers in their network characteristics. These findings have implications for requesters, workers, and platform providers. This talk is based on joint work with Ming Yin, Mary Gray, and Sid Suri.

Abstract 8: **Sewoong Oh: "The Minimax Rate for Adaptive Crowdsourcing" in Crowdsourcing and Machine Learning**, *Oh* 11:00 AM

Adaptive schemes, where tasks are assigned based on the data collected thus far, are widely used in practical crowdsourcing systems to efficiently allocate the budget. However, existing theoretical analyses of crowdsourcing systems suggest that the gain of adaptive task assignments is minimal. To bridge this gap, we propose a new model for representing practical crowdsourcing systems, which strictly generalizes the popular Dawid-Skene model, and characterize the fundamental trade-off between budget and accuracy. We introduce a novel adaptive scheme that matches this fundamental limit. We introduce new techniques to analyze the spectral analyses of non-back-tracking operators, using density evolution techniques from coding theory.

Abstract 18: **Ben Hamner (Kaggle): "Kaggle Competitions and The Future of Reproducible Machine Learning" in Crowdsourcing and Machine Learning**, *Hamner* 03:30 PM

At Kaggle, we've run hundreds of machine learning competitions and seen over 150,000 data scientists make submissions. One thing is clear: winning competitions isn't random. We've learned that certain tools and methodologies work consistently well on different types of problems. Many participants make common mistakes (such as overfitting) that should be actively avoided. Similarly, competition hosts have their own set of pitfalls (such as data leakage). In this talk, I'll share what goes into a winning competition toolkit along with some war stories on what to avoid. Additionally, I'll share what we're seeing on the collaborative side of competitions. Our community is showing an increasing amount of collaboration in developing machine learning models and analytic solutions. As collaboration has grown, we've seen reproducibility as a key pain point in machine learning. It can be incredibly tough to rerun and build on your colleague's work, public work, or even your own past work! We're expanding our focus to build a reproducible data science platform that hits directly at these pain points. It combines versioned data, versioned code, and versioned computational environments (through Docker containers) to create reproducible results.

Adaptive Data Analysis

Vitaly Feldman, Aaditya Ramdas, Aaron Roth, Adam Smith

Room 122 + 123, Fri Dec 09, 08:00 AM

Adaptive data analysis is the increasingly common practice by which insights gathered from data are used to inform further analysis of the same data sets. This is common practice both in machine learning, and in scientific research, in which data-sets are shared and re-used across multiple studies. Unfortunately, most of the statistical inference theory used in empirical sciences to control false discovery rates, and in machine learning to avoid overfitting, assumes a fixed class of hypotheses to test, or family of functions to optimize over, selected independently of the data. If the set of analyses run is itself a function of the data, much of this theory becomes invalid, and indeed, has been blamed as one of the causes of the crisis of reproducibility in empirical science.

Recently, there have been several exciting proposals for how to avoid overfitting and guarantee statistical validity even in general adaptive data analysis settings. The problem is important, and ripe for further advances. The goal of this workshop is to bring together members of different communities (from machine learning, statistics, and theoretical computer science) interested in solving this problem, to share recent results, to discuss promising directions for future research, and to foster collaborations.

Schedule

08:55 AM	**Introductory remarks**	
09:00 AM	**Ruth Heller. Inference following aggregate level hypothesis testing in large scale genomic data**	
09:35 AM	**Weijie Su. Private false discovery rate control and robustness of the Benjamini-Hochberg procedure**	
10:10 AM	**Vitaly Feldman**	*Feldman*
10:20 AM	**Coffee break**	
10:50 AM	**Short talks: Ibrahim Alabdulmohsin, Joshua Loftus, Yu-Xiang Wang, Sam Elder, Aaditya Ramdas, Ryan Rogers**	
12:00 PM	**Lunch break**	
02:30 PM	**Kobbi Nissim. Algorithmic Stability via Differential Privacy**	*nissim*
03:05 PM	**Katrina Ligett. Adaptive Learning with Robust Generalization Guarantees**	
03:50 PM	**Posters**	
04:35 PM	**Lucas Janson. Model-free knockoffs: statistical tools for reproducible selections**	
04:55 PM	**Xiaoying Harris. From Selective Inference to Adaptive Data Analysis**	
05:15 PM	**Peter Grunwald. Safe Testing: An Adaptive Alternative to p-value-based testing**	
05:50 PM	**Aaron Roth**	

Abstracts (8):

Abstract 2: **Ruth Heller. Inference following aggregate level hypothesis testing in large scale genomic data in Adaptive Data Analysis**, 09:00 AM

In many genomic applications, it is common to perform tests using aggregate-level statistics within naturally defined classes for powerful identification of signals. Following aggregate-level testing, it is naturally of interest to infer on the individual units that are within classes that contain signal. Failing to account for class selection will produce biased inference. We develop multiple testing procedures that allow rejection of individual level null hypotheses while controlling for conditional (familywise or false discovery) error rates. We use simulation studies to illustrate validity and power of the proposed procedures in comparison to several possible alternatives. We illustrate the usefulness of our procedures in a natural application involving whole-genome expression quantitative trait loci (eQTL) analysis across 17 tissue types using data from The Cancer Genome Atlas (TCGA) Project.

Joint work with Nilanjan Chatterjee, Abba Krieger, and Jianxin Shi.

Abstract 3: **Weijie Su. Private false discovery rate control and robustness of the Benjamini-Hochberg procedure in Adaptive Data Analysis**, 09:35 AM

We provide the first differentially private algorithms for controlling the false discovery rate (FDR) in multiple hypothesis testing. Our general approach is to adapt a well-known variant of the Benjamini-Hochberg procedure (BHq), making each step differentially private. This destroys the classical proof of FDR control. To prove FDR control of our method, we develop a new proof of the original (non-private) BHq algorithm and its robust variants -- a proof requiring only the assumption that the true null test statistics are independent, allowing for arbitrary correlations between the true nulls and false nulls. This assumption is fairly weak compared to those previously shown in the vast literature on this topic, and explains in part the empirical robustness of BHq.

Abstract 6: **Short talks: Ibrahim Alabdulmohsin, Joshua Loftus, Yu-Xiang Wang, Sam Elder, Aaditya Ramdas, Ryan Rogers in Adaptive Data Analysis**, 10:50 AM

10:50-11:00. Ibrahim Alabdulmohsin. On the Interplay between Information, Stability, and Generalization
11:00-11:10. Joshua Loftus. Significance testing after cross-validation
11:10-11:20. Yu-Xiang Wang*, Jing Lei and Stephen E. Fienberg. A Minimax Theory for Adaptive Data Analysis
11:20-11:30. Sam Elder. Bayesian Adaptive Data Analysis: Challenges and Guarantees
11:30-11:40. Rina Foygel Barber and Aaditya Ramdas*. p-filter: An internally consistent framework for FDR.
11:40-11:50. Ryan Rogers*, Aaron Roth, Adam Smith and Om Thakkar. Max-Information, Differential Privacy, and Post-Selection Hypothesis Testing

Abstract 8: **Kobbi Nissim. Algorithmic Stability via Differential Privacy in Adaptive Data Analysis**, *nissim* 02:30 PM

Adaptively is an important feature of data analysis - the choice of questions to ask about a dataset often depends on previous interactions with the same dataset. However, statistical validity is typically studied in models with limited adaptivity, such as where all questions are specified before the dataset is drawn. A recent line of work by Dwork et al. [STOC, 2015] and Hardt and Ullman [FOCS, 2014] initiated the formal study of this problem and related it to differential privacy [TCC 2006].

In this talk we will explore some of the connections between differential privacy and statistical validity. We will show that algorithms satisfying differential privacy imply low generalization error, and examine some of the implications of this result on private learning and statistical validity with adaptively chosen queries.

Joint work with Raef Bassily, Adam Smith, Thomas Steinke, Uri Stemmer, Jonathan Ullman.

Abstract 9: **Katrina Ligett. Adaptive Learning with Robust Generalization Guarantees in Adaptive Data Analysis**, 03:05 PM

The traditional notion of generalization --- i.e., learning a hypothesis whose empirical error is close to its true error --- is surprisingly brittle. As has recently been noted, even if several algorithms have this guarantee in isolation, the guarantee need not hold if the algorithms are composed adaptively. In this paper, we study three notions of generalization ---increasing in strength--- that are robust to post-processing and amenable to adaptive composition, and examine the relationships between them.

Abstract 11: **Lucas Janson. Model-free knockoffs: statistical tools for reproducible selections in Adaptive Data Analysis**, 04:35 PM

A common problem in modern statistical applications is to select, from a large set of candidates, a subset of variables which are important for determining an outcome of interest. For instance, the outcome may be disease status and the variables may be hundreds of thousands of single nucleotide polymorphisms on the genome. This talk introduces model-free knockoffs, a framework for finding dependent variables while provably controlling the false discovery rate (FDR) in finite samples. FDR control holds no matter the form of the dependence between the response and the covariates, which does not need to be specified in any way. What is required is that we observe i.i.d. samples (X,Y) and know something about the distribution of the covariates although we have shown that the method is robust to unknown/estimated covariate distributions. This framework builds on the knockoff filter of Foygel Barber and Candès introduced a couple of years ago, which was limited to linear models with fewer variables than observations (n ‹ p). In contrast, model-free knockoffs deal with a range of problems far beyond the scope of the original knockoff paper—e.g. it provides valid selections in any generalized linear model including logistic regression---while being more powerful than the original procedure when it applies. Finally, we apply our procedure to data from a case-control study of Crohn’s disease in the United Kingdom, making twice as many discoveries as the original analysis of the same data.

Abstract 12: **Xiaoying Harris. From Selective Inference to Adaptive Data Analysis in Adaptive Data Analysis**, 04:55 PM

Recent development in selective inference has provided a framework of valid inference after some information of the data has been used for model selection. However, most literature concerning selective inference require the practitioners to commit to a pre-specified procedure for model selection. This is rather stringent for applications. In many cases, multiple exploratory data analyses will be performed and the outcome of each will be input to the final model selected by the practitioners. Therefore, we want to develop a framework that allows multiple queries to the data. In a framework similar to that in differential privacy, we allow valid inference after multiple queries to the database. We seek to address this problem from the perspective of “multiple views of the data” and two concrete examples are considered below.

Joint work with Jonathan Taylor.

Abstract 13: **Peter Grunwald. Safe Testing: An Adaptive Alternative to p-value-based testing in Adaptive Data Analysis**, 05:15 PM

Standard p-value based hypothesis testing is not at all adaptive: if our test result is promising but not conclusive (say, p = 0.07) we cannot simply decide to gather a few more data points. While the latter practice is ubiquitous in science, it invalidates p-values and error guarantees.

Here we propose an alternative test based on supermartingales - it has both a gambling and a data compression interpretation. This method allows us to freely combine results from different tests by multiplication (which would be a mortal sin for p-values!), and avoids many other pitfalls of traditional testing as well. If the null hypothesis is simple (a singleton), it also has a Bayesian interpretation, and essentially coincides with a proposal by Vovk (1993) and Berger et al. (1994). Here we work out, for the first time, the case of composite null hypotheses, which allows us to formulate safe, nonasymptotic versions of the most popular tests such as the t-test and the chi square tests. Safe tests for composite H0 are not Bayesian, and initial experiments suggests that they can substantially outperform Bayesian tests (which for composite nulls are not adaptive in general).

Machine Learning for Intelligent Transportation Systems

Li Erran Li, Prof. Darrell

Room 124 + 125, Fri Dec 09, 08:00 AM

Our transportation systems are poised for a transformation as we make progress on autonomous vehicles, vehicle-to-vehicle (V2V) and vehicle-to-everything (V2X) communication infrastructures, and smart road infrastructures such as smart traffic lights. There are many challenges in transforming our current transportation systems to the future vision. For example, how do we achieve near-zero fatality? How do we optimize efficiency through intelligent traffic management and control of fleets? How do we optimize for traffic capacity during rush hours? To meet these requirements in safety, efficiency, control, and capacity, the systems must be automated with intelligent decision making.

Machine learning will be essential to enable intelligent transportation systems. Machine learning has made rapid progress in self-driving, e.g. real-time perception and prediction of traffic scenes, and has started to be applied to ride-sharing platforms such as Uber (e.g. demand forecasting) and crowd-sourced video scene analysis companies such as Nexar (understanding and avoiding accidents). To address the challenges arising in our future transportation system such as traffic management and safety, we need to consider the transportation systems as a whole rather than solving problems in isolation. New machine learning solutions are needed as transportation places specific requirements such as extremely low tolerance on uncertainty and the need to intelligently coordinate self-driving cars through V2V and V2X.

The goal of this workshop is to bring together researchers and practitioners from all areas of intelligent transportations systems to address core challenges with machine learning. These challenges include, but are not limited to: predictive modeling of risk and accidents through telematics, modeling, simulation and forecast of demand and mobility patterns in large scale urban transportation systems, machine learning approaches for control and coordination of traffic leveraging V2V and V2X infrastructures, efficient pedestrian detection, pedestrian intent detection, intelligent decision-making for self-driving cars, scene classification, real-time perception and prediction of traffic scenes, deep reinforcement learning from human drivers, uncertainty propagation in deep neural networks, efficient inference with deep neural networks.

The workshop will include invited speakers, panels, presentations of accepted papers and posters. We invite papers in the form of short, long and position papers to address the core challenges mentioned above. We encourage researchers and practitioners on self-driving cars, transportation systems and ride-sharing platforms to participate.

Schedule

08:30 AM	**Opening Remarks**	
08:45 AM	**Invited Talk: Safe Reinforcement Learning for Robotics (Pieter Abbeel, UC Berkeley and OpenAI)**	*Abbeel*
09:15 AM	**Invited Talk: Active Optimization and Autonomous Vehicles (Jeff Schneider, CMU and Uber ATC)**	
09:45 AM	**Contributed Talks (3 x 15 min)**	
10:30 AM	**Posters and Break**	
11:00 AM	**Invited Talk: Learning Affordance for Direct Perception in Autonomous Driving (JiaoXiong Xiao, AutoX)**	
11:30 AM	**Invited Talk: End to End Learning for Self-Driving Cars (Larry Jackel, NVIDIA)**	
12:00 PM	**Invited Talk: Towards Affordable Self-driving Cars (Raquel Urtasun, University of Toronto)**	*Urtasun*
12:30 PM	**Lunch**	
01:40 PM	**Invited Talk: Visual Understanding of Human Activities for Smart Vehicles and Interactive Environments (Juan Carlos Niebles, Stanford)**	*Niebles*
02:10 PM	**Invited Talk: Autonomous Cars that Coordinate with People (Anca Dragan, Berkeley)**	*Dragan*
02:40 PM	**Lightning Talks (6 x 2 min)**	
03:00 PM	**Posters and Coffe**	
03:30 PM	**Invited Talk: Scene Labeling and more – Deep Neural Nets for Autonomous Vehicles (Uwe Franke, Daimler AG)**	
04:00 PM	**Invited Talk: Efficient Deep Networks for Real-Time Classification in Embedded Platforms (Jose Alvarez, NICTA, Australia)**	
04:30 PM	**Invited Talk: Domain Adaption for Perception and Action (Kate Saenko, Boston University)**	*Saenko*
05:00 PM	**Invited Talk: Learning Adaptive Driving Models from Large-scale Video Datasets (Fisher Yu, Huazhe Xu, Dequan Wang, and Trevor Darrell, Berkeley)**	*Darrell*
05:30 PM	**Discussion**	
06:00 PM	**Closing Remarks**	

Abstracts (13):

Abstract 2: **Invited Talk: Safe Reinforcement Learning for Robotics (Pieter Abbeel, UC Berkeley and OpenAI) in Machine Learning for Intelligent Transportation Systems**, *Abbeel* 08:45 AM

Abstract: Recent advances in deep reinforcement learning have enabled a wide range of capabilities, including learning to play Atari games, learning (simulated) locomotion, and learning (real robot) visuomotor skills. A key issue in the application to real robotics, however, is safety during learning. In this talk I will discuss approaches to make learning safer through incorporation of classical model-predictive control into learning and through safe adaptive transfer of skills from simulated to real environments.

Bio: Pieter Abbeel (Associate Professor, UC Berkeley EECS) works in machine learning and robotics, in particular his research is on making robots learn from people (apprenticeship learning) and how to make robots learn through their own trial and error (reinforcement learning). His robots have learned: advanced helicopter aerobatics, knot-tying, basic assembly, and organizing laundry. He has won various awards, including best paper awards at ICML and ICRA, the Sloan Fellowship, the Air Force Office of Scientific Research Young Investigator Program (AFOSR-YIP) award, the Office of Naval Research Young Investigator Program (ONR-YIP) award, the DARPA Young Faculty Award (DARPA-YFA), the National Science Foundation Faculty Early Career Development Program Award (NSF-CAREER), the Presidential Early Career Award for Scientists and Engineers (PECASE), the CRA-E Undergraduate Research Faculty Mentoring Award, the MIT TR35, the IEEE Robotics and Automation Society (RAS) Early Career Award, and the Dick Volz Best U.S. Ph.D. Thesis in Robotics and

Automation Award.

Abstract 3: **Invited Talk: Active Optimization and Autonomous Vehicles (Jeff Schneider, CMU and Uber ATC) in Machine Learning for Intelligent Transportation Systems**, 09:15 AM

Abstract:
An important property of embedded learning systems is the ever-changing environment they create for all algorithms operating in the system. Optimizing the performance of those algorithms becomes a perpetual on-line activity rather than a one-off task. I will review some of these challenges in autonomous vehicles. I will discuss active optimization methods and their application in robotics and scientific applications, focusing on scaling up the dimensionality and managing multi-fidelity evaluations. I will finish with lessons learned and thoughts on future directions as these methods move into embedded systems.

Bio:
Dr. Jeff Schneider is the engineering lead for machine learning at Uber's Advanced Technologies Center. He is currently on leave from Carnegie Mellon University where he is a research professor in the school of computer science. He has 20 years experience developing, publishing, and applying machine learning algorithms in government, science, and industry. He has over 100 publications and regularly gives talks and tutorials on the subject.

Previously, Jeff was the co-founder and CEO of Schenley Park Research, a company dedicated to bringing machine learning to industry. Later, he developed a machine learning based CNS drug discovery system and commercialized it during two years as Psychogenics' Chief Informatics Officer. Through his research, commercial, and consulting efforts, he has worked with dozens of companies and government agencies around the world.

Abstract 4: **Contributed Talks (3 x 15 min) in Machine Learning for Intelligent Transportation Systems**, 09:45 AM

1. Speeding up Semantic Segmentation for Autonomous Driving (Michael Treml, José Arjona-Medina, Thomas Unterthiner, Rupesh Durgesh, Felix Friedmann, Peter Schuberth, Andreas Mayr, Martin Heusel, Markus Hofmarcher, Michael Widrich, Bernhard Nessler, Sepp Hochreiter)

2. Multi-Path Feedback Recurrent Neural Network for Scene Parsing (Xiaojie Jin, Yunpeng Chen, Zequn Jie, Jiashi Feng, Shuicheng Yan)

3. Increasing the Stability of CNNs using a Denoising Layer Regularized by Local Lipschitz Constant (Hamed H. Aghdam, Elnaz J. Heravi, Domenec Puig)

Abstract 6: **Invited Talk: Learning Affordance for Direct Perception in Autonomous Driving (JiaoXiong Xiao, AutoX) in Machine Learning for Intelligent Transportation Systems**, 11:00 AM

Abstract:
Today, there are two major paradigms for vision-based autonomous driving systems: mediated perception approaches that parse an entire scene to make a driving decision, and behavior reflex approaches that directly map an input image to a driving action by a regressor. In this paper, we propose a third paradigm: a direct perception based approach to estimate the affordance for driving. We propose to map an input image to a small number of key perception indicators that directly relate to the affordance of a road/traffic state for driving. Our representation provides a set of compact yet complete descriptions of the scene to enable a simple controller to drive autonomously. Falling in between the two extremes of mediated perception and behavior reflex, we argue that our direct perception representation provides the right level of abstraction. We evaluate our approach in a virtual racing game as well as real world driving and show that our model can work well to drive a car in a very diverse set of virtual and realistic environments.

Bio:
Jianxiong Xiao (a.k.a., Professor X) is the Founder and CEO of AutoX, Inc., a high-tech startup currently in stealth mode. Previously, he was an Assistant Professor in the Department of Computer Science at Princeton University and the founding director of the Princeton Computer Vision and Robotics Labs from 2013 to 2016. He received his Ph.D. from the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT) in 2013. Before that, he received a BEng. and MPhil. in Computer Science from the Hong Kong University of Science and Technology in 2009. His research focuses on bridging the gap between computer vision and robotics by building extremely robust and dependable computer vision systems for robot perception. In particular, he is a pioneer in the fields of 3D Deep Learning, Autonomous Driving, RGB-D Recognition and Mapping, Big Data, Large-scale Crowdsourcing, and Deep Learning for Robotics. His work has received the Best Student Paper Award at the European Conference on Computer Vision (ECCV) in 2012 and the Google Research Best Papers Award for 2012, and has appeared in the popular press. Jianxiong was awarded the Google U.S./Canada Fellowship in Computer Vision in 2012, the MIT CSW Best Research Award in 2011, and two Google Faculty Awards in 2014 and in 2015 respectively. He co-lead the MIT+Princeton joint team to participate in the Amazon Picking Challenge in 2016, and won the 3rd and 4th place worldwide. More information can be found at: http://www.jianxiongxiao.com.

Abstract 7: **Invited Talk: End to End Learning for Self-Driving Cars (Larry Jackel, NVIDIA) in Machine Learning for Intelligent Transportation Systems**, 11:30 AM

Abstract:

End-to-End Learning has been demonstrated for controlling steering on a drive-by-wire car. The key software component in this system is a Convolutional Neural Network (CNN) that takes as input the stream from a video camera mounted behind the vehicle windshield and then, as output, provides steering commands to the vehicle. The CNN runs on an NVIDIA Drive PX board. The system has successfully driven on divided highways, narrow two lane roads, and roads without lane markings. The CNN was trained using data gathered by capturing on-board video from vehicles driven by humans while simultaneously recording those vehicles steering commands.

Bio:

Larry Jackel is President of North-C Technologies, where he does professional consulting. From 2003-2007 he was a DARPA Program Manager in the IPTO and TTO offices. He conceived and managed programs in Universal Network-Based Document Storage and in Autonomous Ground Robot navigation and Locomotion. For most of his scientific career Jackel was a manager and researcher in Bell Labs and then AT&T Labs. He has created and managed research groups in Microscience and Microfabrication, in Machine Learning and Pattern Recognition, and in Carrier-Scale Telecom Services. Jackel holds a PhD in Experimental Physics from Cornell University with a thesis in superconducting electronics. He is a Fellow of the American Physical Society and the IEEE.

Abstract 8: **Invited Talk: Towards Affordable Self-driving Cars (Raquel Urtasun, University of Toronto) in Machine Learning for Intelligent Transportation Systems**, *Urtasun* 12:00 PM

Abstract:
The revolution of self-driving cars will happen in the near future. Most solutions rely on expensive 3D sensors such as LIDAR as well as hand-annotated maps. Unfortunately, this is neither cost effective nor scalable, as one needs to have a very detailed up-to-date map of the world. In this talk, I'll review our current efforts in the domain of autonomous driving. In particular, I'll present our work on stereo, optical flow, appearance-less localization, 3D object detection as well as creating HD maps from visual information alone. This results in a much more scalable and cost-effective solution to self-driving cars.

Bio:
Raquel Urtasun is an Associate Professor in the Department of Computer Science at the University of Toronto and a Canada Research Chair in Machine Learning and Computer Vision. Prior to this, she was an Assistant Professor at the Toyota Technological Institute at Chicago (TTIC), an academic computer science institute affiliated with the University of Chicago. She received her Ph.D. degree from the Computer Science department at Ecole Polytechnique Federal de Lausanne (EPFL) in 2006 and did her postdoc at MIT and UC Berkeley. Her research interests include machine learning, computer vision and robotics. Her recent work involves perception algorithms for self-driving cars, deep structured models and exploring problems at the intersection of vision and language. She is a recipient of an NVIDIA Pioneers of AI Award, a Ministry of Education and Innovation Early Researcher Award, two Google Faculty Research Awards, a Connaught New Researcher Award and a Best Paper Runner up Prize awarded at the Conference on Computer Vision and Pattern Recognition (CVPR). She is also Program Chair of CVPR 2018, an Editor of the International Journal in Computer Vision (IJCV) and has served as Area Chair of multiple machine learning and vision conferences (i.e., NIPS, UAI, ICML, ICLR, CVPR, ECCV, ICCV).

Abstract 10: **Invited Talk: Visual Understanding of Human Activities for Smart Vehicles and Interactive Environments (Juan Carlos Niebles, Stanford) in Machine Learning for Intelligent Transportation Systems**, *Niebles* 01:40 PM

Abstract:
Future robots, intelligent vehicles, and smart spaces will require a deep and accurate understanding of the environment under diverse and challenging conditions, advanced reasoning capabilities and robust planning under uncertainty. In particular, understanding behaviors of people, such as drivers, road occupants or users interacting with smart machines will be critical for safety, contextualized assistance and natural interfaces. In this talk, I will give an overview of our recent work in computer vision for visual understanding of human actions and behaviors. Our approach extends recent advances in deep learning to enable learning with weak supervision, detecting human actions in videos at scale and anticipating events before they occur.

Bio:
Juan Carlos Niebles received an Engineering degree in Electronics from Universidad del Norte (Colombia) in 2002, an M.Sc. degree in Electrical and Computer Engineering from University of Illinois at Urbana-Champaign in 2007, and a Ph.D. degree in Electrical Engineering from Princeton University in 2011. He is a Senior Research Scientist at the Stanford AI Lab and Associate Director of Research at the Stanford-Toyota Center for AI Research since 2015. He is also an Assistant Professor of Electrical and Electronic Engineering in Universidad del Norte (Colombia) since 2011. His research interests are in computer vision and machine learning, with a focus on visual recognition and understanding of human actions and activities, objects, scenes, and events. He is a recipient of a Google Faculty Research award (2015), the Microsoft Research Faculty Fellowship (2012), a Google Research award (2011) and a Fulbright Fellowship (2005).

Abstract 11: **Invited Talk: Autonomous Cars that Coordinate with People (Anca Dragan, Berkeley) in Machine Learning for Intelligent Transportation Systems**, *Dragan* 02:10 PM

Abstract:
Cars tend to treat people like obstacles whose motion needs to be anticipated, so that the car can best stay out of their way. This results in ultra-defensive cars that cannot coordinate with people, because they miss on a key aspect of coordination: it's not just the car interpreting and responding to the actions of people, people also interpret and respond to the car's actions. We introduce a mathematical formulation of interaction that accounts for this, and show how learning and optimal control can be leveraged to generate car behavior that results in natural coordination strategies, like the car negotiating a merge or inching forward at an intersection to test whether it can go.

Bio: Anca is an Assistant Professor in the EECS Department at UC Berkeley. Her goal is to enable robots to work with, around, and in support of people. She run the InterACT Lab, where they focus on algorithms for human-robot interaction -- algorithms that move beyond the robot's function in isolation, and generate robot behavior that also accounts for interaction and coordination with end-users. She works across different applications, from assistive robots, to manufacturing, to autonomous cars, and draw from optimal control, planning, estimation, learning, and cognitive science. She also helped fund and serve on the steering committee for the Berkeley AI Research (BAIR) Lab, and am a co-PI of the Center for Human-Compatible AI.

Abstract 12: **Lightning Talks (6 x 2 min) in Machine Learning for Intelligent Transportation Systems**, 02:40 PM

1. Similarity Mapping with Enhanced Siamese Network for Multi-Object Tracking
Minyoung Kim, Stefano Alletto, Luca Rigazio

2. End-to-End Deep Reinforcement Learning for Lane Keeping Assist
Ahmad El Sallab, Mohammed Abdou, Etienne Perot and Senthil Yogamani

3. Efficient decomposition method for the stochastic optimization of public transport schedules
Sofia Zaourar-Michel

4. Mapping Occupancy of Dynamic Environments using Big Data Gaussian Process Classification
Ransalu Senanayake, Simon O'Callaghan, Fabio Ramos

5. Nonnegative Matrix Factorisation of Bike Sharing System Temporal Network
Ronan Hamon, Pierre Borgnat, Cédric Févotte, Patrick Flandrin

6. Safe and optimal path planning in uncertain skies
Ashish Kapoor

Abstract 14: **Invited Talk: Scene Labeling and more – Deep Neural Nets for Autonomous Vehicles (Uwe Franke, Daimler AG) in Machine Learning for Intelligent Transportation Systems**, 03:30 PM

Abstract:
For about 80 years, people have been dreaming of cars that are able to drive by themselves. These days, this vision is starting to become reality. For the first time, cars found their way over a long distance in the DARPA Grand Challenge in 2005. Two years later, the famous DARPA Urban Challenge took place. In both events, all finalists based their systems on active sensors, and Google also started their impressive work with a high-end laser scanner accompanied by radars.

In 2013, we let a new S-class vehicle (a.k.a. Bertha) drive itself from Mannheim to Pforzheim, following the route that Bertha Benz took 125 years ago. Bertha's environment perception was based on close to production radars and (stereo) cameras. For the visual object recognition classical box-based classifiers based on HOG and SVM or shallow neural nets were used. The experiment showed that despite the fact that the used stereo system allows for fully autonomous emergency braking in today's Mercedes-Benz production cars, the state-of-the-art in computer vision around 2013 was not sufficient to deliver the deep understanding of the scene that we need for cars driving themselves safely in complex urban traffic. The advent of Deep Neural Networks and the fact that GPUs allow to run powerful nets like the GoogLeNet in real-time totally changed the situation. In our current vision system about 80% of all tasks are solved by DNNs or use information delivered by them. The talk sketches the most important building blocks of this system.

Since we do not believe in a purely box-based recognition system we use a Fully Convolutional Network as the core of our vision system. For training and benchmarking we have introduced the Cityscapes Dataset and benchmark suite, publicly available since early 2016. In September, we registered the 1000th download. Within only one year, the pixel level semantic segmentation performance raised up from 65% IoU to more than 77% (October 2016). The results of the semantic labeling stage are subsequently fused with the stereo based Stixel-World, a super-pixel representation of the depth image using small rectangular shaped regions. The result is a very compact representation of the traffic scene including geometry, motion and semantics. In addition, safety demands to watch out for unexpected small objects (down to a height of 5cm) on the street. We fuse the results of a specially trained FCN with a boosted stereo analysis to detect more than 80% of all targets at distances up to 100m at a false positive rate of 1/min only. If depth is not available from stereo or Lidar, it has to be derived from monocular images. We solve the depth-from-mono problem jointly with scene labeling and instance segmentation. It turns out that these sub-tasks support each other well, resulting in close to ground truth results. All schemes run in real-time on a standard GPU. Given the fact that many suppliers have efficient HW components for CNNs on their roadmap, this raises hope that we can use these powerful techniques in the near future in our cars, both for driver assistance and autonomous driving.

Bio:
Uwe Franke received the Ph.D. degree in electrical engineering from the Technical University of Aachen, Germany, in 1988 for his work on content based image coding.

Since 1989 he has been with Daimler Research and Development and has been constantly working on the development of vision based driver assistance systems. He developed Daimler's lane departure warning system introduced in 2000. Since 2000 he has been head of Daimler's Image Understanding Group. The stereo technology developed by his group is the basis for the Mercedes Benz stereo camera system introduced in 2013. Recent work is on image understanding for autonomous driving, in particular Deep Neural Networks.

He was nominated for the "Deutscher Zukunftspreis", Germany's most prestigious award for Technology and Innovation given by the German President and awarded the Karl-Heinz Beckurts Prize 2012.

Abstract 15: **Invited Talk: Efficient Deep Networks for Real-Time Classification in Embedded Platforms (Jose Alvarez, NICTA, Australia) in Machine Learning for Intelligent Transportation Systems**, 04:00 PM

Abstract:
Convolutional neural networks have achieved impressive success in many tasks in computer vision such as image classification, object detection / recognition or semantic segmentation. While these networks have proven effective in all these applications, they come at a high memory and computational cost, thus not feasible for embedded platforms where power and computational resources are limited. In addition, the process to train the network reduces productivity as it not only requires large computer servers but also takes a significant amount of time (several weeks) with the additional cost of engineering the architecture. Recent works have shown there is significant redundancy in the parameters of deep architectures and therefore, could be replaced by more compact architectures. In this talk, I first introduce our efficient architecture based on filter-compositions and then, a novel approach to automatically determining the optimal number of neurons per layer in the architecture during the training process. As a result, we are able to deliver competitive accuracy and achieve up to 230fps in an embedded platform (Jetson TX-1). Moreover, these networks enable rapid prototyping as their entire training process only requires a few days.

Bio:
Dr. Jose M. Alvarez is a computer vision researcher at Data61 at CSIRO (formerly NICTA) working on efficient methods for large-scale dynamic scene understanding and deep learning. Dr. Alvarez graduated with his Ph.D. from Autonomous University of Barcelona (UAB) in October 2010. During his Ph.D., his research was focused on developing robust road detection algorithms for everyday driving tasks under real-world conditions. Dr. Alvarez visited the ISLA group at the University of Amsterdam (in 2008 and 2009), and the Group Research Electronics at Volkswagen (in 2010). Dr. Alvarez was awarded the best Ph.D. Thesis award in 2010 from the Autonomous University of Barcelona. Subsequently, Dr. Alvarez worked as a postdoctoral researcher at the Courant Institute of Mathematical Science, New York University. In 2012, Dr. Alvarez moved to the computer vision group at NICTA, Australia.

Since 2014, Dr. Alvarez serves as associate editor for IEEE Transactions on Intelligent Transportation Systems.

Abstract 16: **Invited Talk: Domain Adaption for Perception and Action (Kate Saenko, Boston University) in Machine Learning for Intelligent Transportation Systems**, *Saenko* 04:30 PM

Abstract:
Domain adaptation is a branch of machine learning that transfers knowledge from offline training domains to new test domains. Traditional supervised learning suffers from poor generalization when the test data distribution differs from training. This problem arises in many practical applications, including perception for autonomous vehicles. For example, if the perception model is trained on a dataset collected in specific weather conditions and/or geographical locations, its performance is likely to drop significantly in novel test conditions and locations. This is true even for deep neural models that are trained on large scale datasets. I will discuss our recent work focusing on domain adaptation in unsupervised scenarios, where the target domain is assumed to have no annotated labels. Specifically, I will describe a generalized framework based on end-to-end unsupervised domain alignment using domain-adaptive losses, such as the adversarial, maximum mean discrepancy, and correlation alignment losses. This work is in collaboration with the vision group at UC Berkeley.

Bio:
Prof. Kate Saenko is an Assistant Professor at the Computer Science Department at Boston University, and the director of the Computer Vision and Learning Group and member of the IVC group. Previously, she was an Assistant Professor at the UMass Lowell CS department, Postdoctoral Researcher at the International Computer Science Institute, a Visiting Scholar at UC Berkeley EECS and a Visiting Postdoctoral Fellow in the School of Engineering and Applied Science at Harvard University. Her research interests are in developing machine learning for image and language understanding, multimodal perception for autonomous systems, and adaptive intelligent human-computer interfaces.

Abstract 17: **Invited Talk: Learning Adaptive Driving Models from Large-scale Video Datasets (Fisher Yu, Huazhe Xu, Dequan Wang, and Trevor Darrell, Berkeley) in Machine Learning for Intelligent Transportation Systems**, *Darrell* 05:00 PM

Abstract:
Robust perception models should be learned from training data with diverse visual appearances and realistic behaviors. Exising datasets are limited in geographic extend, and can be biased to a source domain. We will overview two recent projects which makes use of a large scale dashcam video dataset. First, we'll present a novel domain adaptive dilation FCN, which adapts and improved performance on unlabeled data. Our model leverages both adversarial domain adaptation losses, and MIL-based boostrapping. We show results adapting from synthetic to real domains, and from classic driving datasets to in-the-wild dashcam data. Second, we'll show a model for end-to-end learning of driving policies from dashcam videos. Current approaches to deep visuomotor policy learning have been generally limited to in-situ models learned from a single vehicle or a simulation environment. We advocate learning a generic vehicle motion model from large scale crowd-sourced video data, and develop an end-to-end trainable architecture for learning to predict a distribution over future vehicle egomotion from instantaneous monocular camera observations and previous vehicle state. Our model incorporates a novel FCN-LSTM architecture, which can be learned from large-scale crowd-sourced vehicle action data, and leverages available scene segmentation side tasks to improve performance under a privileged learning paradigm. We provide a novel large-scale dataset of crowd-sourced driving behavior suitable for training our model, and report results predicting the driver action on held out sequences across diverse conditions.
Bio:
Prof. Darrell is on the faculty of the CS Division of the EECS Department at UC Berkeley and he is also appointed at the UC-affiliated International Computer Science Institute (ICSI). Darrell's group develops algorithms for large-scale perceptual learning, including object and activity recognition and detection, for a variety of applications including multimodal interaction with robots and mobile devices. His interests include computer vision, machine learning, computer graphics, and perception-based human computer interfaces. Prof. Darrell was previously on the faculty of the MIT EECS department from 1999-2008, where he directed the Vision Interface Group. He was a member of the research staff at Interval Research Corporation from 1996-1999, and received the S.M., and PhD. degrees from MIT in 1992 and 1996, respectively. He obtained the B.S.E. degree from the University of Pennsylvania in 1988, having started his career in computer vision as an undergraduate researcher in Ruzena Bajcsy's GRASP lab.

Imperfect Decision Makers: Admitting Real-World Rationality

__Miroslav Karny, David H Wolpert, David Rios Insua, Tatiana V. Guy__

Room 127 + 128, Fri Dec 09, 08:00 AM

The prescriptive (normative) Bayesian theory of decision making under uncertainty has reached a high level of maturity. The assumption that the decision maker is rational (i.e. that they optimize expected utility, in Savage's formulation) is central to this theory. However, empirical research indicates that this central assumption is often violated by real decision-makers. This limits the ability of the prescriptive Bayesian theory to provide a descriptive theory of the real world. One of the reasons that have been proposed for why the assumption of rationality might be violated by real decision makers is the limited cognitive and computational resources of those decision makers, [1]-[5]. This workshop intends to inspect this core assumption and to consider possible ways to modify or complement it.

Many of the precise issues related to this theme – some of which will be addressed in the invited talks - can be formulated as questions:

• Does the concept of rationality require Bayesian reasoning?
• Does quantum probability theory (extending classical Kolmogorov probability) provide novel insights into the relation between decision making and cognition?
• Do the extensions of expected utility (which is a linear function of the relevant probabilities) to nonlinear functions of probabilities enhance the flexibility of decision-making task formulating while respecting the limited cognitive resources of decision makers?
• How can good (meta-)heuristics, so successfully used by real-world decision makers, be elicited?

The list is definitely not complete and we expect that contributed talks, posters and informal discussions will extend it. To stimulate the informal discussions, the invited talks will be complemented by discussants challenging them. Altogether, the workshop aims to bring together diverse scientific communities, to brainstorm possible research directions, and to encourage collaboration among researchers with complementary ideas and expertise. The intended outcome is to understand and diminish the discrepancy between the established prescriptive theory and real-world decision making.

The targeted audience is scientists and students from the diverse scientific communities (decision science, cognitive science, natural science, artificial intelligence, machine learning, social science, economics, etc.) interested in various aspects of rationality.

All accepted submissions will be published in a special issue of the Workshop and Conference Proceedings series of the Journal of Machine Learning Research (JMRL).

[1] H.A. Simon: Theories Of Decision-Making In Economics and Behavioral Science, The American Economic Review, XLIX, 253-283, 1959
[2] C.A. Sims Implications of Rational Inattention, J. of Monetary Economics, 50, 3, 665 -- 690, 2003
[3] A. Tversky, D. Kahneman: Advances in Prospect Theory: Cumulative Representation of Uncertainty, J. of Risk and Uncertainty, 5, 297-323, 1992
[4] 2011 NIPS Workshop on Decision Making with Multiple Imperfect Decision Makers
[5] 2015 NIPS Workshop on Bounded Optimality and Metareasoning

Schedule

08:20 AM	**Opening session**	
08:30 AM	**Bounded Optimality and Rational Metareasoning in Human Cognition**	*Griffiths*
09:00 AM	**Rationality and the Bayesian Paradigm**	*Gilboa*
09:30 AM	**Information-Theoretic Bounded Rationality for Learning and Decision-Making**	*Braun*
10:00 AM	**Poster Spotlights**	
10:30 AM	**Coffee break & Poster session**	
11:00 AM	**Principles and Algorithms for Self-Motivated Behaviour**	*Tishby*
11:30 AM	**Rational beliefs real agents can have -- A logical point of view**	
11:50 AM	**Overcoming temptation: Incentive design for intertemporal choice**	*Mozer*
12:10 PM	**(Ir-)rationality of human decision making**	*Grünwald*
12:30 PM	**Lunch break**	
02:00 PM	**The Rational Status of Quantum Probability Theory Applied to Human Decision Making**	*Pleskac*
02:30 PM	**Quantum Rational Preferences and Desirability**	*Benavoli*
02:50 PM	**Coffee break & Poster session**	
03:30 PM	**Safe Probability**	*Grünwald*
03:50 PM	**What the Recent Revolution in Network Coding Tells Us About the Organization of Social Groups**	*Wolpert*
04:20 PM	**Agency and Causality in Decision Making**	*Ortega*
04:50 PM	**Modelling of Rational Decision Making**	*Wolpert*
05:20 PM	**Closing session**	

Abstracts (12):

Abstract 1: **Opening session in Imperfect Decision Makers: Admitting Real-World Rationality**, 08:20 AM

Introductory comments by the organisers.

Abstract 2: **Bounded Optimality and Rational Metareasoning in Human Cognition in Imperfect Decision Makers: Admitting Real-World Rationality**, *Griffiths* 08:30 AM

Human decision-making is often described as irrational, being the result of applying error-prone heuristics. I will argue that this is partly a consequence of the use of an unrealistic standard of rationality, and that the notion of bounded optimality from the artificial intelligence literature provides a better framework for understanding human behaviour. Within this framework a rational agent seeks to execute the best algorithm for solving a problem, taking into account available computational resources and the cost of time. We find that several classic heuristics from the decision-making literature are bounded optimal, assuming people have access to particular computational resources. This establishes a new problem: how do people find such good heuristics? I will discuss how this problem can be addressed via rational metareasoning, which examines how rational agents should decide what algorithm to use in solving a problem. The result is a view of human decision-making in which people are intelligently and flexibly making the most of their limited computational resources.

Abstract 3: **Rationality and the Bayesian Paradigm in Imperfect Decision Makers: Admitting Real-World Rationality**, *Gilboa* 09:00 AM

It is argued that, contrary to a rather prevalent view within economic theory, rationality does not imply Bayesianism. The note begins by defining these terms and justifying the choice of these definitions, proceeds to survey the main justification for this prevalent view, and concludes by highlighting its weaknesses.

Abstract 4: **Information-Theoretic Bounded Rationality for Learning and Decision-Making in Imperfect Decision Makers: Admitting Real-World Rationality**, *Braun* 09:30 AM

We study an information-theoretic framework of bounded rational decision-making that trades off utility maximization against information-processing costs. We apply the basic principle of this framework to perception-action systems and show how the formation of abstractions and decision-making hierarchies depends on information-processing costs.

Abstract 5: **Poster Spotlights in Imperfect Decision Makers: Admitting Real-World Rationality**, 10:00 AM

Posters:
Marcus Buckmann, Özgür Simsek:
Decision Heuristics For Comparison: How Good Are They?
Sam Ganzfried:
Optimal Number of Choices in Rating Contexts.
Jan Malte Lichtenberg, Özgür Simsek:
Simple Regression Models.
Miroslav Kárný:
Towards Implementable Prescriptive Decision Making.
Krzysztof Drachal:
Forecasting Spot Oil Price Using Google Probabilities.
Özgür Simsek, Marcus Buckmann:
On Learning Decision Heuristics.
Marko Ruman, František Hula, Tatiana V. Guy, Miroslav Kárný:
Real-Life Performance of Deliberation-Aware Responder in Multi-Proposer Ultimatum Game.
Shaudi Mahdavi, Mohammad Amin Rahimian:
Does Hindsight Bias Impede Learning?
Vladimíra Seckarová:
Performance of Kullback-Leibler Based Expert Opinion Pooling for Unlikely Events.
Jakub Štech, T.V.Guy:
Lazy-learning fully probabilistic decision making.
Dimitri Ognibene, Vincenzo G. Fiore, Xiaosi Gu:
Addiction in a Bounded Rational Model: the role of Exploration and Environment Structure.

Abstract 7: **Principles and Algorithms for Self-Motivated Behaviour in Imperfect Decision Makers: Admitting Real-World Rationality**, *Tishby* 11:00 AM

For planning with high uncertainty, or with too many possible end positions as in games like Go or even chess, one can almost never solve the optimal control problem and must use some receding horizon heuristics. One such heuristics is based on the idea of maximizing empowerment, namely, keep the number of possible options maximal. This has been formulated using information theoretic ideas as maximizing the information capacity between the sequence of actions and the possible state of the system at some finite horizon, but no efficient algorithm for calculating this capacity was suggested. In this work we propose a concrete and efficient way for calculating the capacity between a sequence of actions and future states, based on local linearization of the dynamics and Gaussian channel capacity calculation. I will describe the new algorithm and some of its interesting implications.

Abstract 8: **Rational beliefs real agents can have -- A logical point of view in Imperfect Decision Makers: Admitting Real-World Rationality**, 11:30 AM

The purpose of this note is to outline a framework for uncertain reasoning which drops unrealistic assumptions about the agents' inferential capabilities. To do so, we envisage a
pivotal role for the recent research programme of depth-bounded Boolean logics (D'Agostino et al., 2013). We suggest that this can be fruitfully extended to the representation of rational belief under uncertainty. By doing this we lay the foundations for a prescriptive account of rational belief, namely one that realistic agents, as opposed to idealised ones, can feasibly act upon.

Abstract 9: **Overcoming temptation: Incentive design for intertemporal choice in Imperfect Decision Makers: Admitting Real-World Rationality**, *Mozer* 11:50 AM

Individuals are often faced with temptations that can lead them astray from long-term goals. We're interested in developing interventions that steer individuals toward making good initial decisions and then maintaining those decisions over time. In the realm of financial decision making, a particularly successful approach is the prize-linked savings account: individuals are incentivized to make deposits by tying deposits to a periodic lottery that awards bonuses to the savers. Although these lotteries have been very effective in motivating savers across the globe, they are a one-size-fits-all solution. We investigate
whether customized bonuses can be more effective. We formalize a delayed-gratification task as a Markov decision problem and characterize individuals as rational agents subject to temporal discounting, costs associated with effort, and moment-to-moment fluctuations in willpower. Our theory is able to explain key behavioral findings in intertemporal choice. We
created an online delayed-gratification game in which the player scores points by choosing a queue to wait in and patiently advancing to the front. Data collected from the game is fit to the model, and the instantiated model is then used to optimize predicted player performance over a space of incentives. We demonstrate that customized incentive structures can improve goal-directed decision making.

Abstract 12: **The Rational Status of Quantum Probability Theory Applied to Human Decision Making in Imperfect Decision Makers: Admitting Real-World Rationality**, *Pleskac* 02:00 PM

Quantum probability theory (QPT) is a probabilistic framework, alternative to Classic Probability Theory (CPT) that has been employed to model some of the paradoxical phenomena found with human judgments and decisions. One question that arises, however, is why an agent might behave this way especially given that these judgments and decisions appear to deviate from rationality? We will argue that QPT can fulfill the requirement for the Dutch Book theorem, which has been used to justify the rational status of CPT. A second question is how these quantum processes work? We will show how the heuristic processes people use to make judgments and decisions can be modeled with quantum information theory, which perhaps paradoxically provides a better and more parsimonious description of these boundedly rational heuristic processes people use than models grounded in classic information theory. In sum, we will argue that QPT can offer a principled account of the processes people use to make judgments and decisions with their limited computational resources and those judgments and decisions can nevertheless be quite rational.

Abstract 13: **Quantum Rational Preferences and Desirability in Imperfect Decision Makers: Admitting Real-World Rationality**, *Benavoli* 02:30 PM

We develop a theory of quantum rational decision making in the tradition of Anscombe and Aumann's axiomatisation of preferences on horse lotteries. It is essentially the Bayesian decision theory generalised to the space of Hermitian matrices. Among other things, this leads us to give a representation theorem showing that quantum complete rational preferences are obtained by means of expected utility considerations.

Abstract 15: **Safe Probability in Imperfect Decision Makers: Admitting Real-World Rationality**, *Grünwald* 03:30 PM

We formalize the idea of probability distributions that lead to reliable predictions about some, but not all aspects of a domain. The resulting notion of `safety' provides a fresh perspective on foundational issues in statistics, providing a middle ground between imprecise probability and multiple-prior models on the one hand and strictly Bayesian approaches on the other. It also allows us to formalize fiducial distributions in terms of the set of random variables that they can safely predict, thus taking some of the sting out of the fiducial idea. By restricting probabilistic inference to safe uses, one also automatically avoids paradoxes such as the Monty Hall problem. Safety comes in a variety of degrees, such as `validity' (the strongest notion), `calibration', `confidence safety' and `unbiasedness' (almost the weakest notion).

Abstract 17: **Agency and Causality in Decision Making in Imperfect Decision Makers: Admitting Real-World Rationality**, *Ortega* 04:20 PM

We review the distinction between evidential and causal decision-making and the challenges that this distinction poses to the application of the expected utility principle. We furthermore establish firm connections between causality, information-theory, and game-theoretic concepts. Finally, we show how to use the aforementioned connections to construct adaptive agents that are universal over a given class of stochastic environments - such as Thompson sampling.

Challenges in Machine Learning: Gaming and Education

Isabelle Guyon, Evelyne Viegas, Balázs Kégl, Ben Hamner, Sergio Escalera

Room 129 + 130, Fri Dec 09, 08:00 AM

Challenges in machine learning and data science are competitions running over several weeks or months to resolve problems using provided datasets or simulated environments. The playful nature of challenges naturally attracts students, making challenge a great teaching resource. For this third edition of the CiML workshop at NIPS we want to explore more in depth the opportunities that challenges offer as teaching tools. The workshop will give a large part to discussions around several axes: (1) benefits and limitations of challenges to give students problem-solving skills and teach them best practices in machine learning; (2) challenges and continuous education and up-skilling in the enterprise; (3) design issues to make challenges more effective teaching aids; (4) curricula involving students in challenge design as a means of educating them about rigorous experimental design, reproducible research, and project leadership. CiML is a forum that brings together workshop organizers, platform providers, and participants to discuss best practices in challenge organization and new methods and application opportunities to design high impact challenges. Following the success of last year's workshop (http://ciml.chalearn.org/), in which a fruitful exchange led to many innovations, we propose to reconvene and discuss new opportunities for challenges in education, one of the hottest topics identified in last year's discussions. We have invited prominent speakers in this field. We will also reserve time to an open discussion to dig into other topic including open innovation, coopetitions, platform interoperability, and tool mutualisation.

Schedule

08:00 AM	**Welcome**	*Viegas*
08:30 AM	**Gathering common sense knowledge: how to game it?**	*Zitnick*
09:10 AM	**The Michigan Data Science Team: A Student Organization for Machine Learning Challenges**	*Stroud*
09:30 AM	**Energy generation prediction: Lessons learned from the use of Kaggle in Machine Learning Course**	*Fernandez-Bes*
09:50 AM	**Learning to improve learning: ML in the classroom**	*Brunskill*
11:00 AM	**Challenges in education**	*Kégl, Hamner*
12:00 PM	**Lunch, posters and discussions**	
02:00 PM	**OpenML in research and education**	*Vanschoren*
02:40 PM	**ImageCLEF 2017 LifeLog task**	*Dang Nguyen*
03:30 PM	**Evaluation-as-a-Service: a serious game**	*Mueller*
04:10 PM	**Reproducible Research: moving to the BEAT**	*Marcel*
04:50 PM	**CAFA: a Challenge Dedicated to Understanding the Function of Biological Macromolecules**	*Radivojac*
05:10 PM	**Interactive Machine Learning (iML): a challenge for Game-based approaches**	*Holzinger*
05:30 PM	**Gaming challenges and encouraging collaborations**	*Escalera, Guyon*

Private Multi-Party Machine Learning

Borja Balle, Aurélien Bellet, David Evans, Adrià Gascón

Room 131 + 132, Fri Dec 09, 08:00 AM

The workshop focuses on the problem of privacy-preserving machine learning in scenarios where sensitive datasets are distributed across multiple data owners. Such distributed scenarios occur quite often in practice, for example when different parties contribute different records to a dataset, or information about each record in the dataset is held by different owners. Different communities have developed approaches to deal with this problem, including differential privacy-like techniques where noisy sketches are exchanged between the parties, homomorphic encryption where operations are performed on encrypted data, and tailored approaches using techniques from the field of secure multi-party computation. The workshop will serve as a forum to unify different perspectives on this problem and explore the relative merits of each approach. The workshop will also serve as a venue for networking researchers from the machine learning and secure multi-party computation communities interested in private learning, and foster fruitful long-term collaborations. The workshop will have a particular emphasis in the decentralization aspect of privacy-preserving machine learning. This includes a large number of realistic scenarios where the classical setup of differential privacy with a "trusted curator" that prepares the data cannot be directly applied. The problem of privacy-preserving computation gains relevance in this model, and effectively leveraging the tools developed by the cryptographic community to develop private multi-party learning algorithms poses a remarkable challenge. Our program will include an introductory tutorial to secure multi-party computation for a machine learning audience, and talks by world-renowned experts from the machine learning and cryptography communities who have made high quality contributions to this problem.

Schedule

09:00 AM	**Kobbi Nissim**
09:45 AM	**Mariana Raykova**
11:00 AM	**Jack Doerner**
11:30 AM	**Stratis Ioannidis**
12:15 PM	**Poster Spotlights**
02:30 PM	**TBD 1**
03:30 PM	**Poster Session**
04:30 PM	**Dawn Song**
05:15 PM	**TBD 2**
05:35 PM	**TBD 3**
05:55 PM	**Nina Balcan**

Learning, Inference and Control of Multi-Agent Systems

Thore Graepel, Marc Lanctot, Joel Z Leibo, Guy Lever, Janusz Marecki, Frans A Oliehoek, Karl Tuyls, Vicky Holgate

Room 133 + 134, Fri Dec 09, 08:00 AM

We live in a multi-agent world and to be successful in that world, agents, and in particular, artificially intelligent agents, will need to learn to take into account the agency of others. They will need to compete in market places, cooperate in teams, communicate with others, coordinate their plans, and negotiate outcomes. Examples include self-driving cars interacting in traffic, personal assistants acting on behalf of humans and negotiating with other agents, swarms of unmanned aerial vehicles, financial trading systems, robotic teams, and household robots.

Furthermore, the evolution of human intelligence itself presumably depended on interaction among human agents, possibly starting out with confrontational scavenging [1] and culminating in the evolution of culture, societies, and language. Learning from other agents is a key feature of human intelligence and an important field of research in machine learning [2]. It is therefore conceivable that exposing learning AI agents to multi-agent situations is necessary for their development towards intelligence.

We can also think of multi-agent systems as a design philosophy for complex systems. We can analyse complex systems in terms of agents at multiple scales. For example, we can view the system of world politics as an interaction of nation state agents, nation states as an interaction of organizations, and further down into departments, people etc. Conversely, when designing systems we can think of agents as building blocks or modules interacting to produce the behaviour of the system, e.g. [3].

Multi-agent systems can have desirable properties such as robustness and scalability, but their design requires careful consideration of incentive structures, learning, and communication. In the most extreme case, agents with individual views of the world, individual actuators, and individual incentive structures need to coordinate to achieve a common goal. To succeed they may need a Theory of Mind that allows them to reason about other agents' intentions, beliefs, and behaviours [4]. When multiple learning agents are interacting, the learning problem from each agent's perspective may become non-stationary, non-Markovian, and only partially observable. Studying the dynamics of learning algorithms could lead to better insight about the evolution and stability of such systems [5].

Problems involving competing or cooperating agents feature in recent AI breakthroughs in competitive games [6,7], current ambitions of AI such as robotic football teams [8], and new research into emergent language and agent communication in reinforcement learning [9,10].

In summary, multi-agent learning will be of crucial importance to the future of computational intelligence and pose difficult and fascinating problems that need to be addressed across disciplines. The paradigm shift from single-agent to multi-agent systems will be pervasive and will require efforts across different fields including machine learning, cognitive science, robotics, natural computing, and (evolutionary) game theory. In this workshop we aim to bring together researchers from these different fields to discuss the current state of the art, future avenues and visions for work regarding theory and practice of multi-agent learning, inference, and decision-making.

Topics we consider for inclusion in the workshop include multi-agent reinforcement learning; deep multi-agent learning; theory of mind; multi-agent communication; POMDPs, Dec-POMDPS and partially observable stochastic games; multi-agent robotics, human-robot collaboration, swarm robotics; game theory, mechanism design, algorithms for computing nash equilibria and other solution concepts; bioinspired approaches, swarm intelligence and collective intelligence; co-evolution, evolutionary dynamics and culture; ad hoc teamwork.

[1] 'Confrontational scavenging as a possible source for language and cooperation', Derek Bickerton and Eörs Szathmáry, BMC Evolutionary Biology 2011

[2] ‘Apprenticeship Learning via Inverse Reinforcement Learning’, Pieter Abbeel and Andrew Y. Ng, ICML 2004
[3] ‘The Society of Mind’, Marvin Minsky, 1986
[4] ‘Building Machines That Learn and Think Like People’, Brenden M. Lake et al., CBMM Memo 2016
[5] ‘Evolutionary Dynamics of Multi-Agent Learning: A Survey’, Daan Bloembergen et al., JAIR 2015
[6] 'Mastering the game of Go with deep neural networks and tree search', David Silver et al., Nature 2016
[7] 'Heads-up limit hold’em poker is solved', Michael Bowling et al., Science 2015
[8] RoboCup, http://www.robocup.org/
[9] 'Learning to Communicate with Deep Multi-Agent Reinforcement Learning', Jakob N. Foerster et al., Arxiv 2016
[10] 'Learning Multiagent Communication with Backpropagation', Sainbayar Sukhbaatar et al. Arxiv 2016

Schedule

08:30 AM	**Introduction**	*Graepel, Tuyls, Oliehoek*
08:50 AM	**Learning to Communicate with Deep Multi–Agent Reinforcement Learning**	*Whiteson*
09:40 AM	**Computer Curling: AI in Sports Analytics**	*Bowling*
11:00 AM	**Reverse engineering human cooperation (or, How to build machines that treat people like people)**	*Tenenbaum, Kleiman-Weiner*
11:50 AM	**Spotlight Session**	
12:40 PM	**Lunch**	
01:30 PM	**Poster Session**	
02:10 PM	**Multi-Agent and Multi-Robot Coordination with Uncertainty and Limited Communication**	
03:00 PM	**Coffee Break**	
03:30 PM	**Safe, Multi-Agent, Reinforcement Learning for Autonomous Driving**	
03:50 PM	**A Study of Value Iteration with Non-Stationary Strategies in General Sum Markov Games**	*Pérolat*
04:10 PM	**Learning to Assemble Objects with Robot Swarms**	*Neumann*
04:30 PM	**Break**	
04:50 PM	**Challenges on the way to fully autonomous swarms of drones**	*de Croon*
05:40 PM	**Discussion Panel**	
06:20 PM	**Concluding Remarks**	*Graepel, Oliehoek, Tuyls*

Abstracts (1):

Abstract 14: **Challenges on the way to fully autonomous swarms of drones in Learning, Inference and Control of Multi-Agent Systems**, *de Croon* 04:50 PM

While a single, small robot is limited in its capabilities to perform complex tasks, large groups or "swarms" of such robots have a much bigger potential. Physically, they can collaborate to move heavier things, cross gaps bigger than a single robot body length, or explore unknown areas much quicker. Mentally, they can take in and process much more information than a single robot could, even if communication is extremely limited. In the NIPS 2016 workshop on multi-agent systems, it is suggested that true Artificial Intelligence can only be reached by having robots interact with each other, and it is well-known that groups of robots potentially have a much larger collective learning potential than animals or humans.

So, why are we not yet seeing many such robotic swarms in the real world or even in academia? In my talk I will go into the challenges of making an autonomous swarm of tiny drones explore an unknown building. These drones are < 50 grams and have to fly around, avoid obstacles, navigate, and work together for the most efficient exploration. I will highlight how complex these various challenges are and report on a specific study in which we have drones use their bluetooth modules to avoid each other, should they find themselves in the same small indoor space. This case study will illustrate what are in my eyes the major challenges towards the promised autonomous robotic swarms.

Brains and Bits: Neuroscience meets Machine Learning

Eva L Dyer, Allie Fletcher, Jascha Sohl-Dickstein, Joshua T Vogelstein, Konrad Koerding, Jakob H Macke

Room 211, Fri Dec 09, 08:00 AM

The goal of this workshop is to bring together researchers from neuroscience, deep learning, machine learning, computer science theory, and statistics for a rich discussion about how computer science and neuroscience can inform one another as these two fields rapidly move forward. We invite high quality submissions and discussion on topics including, but not limited to, the following fundamental questions: a) shared approaches for analyzing biological and artificial neural systems, b) how insights and challenges from neuroscience can inspire progress in machine learning, and c) methods for interpreting the revolutionary large scale datasets produced by new experimental neuroscience techniques.

Experimental methods for measuring neural activity and structure have undergone recent revolutionary advances, including in high-density recording arrays, population calcium imaging, and large-scale reconstructions of anatomical circuitry. These developments promise unprecedented insights into the collective dynamics of neural populations and thereby the underpinnings of brain-like computation. However, these next-generation methods for measuring the brain’s architecture and function produce high-dimensional, large scale, and complex datasets, raising challenges for analysis. What are the machine learning and analysis approaches that will be indispensable for analyzing these next-generation datasets? What are the computational bottlenecks and challenges that must be overcome?

In parallel to experimental progress in neuroscience, the rise of deep learning methods has shown that hard computational problems can be solved by machine learning algorithms that are inspired by biological neural networks, and built by cascading many nonlinear units. In contrast to the brain, artificial neural systems are fully observable, so that experimental data-collection constraints are not relevant. Nevertheless, it has proven challenging to develop a theoretical understanding of how neural networks solve tasks, and what features are critical to their performance. Thus, while deep networks differ from biological neural networks in many ways, they provide an interesting testing ground for evaluating strategies for understanding neural processing systems. Are there synergies between analysis methods for biological and artificial neural systems? Has the resurgence of deep learning resulted in new hypotheses or strategies for trying to understand biological neural networks? Conversely, can neuroscience provide inspiration for the next generation of machine-learning algorithms?

We welcome participants from a range of disciplines in statistics, applied physics, machine learning, and both theoretical and experimental neuroscience, with the goal of fostering interdisciplinary insights. We hope that active discussions among these groups can set in motion new collaborations and facilitate future breakthroughs on fundamental research problems.

Schedule

08:30 AM	**Welcome and Opening Remarks**	*Dyer, Fletcher*
08:45 AM	**Christos Papadimitriou : A computer scientist thinks about the brain**	
09:30 AM	**Cristina Savin : Spike-Based Probabilistic Computation**	
10:00 AM	**Mitya Chklovskii : Toward Biologically Plausible Machine Learning: A Similarity Matching Approach**	
10:30 AM	**Coffee Break 1a (plus posters)**	
11:00 AM	**Jonathan Pillow : Scalable Inference for Structured Hierarchical Receptive Field Models**	
11:30 AM	**Emily Fox : Functional Connectivity in MEG via Graphical Models of Time Series**	
12:00 PM	**Srini Turaga**	*Turaga*
12:30 PM	**Lunch Day 1**	
02:00 PM	**Il Memming Park : Dynamical Systems Interpretation of Neural Trajectories**	
02:30 PM	**David Sussillo : LFADS - Latent Factor Analysis via Dynamical Systems**	
03:00 PM	**Coffee Break 1b (plus posters)**	
03:30 PM	**Poster Session 1**	
04:00 PM	**TBA**	
04:30 PM	**Michael Buice : The Allen Brain Observatory**	
05:00 PM	**Stefan Mihalas : Modeling Optimal Context Integration in Cortical Columns**	
05:30 PM	**Breakout Discussion Afternoon Session**	*Koerding, Dyer*
08:30 AM	**Opening Remarks**	*Dyer, Sohl-Dickstein*
08:45 AM	**Yoshua Bengio : Toward Biologically Plausible Deep Learning**	
09:30 AM	**Surya Ganguli : Deep Neural Models of the Retinal Response to Natural Stimuli**	
10:00 AM	**Max Welling : Making Deep Learning Efficient Through Sparsification**	
10:30 AM	**Coffee Break 2a (plus posters)**	
11:00 AM	**David Cox : Predictive Coding for Unsupervised Feature Learning**	
11:30 AM	**From Brains to Bits and Back Again**	*Bengio, Sejnowski, Papadimitriou, Macke, Denève, Hassabis, Fletcher, Koerding*
12:30 PM	**Lunch Day 2**	
02:00 PM	**Fred Hamprecht : Motif Discovery in Functional Brain Data**	
02:30 PM	**Anima Anandkumar**	*Anandkumar*
03:00 PM	**Coffee Break 2b (plus posters)**	
03:30 PM	**Poster Session 2**	
04:00 PM	**Kanitschneider : Training Recurrent Networks to Generate Hypotheses About How the Brain Solves Hard Navigation Problems**	

04:30 PM	**Jorg Lucke : Probabilistic Inference and the Brain: Towards General, Scalable, and Deep Approximations**

Machine Intelligence @ NIPS

Tomas Mikolov, Baroni Marco, Armand Joulin, Germán Kruszewski, Angeliki Lazaridou, Klemen Simonic

Room 212, Fri Dec 09, 08:00 AM

Recent years have seen the success of machine learning systems, in particular deep learning architectures, on specific challenges such as image classification and playing Go. Nevertheless, machines still fail on hallmarks of human intelligence such as the flexibility to quickly switch between a number of different tasks, the ability to creatively combine previously acquired skills in order to perform a more complex goal, the capacity to learn a new skill from just a few examples, or the use of communication and interaction to extend one's knowledge in order to accomplish new goals. This workshop aims to stimulate theoretical and practical advances in the development of machines endowed with human-like general-purpose intelligence, focusing in particular on benchmarks to train and evaluate progress in machine intelligence. The workshop will feature invited talks by top researchers from machine learning, AI, cognitive science and NLP, who will discuss with the audience their ideas about what are the most pressing issues we face in developing true AI and the best methods to measure genuine progress. We are moreover calling for position statements from interested researchers to complement the workshop program. The workshop will also introduce the new Environment for Communication-Based AI to the research community, encouraging discussion on how to make it the ultimate benchmark for machine intelligence. The Environment aims at being an interactive playground where systems can only succeed if they possess the hallmarks of intelligence we listed above. In September, we will make a prototype of the Environment available, so that researchers interested in submitting position statements to the workshop can experiment with it and take it into account in their proposals.

Schedule

08:30 AM	**Klemen Simonic - Introduction**
09:00 AM	**Marco Baroni - A roadmap for communication-based AI**
09:20 AM	**Allan Jabri - The commAI-env environment for communication-based AI**
09:30 AM	**Raquel Fernandez - Human-like dialogue: Key challenges for AI**
09:55 AM	**Jürgen Schmidhuber: Learning incrementally to become a general problem solver**
10:20 AM	**Rudolf Kadlec, Ondrej Bajgar, Jan Kleindienst - From particular to general: A preliminary case study of transfer learning in reading comprehension**
10:30 AM	**Coffee break**
11:00 AM	**Marek Rosa, Jan Feyereisl - Consolidating the search for general AI**
11:10 AM	**Alex Peysakhovich - Gaining insights from game theory about the emergence of communication**
11:20 AM	**Tomo Lazovich, Matthew C. Graham, Troy M. Lau, Joshua C. Poore - Socially constructed machine intelligence**
11:30 AM	**Douwe Kiela, Luana Bulat, Anita L. Vero, Stephen Clark - Virtual embodiment: A scalable long-term strategy for Artificial Intelligence research**
11:40 AM	**Panel on Basic requirements for Machine Intelligence: Angeliki Lazaridou (moderator), Katja Hofmann, Brenden Lake, Jürgen Schmidhuber, Arthur Szlam, Jan Feyereisl, Rudolf Kadlec, Armand Joulin**
12:30 PM	**Lunch Break**
02:00 PM	**Brenden Lake - Building machines that learn and think like people**
02:25 PM	**Fernando Diaz - Malmo: Flexible and scalable evaluation in Minecraft**
02:50 PM	**Jon Gauthier, Igor Mordatch - A paradigm for situated and goal-driven language learning**
03:00 PM	**Coffee break 2**
03:30 PM	**Arthur Szlam - In praise of fake AI**
03:55 PM	**Emmanuel Dupoux - An evolutionary perspective on machine intelligence**

04:20 PM	**Julian Togelius - Are video games the perfect environments for developing artificial general intelligence? Which kind of general intelligence?**	
04:45 PM	**Steven Hansen - Minimally naturalistic Artificial Intelligence**	
04:55 PM	**Gemma Boleda - Remarks on the CommAI-env**	
05:05 PM	**Panel on CommAI-env and other environments for the development of AI: Germán Kruszewski(moderator), Julian Togelius, Tomas Mikolov, Emmanuel Dupoux, Raquel Fernandez, Alex Peysakhovich, Gemma Boleda, Igor Mordatch**	

People and machines: Public views on machine learning, and what this means for machine learning researchers

Peter Donnelly, Jessica Montgomery, Susannah Odell, Sabine Hauert, Zoubin Ghahramani, Katherine Gorman

VIP Room, Fri Dec 09, 12:00 PM

The Royal Society is currently carrying out a major programme of work on machine learning, to assess its potential over the next 5-10 years, barriers to realising that potential, and the legal, ethical, social and scientific questions which arise as machine learning becomes more pervasive.
As part of this work, the Royal Society has carried out a public dialogue exercise to explore public awareness of, and attitudes towards, machine learning and its applications. The results of this work illustrate some of the key questions people have about machine learning; about why it is used, for what purpose, and with what pattern of benefits and disbenefits. It draws attention to the need to enable informed public debate that engages with specific applications.

In addition, machine learning is put to use in a range of different applications, it reframes existing social and ethical challenges, such as those relating to privacy and stereotyping, and also creates new challenges, such as interpretability, robustness and human-machine interaction. Many of these form the basis of active and stimulating areas of research, which can both move forward the field of machine learning and help address key governance issues.

The UK’s experience with other emerging technologies shows that it is possible to create arrangements that enable a robust public consensus on the safe and valuable use of even the most potentially contentious technologies. An effective dialogue process with the public can help to create these arrangements. From Twitter to Ted Talks, machine learning researchers have a range of ways in which they can engage with the public, and take an active role in public discussions about this technology. Yet, much of what the public hears about machine learning from the media focuses on accidents involving autonomous machines, or fears about labour market changes caused by direct substitution of people for machines.

This lunchtime session will present new research on the public’s view of machine learning, alongside a discussion of how research can help address some of the broader social challenges associated with machine learning.

Speakers: Dr Sabine Hauert speak about the Royal Society's recent public dialogues on machine learning and why it is important to engage with the public. Professor Zoubin Ghahramani will then explore the role of machine learning research in addressing areas of social concern, such as transparency and interpretability. Katherine Gorman will then discuss tools for communicating research to the public.

Lunch will be provided for attendees.

Schedule

12:00 PM	**Introduction by Chair**	*Donnelly*
12:10 PM	**Understanding the Public's Views of Social Benefit and Social Risk: Lessons from the Royal Society's Public Dialogue and What This Means for Science Communication**	*Hauert*
12:40 PM	**How Machine Learning Research Can Address Key Societal and Governance Issues**	*Ghahramani*
01:10 PM	**Crafting a Story to Communicate Your Research to the Public Using the 'Algorithm Toolkit'**	*Gorman*
01:30 PM	**Extended Q&A, including Questions from Twitter: #RSmachinelearning**	

Neurorobotics: A Chance for New Ideas, Algorithms and Approaches

Elmar Rueckert, Martin Riedmiller

VIP Room, Fri Dec 09, 14:30 PM

Workshop webpage: http://www.neurorobotic.eu

Modern robots are complex machines with many compliant actuators and various types of sensors including depth and vision cameras, tactile electrodes and dozens of proprioceptive sensors. The obvious challenges are to process these high dimensional input patterns, memorize low dimensional representations of them and to generate the desired motor commands to interact in dynamically changing environments. Similar challenges exist in brain machine interfaces (BMIs) where complex prostheses with perceptional feedback are controlled, or in motor neuroscience where in addition cognitive features need to be considered. Despite this broad research overlap the developments happened mainly in parallel and were not ported or exploited in the related domains. The main bottleneck for collaborative studies has been a lack of interaction between the core robotics, the machine learning and the neuroscience communities.

Why is it now just the right time for interactions?

- Latest developments based on deep neural networks have advanced the capabilities of robotic systems by learning control policies directly from the high dimensional sensor readings.
- Many variants of networks have been recently developed including the integration of feedback through recurrent connections, the projection to different feature spaces, may be trained at different time scales and can be modulated through additional inputs.
- These variants can be the basis for new models and concepts in motor neuroscience, where simple feed forward structures were not sufficiently powerful.
- Robotic applications demonstrated the feasibility of such networks for real time control of complex systems, which can be exploited in BMIs.
- Modern robots and new sensor technologies require models that can integrate a huge amount of inputs of different dimension, at different rates and with different noise levels. The neuroscience communities face such challenges and develop sophisticated models that can be evaluated in robotic applications used as benchmarks.
- New learning rules can be tested on real systems in challenging environments.

Topics:

- Convolutional Networks and Real-time Robotic and Prosthetic applications
- Deep Learning for Robotics and Prosthetics
- End-to-End Robotics / Learning
- Feature Representations for Big Data
- Movement Representations, Movement Primitives and Muscle Synergies
- Neural Network Hardware Implementation, Neuromorphic Hardware
- Recurrent Networks and Reservoirs for Control of high dimensional systems
- Reinforcement Learning and Bayesian Optimization in Neural Networks from multiple reward sources
- Sampling Methods and Spiking Networks for Robotics
- Theoretical Learning Concepts, Synaptic Plasticity Rules for Neural Networks

Schedule

02:30 PM	**Juergen Schmidhuber (Scientific Director of the Swiss AI Lab IDSIA)**	*Schmidhuber*
03:30 PM	**Sergey Levine (University of California, Berkeley)**	*Levine*
04:00 PM	**Pieter Abbeel (University of California, Berkeley)**	*Abbeel*
04:30 PM	**Johanni Brea (École polytechnique fédérale de Lausanne, EPFL)**	
05:20 PM	**Paul Schrater (University of Minnesota)**	*Schrater*
05:45 PM	**Frank Hutter (University Freiburg)**	*Hutter*
06:10 PM	**Raia Hadsell (Google DeepMind)**	
06:35 PM	**Panel Discussion, Session One: Reinforcement Learning, Imitation, and Active Learning**	
08:30 AM	**Introduction by Elmar Rueckert and Martin Riedmiller**	
08:35 AM	**Robert Legenstein (Graz University of Technology)**	*Legenstein*
09:05 AM	**Sylvain Calinon (Idiap Research Institute, EPFL Lausanne)**	
09:35 AM	**Chelsea Finn (University of California, Berkeley)**	*Finn*
10:05 AM	**Peter Stone (University of Texas at Austin)**	*Stone*
11:00 AM	**Paul Verschure (Catalan Institute of Advanced Research)**	
11:30 AM	**Tobi Delbrück (University of Zurich and ETH Zurich)**	*Delbruck*
12:00 PM	**Moritz Grosse-Wentrup (Max Planck Institute Tuebingen)**	*Grosse-Wentrup*
12:30 PM	**Kristian Kersting (Technische Universität Dortmund)**	
02:00 PM	**Emo Todorov (University of Washington)**	*Todorov*
02:30 PM	**Richard Sutton (University of Alberta)**	*Sutton*
03:30 PM	**Bert Kappen (Radboud University)**	*Kappen*
04:00 PM	**Jean-Pascal Pfister (University of Zurich and ETH Zurich)**	
05:00 PM	**Jan Babic (Josef Stefan Institute Ljubijana)**	*Babic*
05:30 PM	**Martin Giese (University Clinic Tübingen)**	
06:00 PM	**Panel Discussion, Session Two and Session Three**	

Dec. 10, 2016

Bayesian Deep Learning

Yarin Gal, Christos Louizos, Zoubin Ghahramani, Kevin P Murphy, Max Welling

Area 1, Sat Dec 10, 08:00 AM

While deep learning has been revolutionary for machine learning, most modern deep learning models cannot represent their uncertainty nor take advantage of the well studied tools of probability theory. This has started to change following recent developments of tools and techniques combining Bayesian approaches with deep learning. The intersection of the two fields has received great interest from the community over the past few years, with the introduction of new deep learning models that take advantage of Bayesian techniques, as well as Bayesian models that incorporate deep learning elements.

In fact, the use of Bayesian techniques in deep learning can be traced back to the 1990s', in seminal works by Radford Neal, David MacKay, and Dayan et al.. These gave us tools to reason about deep models confidence, and achieved state-of-the-art performance on many tasks. However earlier tools did not adapt when new needs arose (such as scalability to big data), and were consequently forgotten. Such ideas are now being revisited in light of new advances in the field, yielding many exciting new results.

This workshop will study the advantages and disadvantages of such ideas, and will be a platform to host the recent flourish of ideas using Bayesian approaches in deep learning and using deep learning tools in Bayesian modelling. The program will include a mix of invited talks, contributed talks, and contributed posters. Also, the historic context of key developments in the field will be explained in an invited talk, followed by a tribute talk to David MacKay's work in the field. Future directions for the field will be debated in a panel discussion.

Schedule

08:30 AM	**BNNs for RL: A Success Story and Open Questions**	*Doshi-Velez*
08:55 AM	**Categorical Reparameterization with Gumbel-Softmax**	*Jang*
09:10 AM	**History of Bayesian neural networks**	*Ghahramani*
09:40 AM	**Poster spotlights**	
09:55 AM	**Discussion over coffee and poster session I**	
10:55 AM	**Deep exponential families**	*Blei*
11:20 AM	**Relativistic Monte Carlo**	*LU*
11:35 AM	**Alpha divergence minimization for Bayesian deep learning**	*Hernández-Lobato*
12:00 PM	**Lunch**	
01:30 PM	**A Tribute to David MacKay**	*Adams*
02:00 PM	**Adversarial Approaches to Bayesian Learning and Bayesian Approaches to Adversarial Robustness**	*Goodfellow*
02:25 PM	**Learning to Draw Samples: With Application to Amortized MLE for Generative Adversarial Training**	
02:40 PM	**Discussion over coffee and poster session II**	
03:35 PM	**Bayesian Agents: Bayesian Reasoning and Deep Learning in Agent-based Systems**	*Mohamed*
04:00 PM	**Panel Discussion**	*Mohamed, Blei, Adams, Hernández-Lobato, Goodfellow, Gal*
05:00 PM	**Discussion over coffee and poster session III**	

Optimizing the Optimizers

Maren Mahsereci, Alex J Davies, Philipp Hennig

Area 2, Sat Dec 10, 08:00 AM

http://www.probabilistic-numerics.org/meetings/NIPS2016/

Optimization problems in machine learning have aspects that make them more challenging than the traditional settings, like stochasticity, and parameters with side-effects (e.g., the batch size and structure). The field has invented many different approaches to deal with these demands. Unfortunately - and intriguingly - this extra functionality seems to invariably necessitate the introduction of tuning parameters: step sizes, decay rates, cycle lengths, batch sampling distributions, and so on. Such parameters are not present, or at least not as prominent, in classic optimization methods. But getting them right is frequently crucial, and necessitates inconvenient human “babysitting”.

Recent work has increasingly tried to eliminate such fiddle factors, typically by statistical estimation. This also includes automatic selection of external parameters like the batch-size or -structure, which have not traditionally been treated as part of the optimization task. Several different strategies have now been proposed, but they are not always compatible with each other, and lack a common framework that would foster both conceptual and algorithmic interoperability. This workshop aims to provide a forum for the nascent community studying automating parameter-tuning in optimization routines.

Among the questions to be addressed by the workshop are:

* Is the prominence of tuning parameters a fundamental feature of stochastic optimization problems? Why do classic optimization methods manage to do well with virtually no free parameters?
* In which precise sense can the "optimization of optimization algorithms" be phrased as an inference / learning problem?
* Should, and can, parameters be inferred at design-time (by a human), at compile-time (by an external compiler with access to a meta-description of the problem) or run-time (by the algorithm itself)?
* What are generic ways to learn parameters of algorithms, and inherent difficulties for doing so? Is the goal to specialize to a particular problem, or to generalize over many problems?

In addition to the invited and already confirmed speakers, we will also invite contributed work from the community. Topics of interest include, but are not strictly limited to,

* Parameter adaptation for optimization algorithms
* Stochastic optimization methods
* Optimization methods adapted for specific applications
* Batch selection methods
* Convergence diagnostics for optimization algorithms

Schedule

09:00 AM	**opening remarks and introduction**	
09:10 AM	**Matt Hoffman (DeepMind)**	
09:30 AM	**David Duvenaud (U of Toronto)**	
10:00 AM	**Stephen J Wright (U of Wisconsin)**	
10:30 AM	***coffee break***	
11:00 AM	**Samantha Hansen (Spotify)**	
11:30 AM	**spotlights**	
12:00 PM	**poster session**	
12:45 PM	***lunch break***	
02:15 PM	**Matteo Pirotta (Politecnico di Milano)**	
02:40 PM	**Ameet Talwalker (UCLA)**	
03:00 PM	***coffe break***	
03:30 PM	**Ali Rahimi (Google)**	
03:50 PM	**Mark Schmidt (UBC)**	
04:20 PM	***panel discussion***	

Deep Learning for Action and Interaction

Chelsea Finn, Raia Hadsell, David Held, Sergey Levine, Percy S Liang

Area 3, Sat Dec 10, 08:00 AM

Deep learning systems that act in and interact with an environment must reason about how actions will change the world around them. The natural regime for such real-world decision problems involves supervision that is weak, delayed, or entirely absent, and the outputs are typically in the context of sequential decision processes, where each decision affects the next input. This regime poses a challenge for deep learning algorithms, which typically excel with: (1) large amounts of strongly supervised data and (2) a stationary distribution of independently observed inputs. The algorithmic tools for tackling these challenges have traditionally come from reinforcement learning, optimal control, and planning, and indeed the intersection of reinforcement learning and deep learning is currently an exciting and active research area. At the same time, deep learning methods for interactive decision-making domains have also been proposed in computer vision, robotics, and natural language processing, often using different tools and algorithmic formalisms from classical reinforcement learning, such as direct supervised learning, imitation learning, and model-based control. The aim of this workshop will be to bring together researchers across these disparate fields. The workshop program will focus on both the algorithmic and theoretical foundations of decision making and interaction with deep learning, and the practical challenges associated with bringing to bear deep learning methods in interactive settings, such as robotics, autonomous vehicles, and interactive agents.

Schedule

09:00 AM	**Introductions**	
09:15 AM	**Joelle Pineau**	*Pineau*
09:40 AM	**Honglak Lee**	
10:05 AM	**Chris Summerfield**	
10:30 AM	**Morning coffee break**	
11:00 AM	**Jianxiong Xiao**	*Xiao*
11:25 AM	**Morning contributor spotlights**	
11:45 AM	**Morning poster session**	
12:00 PM	**Lunch break**	
02:00 PM	**Abhinav Gupta**	*Gupta*
02:25 PM	**Afternoon contributor spotlights**	
02:45 PM	**Afternoon poster session**	
03:00 PM	**Afternoon coffee break**	
03:30 PM	**Tim Lillicrap**	*Lillicrap*
03:55 PM	**Raquel Urtasun**	*Urtasun*
04:20 PM	**Jason Weston**	*Weston*
04:45 PM	**Contributor “pitch” session**	
05:30 PM	**Panel and audience discussion**	

Learning with Tensors: Why Now and How?

Anima Anandkumar, Rong Ge, Yan Liu, Maximilian Nickel, Rose Yu

Area 5 + 6, Sat Dec 10, 08:00 AM

Real world data in many domains is multimodal and heterogeneous, such as healthcare, social media, and climate science. Tensors, as generalizations of vectors and matrices, provide a natural and scalable framework for handling data with inherent structures and complex dependencies. Recent renaissance of tensor methods in machine learning ranges from academic research on scalable algorithms for tensor operations, novel models through tensor representations, to industry solutions including Google TensorFlow and Tensor Processing Unit (TPU). In particular, scalable tensor methods have attracted considerable amount of attention, with successes in a series of learning tasks, such as learning latent variable models [Anandkumar et al., 2014; Huang et al., 2015, Ge et al., 2015], relational learning [Nickle et al., 2011, 2014, 2016], spatio-temporal forecasting [Yu et al., 2014, 2015, 2016] and training deep neural networks [Alexander et al., 2015].

These progresses trigger new directions and problems towards tensor methods in machine learning. The workshop aims to foster discussion, discovery, and dissemination of research activities and outcomes in this area and encourages breakthroughs. We will bring together researchers in theories and applications who are interested in tensors analysis and development of tensor-based algorithms. We will also invite researchers from related areas, such as numerical linear algebra, high-performance computing, deep learning, statistics, data analysis, and many others, to contribute to this workshop. We believe that this workshop can foster new directions, closer collaborations and novel applications. We also expect a deeper conversation regarding why learning with tensors at current stage is important, where it is useful, what tensor computation softwares and hardwares work well in practice and, how we can progress further with interesting research directions and open problems.

Schedule

08:30 AM	**Opening Remarks**
08:40 AM	**On Depth Efficiency of Convolutional Networks: the use of Hierarchical Tensor Decomposition for Network Design and Analysis**
09:20 AM	**Contributed Talks**
10:00 AM	**Poster Spotlight 1**
10:30 AM	**Coffee Break and Poster Session 1**
11:00 AM	**Tensor Network Ranks**
11:40 AM	**Computational Phenotyping using Tensor Factorization**
11:40 AM	**Keynote Speech by Jimeng Sun**
12:20 PM	**Lunch**
02:00 PM	**Orthogonalized Alternating Least Squares: A theoretically principled tensor factorization algorithm for practical use**
02:40 PM	**Poster Spotlight**
02:40 PM	**Poster Spotlight 2**
03:00 PM	**Coffee Break and Poster Session**
03:00 PM	**Coffee Break and Poster Session 2**
03:30 PM	**Tensor decompositions for big multi-aspect data analytics**
04:10 PM	**PhD Symposium**
05:00 PM	**Panel Discussion and Closing Remarks**

Abstracts (3):

Abstract 2: **On Depth Efficiency of Convolutional Networks: the use of Hierarchical Tensor Decomposition for Network Design and Analysis in Learning with Tensors: Why Now and How?**, 08:40 AM

Our formal understanding of the inductive bias that drives the success of deep convolutional networks on computer vision tasks is limited. In particular, it is unclear what makes hypotheses spaces born from convolution and pooling operations so suitable for natural images.
I will present recent work that derive an equivalence between convolutional networks and hierarchical tensor decompositions. Under this equivalence, the structure of a network corresponds to the type of decomposition, and the network weights correspond to the decomposition parameters. This allows analyzing hypotheses spaces of networks by studying tensor spaces of corresponding decompositions, facilitating the use of algebraic and measure theoretical tools.
Specifically, the results I will present include showing how exponential depth efficiency is achieved in a family of deep networks called Convolutional Arithmetic Circuits, show that CAC is equivalent to SimNets, show that depth efficiency is superior to conventional ConvNets and show how inductive bias is tied to correlations between regions of the input image. In particular,
correlations are formalized through the notion of separation rank, which for a given input partition, measures how far a function is from being separable.
I will show that a polynomially sized deep network supports exponentially high separation ranks for certain input partitions, while being limited to polynomial separation ranks for others.
The network's pooling geometry effectively determines which input partitions are favored, thus serves as a means for controlling the inductive bias.
Contiguous pooling windows as commonly employed in practice favor interleaved partitions over coarse ones, orienting the inductive bias towards the statistics of natural images.
In addition to analyzing deep networks, I will show that shallow ones support only linear separation ranks, and by this gain insight into the benefit of functions brought forth by depth -- they are able to efficiently

model strong correlation under favored partitions of the input.

This work covers material recently presented in COLT, ICML and CVPR including recent Arxiv submissions. The work was jointly done with doctoral students Nadav Cohen and Or Sharir.

Abstract 10: **Orthogonalized Alternating Least Squares: A theoretically principled tensor factorization algorithm for practical use in Learning with Tensors: Why Now and How?**, 02:00 PM

From a theoretical perspective, low-rank tensor factorization is an algorithmic miracle, allowing for (provably correct) reconstruction and learning in a number of settings. From a practical standpoint, we still lack sufficiently robust, versatile, and efficient tensor factorization algorithms, particularly for large-scale problems. Many of the algorithms with provable guarantees either suffer from an expensive initialization step, and require the iterative removal of rank-1 factors, destroying any sparsity that might be present in the original tensor. On the other hand, the most commonly used algorithm in practice is "alternating least squares" [ALS], which iteratively fixes all but one mode, and optimizes the remaining mode. This algorithm is extremely efficient, but often converges to bad local optima, particularly when the weights of the factors are non-uniform. In this work, we propose a modification of the ALS approach that enjoys practically viable efficiency, as well as provable recovery (assuming the factors are random or have small pairwise inner products) even for highly non-uniform weights. We demonstrate the significant superiority of our recovery algorithm over the traditional ALS on both random synthetic data, and on computing word embeddings from a third-order word tri-occurrence tensor.

This is based on joint work with Vatsal Sharan.

Abstract 15: **Tensor decompositions for big multi-aspect data analytics in Learning with Tensors: Why Now and How?**, 03:30 PM

Tensors and tensor decompositions have been very popular and effective tools for analyzing
multi-aspect data in a wide variety of fields, ranging from Psychology to Chemometrics,
and from Signal Processing to Data Mining and Machine Learning.

Using tensors in the era of big data poses the challenge of scalability and efficiency.
In this talk, I will discuss recent techniques on tackling this challenge by parallelizing and speeding up tensor decompositions, especially for very sparse datasets (such as the ones encountered for example in online social network analysis).

In addition to scalability, I will also touch upon the challenge of unsupervised quality assessment, where in absence of ground truth, we seek
to automatically select the decomposition model that captures best the structure
in our data.

The talk will conclude with a discussion on future research directions and open problems in tensors for big data analytics.

Continual Learning and Deep Networks

Razvan Pascanu, Mark Ring, Tom Schaul

Area 7 + 8, Sat Dec 10, 08:00 AM

Humans have the extraordinary ability to learn continually from experience. Not only can we apply previously learned knowledge and skills to new situations, we can also use these as the foundation for later learning. One of the grand goals of AI is building an artificial "continual learning" agent that constructs a sophisticated understanding of the world from its own experience, through the autonomous incremental development of ever more complex skills and knowledge.

Hallmarks of continual learning include: interactive, incremental, online learning (learning occurs at every moment, with no fixed tasks or data sets); hierarchy or compositionality (previous learning can become the foundation far later learning); "isolaminar" construction (the same algorithm is used at all stages of learning); resistance to catastrophic forgetting (new learning does not destroy old learning); and unlimited temporal abstraction (both knowledge and skills may refer to or span arbitrary periods of time).

Continual learning is an unsolved problem which presents particular difficulties for the deep-architecture approach that is currently the favored workhorse for many applications. Some strides have been made recently, and many diverse research groups have continual learning on their road map. Hence we believe this is an opportune moment for a workshop focusing on this theme. The goals would be to define the different facets of the continual-learning problem, to tease out the relationships between different relevant fields (such as reinforcement learning, deep learning, lifelong learning, transfer learning, developmental learning, computational neuroscience, etc.) and to propose and explore promising new research directions.

Schedule

08:30 AM	**Introduction to the workshop**
08:50 AM	**Invited talk - Richard Sutton**
09:20 AM	**Spotlight #1**
09:30 AM	**Spotlight #2**
09:40 AM	**Invited talk - Claudia Clopath**
10:40 AM	**Invited talk - Satinder Singh Baveja**
11:10 AM	**Spotlight #3**
11:20 AM	**Posters**
12:20 PM	**Lunch**
02:00 PM	**Invited talk - Honglak Lee**
02:30 PM	**Spotlight #4**
02:40 PM	**Spotlight #5**
02:50 PM	**Invited talk - Eric Eaton**
03:20 PM	**DARPA funding opportunities - Hava Siegelmann**

03:30 PM	**Coffee Break**
04:00 PM	**Invited talk - Raia Hasdell**
04:30 PM	**Invited talk - Doina Precup**
05:00 PM	**Panel (all invited speakers)**

End-to-end Learning for Speech and Audio Processing

John Hershey, Philemon Brakel

Hilton Diag. Mar, Blrm. A, Sat Dec 10, 08:00 AM

This workshop focuses on recent advances to end-to-end methods for speech and more general audio processing. Deep learning has transformed the state of the art in speech recognition, and audio analysis in general. In recent developments, new deep learning architectures have made it possible to integrate the entire inference process into an end-to-end system. This involves solving problems of an algorithmic nature, such as search over time alignments between different domains, and dynamic tracking of changing input conditions. Topics include automatic speech recognition systems (ASR) and other audio procssing systems that subsume front-end adaptive microphone array processing and source separation as well as back-end constructs such as phonetic context dependency, dynamic time alignment, or phoneme to grapheme modeling. Other end-to-end audio applications include speaker diarization, source separation, and music transcription. A variety of architectures have been proposed for such systems, ranging from shift-invariant convolutional pooling to connectionist temporal classification (CTC) and attention based mechanisms, or other novel dynamic components. However there has been little comparison yet in the literature of the relative merits of the different approaches. This workshop delves into questions about how different approaches handle various trade-offs in terms of modularity and integration, in terms of representation and generalization. This is an exciting new area and we expect significant interest from the machine learning and speech and audio processing communities.

Schedule

09:30 AM	**Jan Chorowski**
10:30 AM	**Li Deng**
11:00 AM	**Coffee Break**
11:00 AM	**Andrew Maas**
11:30 AM	**Florian Metze**
11:30 AM	**Florian Metze: End-to-end learning for language universal speech recognition**
12:00 PM	**Lunch**
02:00 PM	**Tara Sainath**
02:30 PM	**Oriol Vinyals**
03:30 PM	**Discussion Panel**
04:30 PM	**Spotlights and Posters**

Abstracts (1):

Abstract 6: **Florian Metze: End-to-end learning for language universal speech recognition in End-to-end Learning for Speech and Audio Processing**, 11:30 AM

One of the great successes of end-to-end learning strategies such as Connectionist Temporal Classification in automatic speech recognition is the ability to train very powerful models that map directly from features to characters or context independent phones. Traditional hybrid models, or even GMMs usually require context dependent states and a Hidden Markov Model in order to achieve good performance. By contrast, with CTC, it thus becomes possible to train a multi-lingual RNN that can directly predict phones in multiple languages (multi-task training), language independent articulatory features, and language universal phones, allowing for the recognition of speech in languages for which no acoustic training data is available.

Machine Learning for Spatiotemporal Forecasting

Florin Popescu, Sergio Escalera, Xavier Baró, Stephane Ayache, Isabelle Guyon

Hilton Diag. Mar, Blrm. B, Sat Dec 10, 08:00 AM

A crucial, high impact application of learning systems is forecasting. While machine learning has already been applied to time series analysis and signal processing, the recent big data revolution allows processing and prediction of vast data flows and forecasting of high dimensional, spatiotemporal series using massive multi-modal streams as predictors. Wider data bandwidths allow machine learning techniques such as connectionist and deep learning methods to assist traditional forecasting methods from fields such as engineering and econometrics, while probabilistic methods are uniquely suited to address the stochastic nature of many processes requiring forecasting.

This workshop will bring together multi-disciplinary researchers from signal processing, statistics, machine learning, computer vision, economics and causality looking to widen their application or methodological scope. It will begin by providing a forum to discuss pressing application areas o forecasting: video compression and understanding, energy and and smart grid management, economics and finance, environmental and health policy (e.g. epidemiology), as well as introduce challenging new datasets. A large dataset, created for an industry-driven data competition, will be presented - this dataset not only helps develop and compare new methods for forecasting, but also addresses deeper underlying learning theory questions: do effective learning systems truly infer underlying structure or merely output accuracy in data streams?, is transfer learning available at no loss to specificity? and is semi-supervised learning is an inherent property of powerful, accurate, learning machines? What strategies are scalable so they perform well on sparse as well as big data? What exactly is a good forecasting machine? Therefore a forum is also planned to discuss such pressing issues,- dedicated poster sessions and panels are scheduled. We plan for a varied list of reknowned speakers, presenting data sources, rich open-source platforms for forecasting, prediction performance evaluation metrics, past forecasting competitions and state-of-the-art methods.

Schedule

09:00 AM	**Welcome and introduction to spatiotemporal forecasting: platforms, tools, datasets and challenges. Florin Popescu (Fraunhofer Institute, Germany).**
09:30 AM	**Forecasting for Electrical Transmission Grid: Antoine Marot, (RTE: Réseau de transport d'électricité, FR).**
10:00 AM	**Forecasting using machine learning techniques in energy and agriculture. Danny Silver (Acadia University, CA)**
10:00 AM	**Break**
11:00 AM	**Application areas of advanced forecasting methods.**
12:00 PM	**Lunch and Poster Session**
02:30 PM	**Financial Risk Forecasting: Alexander Statnikov (American Express)**
03:00 PM	**Cofee Break**
03:30 PM	**Platforms for forecasting**
04:30 PM	**From Hype-Cycle to Reality of Predictive Analytics (a Time Series Forecasting perspective): Sven Crone (Lancaster Centre for Forecasting, Lancaster Univ.)**
05:00 PM	**Comparing forecasting methods: how?**
05:00 PM	**Spatiotemporal Online Learning with Expert Advice and Applications in Climate Science and Finance: Claire Monteleoni (George Washington University)**
05:45 PM	**Final discusison (mod. Program Committe)**

Let's Discuss: Learning Methods for Dialogue

Hal Daume III, Paul Mineiro, Amanda Stent, Jason E Weston

Hilton Diag. Mar, Blrm. C, Sat Dec 10, 08:00 AM

Humans conversing naturally with machines is a staple of science fiction. Building agents capable of mutually coordinating their states and actions via communication, in conjunction with human agents, would be one of the Average engineering feats of human history. In addition to the tremendous economic potential of this technology, the ability to converse appears intimately related to the overall goal of AI.

Although dialogue has been an active area within the linguistics and NLP communities for decades, the wave of optimism in the machine learning community has inspired increased interest from researchers, companies, and foundations. The NLP community has enthusiastically embraced and innovated neural information processing systems, resulting in substantial relevant activity published outside of NIPS. A forum for increased interaction (dialogue!) with these communities at NIPS will accelerate creativity and progress.

We plan to focus on the following issues:

1. How to be data-driven
a. What are tractable and useful intermediate tasks on the path to truly conversant machines? How can we leverage existing benchmark tasks and competitions? What design criteria would we like to see for the next set of benchmark tasks and competitions?
b. How do we assess performance? What can and cannot be done with offline evaluation on fixed data sets? How can we facilitate development of these offline evaluation tasks in the public domain? What is the role of online evaluation as a benchmark, and how would we make it accessible to the general community? Is there a role for simulated environments, or tasks where machines communicate solely with each other?
2. How to build applications
a. What unexpected problem aspects arise in situated systems? human-hybrid systems? systems learning from adversarial inputs?
b. Can we divide and conquer? Do we need to a irreducible end-to-end system, or can we define modules with abstractions that do not leak?
c. How do we ease the burden on the human designer of specifying or bootstrapping the system?
3. Architectural and algorithmic innovation
a. What are the associated requisite capabilities for learning architectures, and where are the deficiencies in our current architectures? How can we leverage recent advances in reasoning, attention, and memory architectures? How can we beneficially incorporate linguistic knowledge into our architectures?
b. How far can we get with current optimization techniques? To learn requisite competencies, do we need advances in discrete optimization? curriculum learning? (inverse) reinforcement learning?

Schedule

08:20 AM	**Opening**	
08:25 AM	**Building Complete Systems**	
08:30 AM	**Evolvable Dialogue Systems**	*Gasic*
09:10 AM	**The Missing Pieces for a Full-Fledged Dialog Agent**	*Marco*
09:50 AM	**Authoring End-to-End Dialog Systems**	*Williams*
10:30 AM	**Break**	

11:00 AM	**Panel Session 1**	
11:20 AM	**Lunch**	
12:55 PM	**Leveraging Linguistics**	
01:00 PM	**Coordination and Learning in Human Dialogue**	*Fernández*
01:40 PM	**Domain Adaptation using Linguistic Knowledge**	*Dethlefs*
02:20 PM	**Bootstrapping Incremental Dialogue Systems: Using Linguistic Knowledge to Learn from Minimal Data**	*Kalatzis, Eshghi*
02:40 PM	**Multi-Agent Communication and the Emergence of (Natural) Language**	*Lazaridou*
03:00 PM	**Break 2**	
03:30 PM	**Panel Session 2**	
03:45 PM	**Modeling Techniques**	
03:50 PM	**Evaluating End-to-End Goal Oriented Dialog Systems**	*Bordes*
04:30 PM	**Awkward Silence? The evaluation of non-goal oriented dialogue systems**	*Hastie*
05:10 PM	**Learning Goal-oriented Dialog using Gated End-to-End Memory Networks**	*Perez*
05:30 PM	**Generative Deep Neural Networks for Dialogue: A Short Review**	*Serban*
05:50 PM	**Mini Break**	
06:00 PM	**Panel Session 3**	
06:20 PM	**La Fin**	

Abstracts (9):

Abstract 2: **Building Complete Systems in Let's Discuss: Learning Methods for Dialogue**, 08:25 AM

This set of talks and panel session is organized around the theme of building end-to-end dialog systems.

Abstract 6: **Break in Let's Discuss: Learning Methods for Dialogue**, 10:30 AM

Workshop coffee break.

Abstract 9: **Leveraging Linguistics in Let's Discuss: Learning Methods for Dialogue**, 12:55 PM

This set of talks and panel session is organized around the theme of leveraging linguistics to build, improve, and understand dialog systems.

Abstract 14: **Break 2 in Let's Discuss: Learning Methods for Dialogue**, 03:00 PM

Workshop coffee break.

Abstract 16: **Modeling Techniques in Let's Discuss: Learning Methods for Dialogue**, 03:45 PM

This set of talks and panel session is organized around the theme of modeling dialogue using machine learning techniques: in particular, what architectures to use, and how to evaluate performance.

Abstract 17: **Evaluating End-to-End Goal Oriented Dialog Systems in Let's Discuss: Learning Methods for Dialogue**, *Bordes* 03:50 PM

Traditional dialog systems used in goal-oriented applications require a lot of domain-specific handcrafting, which hinders scaling up to new domains. End- to-end dialog systems, in which all components are trained from the dialogs themselves, escape this limitation. But the encouraging successes recently obtained in chit-chat dialog may not carry over to goal-oriented settings. In this talk, we will discuss how to evaluate end-to-end goal oriented dialog systems in a robust and reproducible manner. We will also present a new testbed designed to that end. On this new dataset, we show that an end-to-end dialog system based on Memory Networks can reach promising, yet imperfect, performance and learn to perform non-trivial operations. We confirm those results by comparing our system to a hand-crafted slot-filling baseline on data from the second Dialog State Tracking Challenge (Henderson et al., 2014a) and show similar result patterns on data extracted from an online concierge service.

Abstract 18: **Awkward Silence? The evaluation of non-goal oriented dialogue systems in Let's Discuss: Learning Methods for Dialogue**, *Hastie* 04:30 PM

Non-goal orientated dialogue systems can provide users with key information, support decision making, facilitate user action or simply chat for the sake of having some company. The key question is how do we define what it means to be an engaging conversationalist and how can we measure this automatically? This talk will take a multidisciplinary approach to evaluating data-driven, non-goal orientated dialogue systems, taking insights from the gaming industry, HCI, psychology and cognitive theory.

Abstract 19: **Learning Goal-oriented Dialog using Gated End-to-End Memory Networks in Let's Discuss: Learning Methods for Dialogue**, *Perez* 05:10 PM

In this paper, we introduce a novel memory network model using an end-to-end differentiable memory access regulation mechanism. It is inspired by the current progress on the connection short-cutting principle in the field of computer vision. We name it Gated End-to-End Memory Network (GMemN2N). From the machine learning perspective, this new capability is learned in an end-to-end fashion without the use of any additional supervision signal which is, as far as our knowledge goes, the first of its kind. Our experiments show improvements on all of the Dialog bAbI tasks, particularly on the real human-bot conversion-based Dialog State Tracking Challenge (DSTC2) dataset. This method does not require the use of any domain knowledge. Our model sets a new state of the art of end-to-end trainable dialog systems on this dataset.

Abstract 20: **Generative Deep Neural Networks for Dialogue: A Short Review in Let's Discuss: Learning Methods for Dialogue**, *Serban* 05:30 PM

Researchers have recently started investigating deep neural networks for dialogue applications. In particular, generative sequence-to-sequence (Seq2Seq) models have shown promising results for unstructured tasks, such as word-level dialogue response generation. The hope is that such models will be able to leverage massive amounts of data to learn meaningful natural language representations and response generation strategies, while requiring a minimum amount of domain knowledge and hand-crafting. We review recently proposed models based on generative encoder-decoder neural network architectures, and show that these models have better ability to incorporate long-term dialogue history, to model uncertainty and ambiguity in dialogue, and to generate responses with high-level compositional structure.

Large Scale Computer Vision Systems

Manohar Paluri, Lorenzo Torresani, Gal Chechik, Dario Garcia, Du Tran

Room 111, Sat Dec 10, 08:00 AM

Computer Vision is a mature field with long history of academic research, but recent advances in deep learning provided machine learning models with new capabilities to understand visual content. There have been tremendous improvements on problems like classification, detection, segmentation, which are basic proxies for the ability of a model to understand the visual content. These are accompanied by a steep rise of Computer Vision adoption in industry at scale, and by more complex tasks such as Image Captioning and Visual Q&A. These go well beyond the classical problems and open the doors to a whole new world of possibilities. As industrial applications mature, the challenges slowly shift towards challenges in data, in scale, and in moving from purely visual data to multi-modal data.

The unprecedented adoption of Computer Vision to numerous real world applications processing billions of "live" media content daily, raises a new set of challenges, including:

1. Efficient Data Collection (Smart sampling, weak annotations, ...)
2. Evaluating performance in the wild (long tails, embarrassing mistakes, calibration)
3. Incremental learning: Evolve systems incrementally in complex environments (new data, new categories, federated architectures ...)
4. Handling tradeoffs: Computation vs Accuracy vs Supervision
5. Outputs are various types (Binary predictions, embeddings etc.)
6. Machine learning feedback loops
7. Minimizing technical debt as system matures
8. On-device vs On-cloud vs Split
9. Multi-modal content understanding

We will bring together researchers and practitioners who are interested to address this new set of challenges and provide a venue to share how industry and academia approach these problems. We will invite prominent speakers from academia and industry to give their perspectives on these challenges. In addition, we will have 5 minute spotlights for selected papers submitted to the workshop and a poster session for all selected submissions. The topics of submissions should be related to the above mentioned list of challenges. We will end the session with a panel discussion including the speakers on the future of large scale vision and its applications in the wild.

In the second part we will looke at how specifically this applies to video understanding. Video understanding aims at developing computer methods that can interpret videos at different semantic levels. Applications include video categorization, event detection, semantic segmentation, description, summarization, tagging, content-based retrieval, surveillance, and many more. Although in the last two decades the field of video analytics has witnessed significant progress, most problems in this area still remain largely unsolved. In recent years video understanding has become an even more critical and timely problem to address because of the tremendous growth of videos on the Internet, most of which do not contain tags or descriptions and thus necessitate automatic analysis to become searchable or browsable. At the same time the rise of online video repositories represents an opportunity for the creation of new pivotal large-scale datasets for research in this area. Given the recent breakthroughs achieved by deep learning in other big data domains, we believe that video understanding may very well be on the verge of a technical revolution that will spur significant advances in this area.

In order to foster further progress by the research community, we propose to organize a one-day workshop to discuss emerging innovations and ideas about the problems and challenges related to video understanding. The workshop will consist of a series of invited talks from researchers in this area. In addition, we will publicly announce and present a new large-scale benchmark for video comprehension [1] that has the potential to become an instrumental resource for future research in this field. Compared to existing video datasets, our proposed benchmark has much bigger scale and it casts video understanding in the novel form of multiple choice tests that assess the ability of the algorithm to comprehend the semantics of the video.

This workshop will be the first of a series of annual meetings that we will organize to stimulate steady progress in this area. In each subsequent edition of this workshop, we will host an annual challenge on our continuously expanding video comprehension benchmark in order to motivate students and researchers to push the envelope on this problem. We hope to bring together researchers with common interests in video analysis to share, learn, and make good progress toward better video understanding methods.

[1] D. Tran, M. Paluri, and L. Torresani, "ViCom: Benchmark and Methods for Video Comprehension," CoRR, abs/1606.07373, July 2016, http://arxiv.org/abs/1606.07373

Schedule

09:00 AM	**Invited Talk 1**	*Torralba*
09:30 AM	**Invited Talk 2**	*Schmid*
10:00 AM	**CV @ Scale Challenges**	*Garcia, Paluri, Chechik*
11:00 AM	**ViCom: Benchmark and Methods for Video Comprehension**	*Tran, Bolonkin, Paluri, Torresani*

11:30 AM	**Knowledge Acquisition for Visual Question Answering via Iterative Querying**	*Zhu, Lim, Li*
11:45 AM	**Tag Prediction at Flickr: a View from the Darkroom**	*Boakye*
02:00 PM	**Invited Talk 3**	*Urtasun*
02:30 PM	**What makes ImageNet good for Transfer Learning?**	*Huh, Agrawal, Efros*
02:45 PM	**PororoQA: Cartoon Video Series Dataset for Story Understanding**	*Kim, Heo, Zhang*
03:00 PM	**Poster Presentations**	
04:00 PM	**Invited Talk 4**	*Gupta*
04:30 PM	**Invited Talk 5**	*Wang*

OPT 2016: Optimization for Machine Learning

Suvrit Sra, Francis Bach, Sashank J. Reddi, Niao He

Room 112, Sat Dec 10, 08:00 AM

As the ninth in its series, OPT 2016 builds on remarkable precedent established by the highly successful series of workshops: OPT 2008--OPT 2015, which have been instrumental in bridging the OPT and ML communities closer together.

The previous OPT workshops enjoyed packed to overpacked attendance. This huge interest is no surprise: optimization is the 2nd largest topic at NIPS and is indeed foundational for the wider ML community.

Looking back over the past decade, a strong trend is apparent: The intersection of OPT and ML has grown monotonically to the point that now several cutting-edge advances in optimization arise from the ML community. The distinctive feature of optimization within ML is its departure from textbook approaches, in particular, by having a different set of goals driven by “big-data,” where both models and practical implementation are crucial.

This intimate relation between OPT and ML is the core theme of our workshop. We wish to use OPT2016 as a platform to foster discussion, discovery, and dissemination of the state-of-the-art in optimization as relevant to machine learning. And even beyond that, as a platform to identify new directions and challenges that will drive future research.

How OPT differs from other related workshops:

Compared to the other optimization focused workshops that we are aware of, the distinguishing features of OPT are: (a) it provides a unique bridge between the ML community and the wider optimization community; (b) it encourages theoretical work on an equal footing with practical efficiency; and (c) it caters to a wide body of NIPS attendees, experts and beginners alike (some OPT talks are always of a more “tutorial” nature).

Extended abstract

The OPT workshops have previously covered a variety of topics, such as frameworks for convex programs (D. Bertsekas), the intersection of ML and optimization, classification (S. Wright), stochastic gradient and its tradeoffs (L. Bottou, N. Srebro), structured sparsity (Vandenberghe), randomized methods for convex optimization (A. Nemirovski), complexity theory of convex optimization (Y. Nesterov), distributed optimization (S. Boyd), asynchronous stochastic gradient (B. Recht), algebraic techniques (P. Parrilo), nonconvex optimization (A. Lewis), sums-of-squares techniques (J. Lasserre), deep learning tricks (Y. Bengio), stochastic convex optimization (G. Lan), new views on interior point (E. Hazan), among others.

Several ideas propounded in OPT have by now become important research topics in ML and optimization --- especially in the field of randomized algorithms, stochastic gradient and variance reduced stochastic gradient methods. An edited book "Optimization for Machine Learning" (S. Sra, S. Nowozin, and S. Wright; MIT Press, 2011) grew out of the first three OPT workshops, and contains high-quality contributions from many of the speakers and attendees, and there have been sustained requests for the next edition of such a volume.

Much of the recent focus has been on large-scale first-order convex optimization algorithms for machine learning, both from a theoretical and methodological point of view. Covered topics included stochastic gradient algorithms, (accelerated) proximal algorithms, decomposition and coordinate descent algorithms, parallel and distributed optimization. Theoretical and practical advances in these methods remain a topic of core interest to the workshop. Recent years have also seen interesting advances in non-convex optimization such as a growing body of results on alternating minimization, tensor factorization etc.

We also do not wish to ignore the not particularly large scale setting, where one does have time to wield substantial computational resources. In this setting, high-accuracy solutions and deep understanding of the lessons contained in the data are needed. Examples valuable to MLers may be exploration of genetic and environmental data to identify risk factors for disease; or problems dealing with setups where the amount of observed data is not huge, but the mathematical model is complex. Consequently, we encourage optimization methods on manifolds, ML problems with differential geometric antecedents, those using advanced algebraic techniques, and computational topology, for instance.

At this point, we would like to emphasize again that OPT2016 is one of the few optimization+ML workshops that lies at the intersection of theory and practice: both actual efficiency of algorithms in practice as well as their theoretical analysis are given equal value.

Schedule

08:15 AM	**Opening Remarks**
08:30 AM	**Invited Talk: Online Optimization, Smoothing, and Worst-case Competitive Ratio (Maryam Fazel, University of Washington)**

09:15 AM	**Spotlight: Markov Chain Lifting and Distributed ADMM**
09:30 AM	**Poster Session 1**
10:30 AM	**Coffee Break 1**
11:00 AM	**Invited Talk: Kernel-based Methods for Bandit Convex Optimization (Sébastien Bubeck, Microsoft Research)**
11:45 AM	**Spotlight: Frank-Wolfe Algorithms for Saddle Point Problems**
12:00 PM	**Lunch Break**
02:00 PM	**Invited Talk: Semidefinite Programs with a Dash of Smoothness: Why and When the Low-Rank Approach Works (Nicolas Boumal, Princeton University)**
02:45 PM	**Spotlight: QuickeNing: A Generic Quasi-Newton Algorithm for Faster Gradient-Based Optimization**
03:00 PM	**Coffee Break 2**
03:30 PM	**Poster Session 2**
04:30 PM	**Invited Talk: Oracle Complexity of Second-Order Methods for Finite-Sum Problems (Ohad Shamir, Weizmann Institute of Science)**
05:15 PM	**Spotlight: Reliably Learning the ReLU in Polynomial Time**

Abstracts (8):

Abstract 2: **Invited Talk: Online Optimization, Smoothing, and Worst-case Competitive Ratio (Maryam Fazel, University of Washington) in OPT 2016: Optimization for Machine Learning**, 08:30 AM

In Online Optimization, the data in an optimization problem is revealed over time and at each step a decision variable needs to be set without knowing the future data. This setup covers online resource allocation, from classical inventory problems to the Adwords problem popular in online advertising.
In this talk, we prove bounds on the competitive ratio of two primal-dual algorithms for a broad class of online convex optimization problems. We give a sufficient condition on the objective function that guarantees a constant worst-case competitive ratio for monotone functions. We show how smoothing the objective can improve the competitive ratio of these algorithms, and for separable functions, we show that the optimal smoothing can be derived by solving a convex optimization problem. This result allows us to directly optimize the competitive ratio bound over a class of smoothing functions, and hence design effective smoothing customized for a given cost function.

Abstract 3: **Spotlight: Markov Chain Lifting and Distributed ADMM in OPT 2016: Optimization for Machine Learning**, 09:15 AM

The time to converge to the steady state of a finite Markov chain can be greatly reduced by a lifting operation, which creates a new Markov chain on an expanded state space. For a class of quadratic objectives, we show an analogous behavior whereby a distributed ADMM algorithm can be seen as a lifting of Gradient Descent. This provides a deep insight for its faster convergence rate under optimal parameter tuning. We conjecture that this gain is always present, contrary to when lifting a Markov chain, where sometimes we only obtain a marginal speedup.

Abstract 6: **Invited Talk: Kernel-based Methods for Bandit Convex Optimization (Sébastien Bubeck, Microsoft Research) in OPT 2016: Optimization for Machine Learning**, 11:00 AM

A lot of progress has been made in recent years on extending classical multi-armed bandit strategies to very large set of actions. A particularly challenging setting is the one where the action set is continuous and the underlying cost function is convex, this is the so-called bandit convex optimization (BCO) problem. I will tell the story of BCO and explain some of the new ideas that we recently developed to solve it. I will focus on three new ideas from our recent work http://arxiv.org/abs/1607.03084 with Yin Tat Lee and Ronen Eldan: (i) a new connection between kernel methods and the popular multiplicative weights strategy; (ii) a new connection between kernel methods and one of Erdos' favorite mathematical object, the Bernoulli convolution, and (iii) a new adaptive (and increasing!) learning rate for multiplicative weights. These ideas could be of broader interest in learning/algorithm's design

Abstract 7: **Spotlight: Frank-Wolfe Algorithms for Saddle Point Problems in OPT 2016: Optimization for Machine Learning**, 11:45 AM

We extend the Frank-Wolfe (FW) optimization algorithm to solve constrained smooth convex-concave saddle point (SP) problems. Remarkably, the method only requires access to linear minimization oracles. Leveraging recent advances in FW optimization, we provide the first proof of convergence of a FW-type saddle point solver over polytopes, thereby partially answering a 30 year-old conjecture. We verify our convergence rates empirically and observe that by using a heuristic step-size, we can get empirical convergence under more general conditions, paving the way for future theoretical work.

Abstract 9: **Invited Talk: Semidefinite Programs with a Dash of Smoothness: Why and When the Low-Rank Approach Works (Nicolas Boumal, Princeton University) in OPT 2016: Optimization for Machine Learning**, 02:00 PM

Semidefinite programs (SDPs) can be solved in polynomial time by interior point methods, but scalability can be an issue. To address this shortcoming, over a decade ago, Burer and Monteiro proposed to solve SDPs with few equality constraints via low-rank, non-convex surrogates. Remarkably, for some applications, local optimization methods seem to converge to global optima of these non-convex surrogates reliably.

In this presentation, we show that the Burer-Monteiro formulation of SDPs in a certain class almost never has any spurious local optima, that is: the non-convexity of the low-rank formulation is benign (even saddles are strict). This class of SDPs covers applications such as max-cut, community detection in the stochastic block model, robust PCA, phase retrieval and synchronization of rotations.

The crucial assumption we make is that the low-rank problem lives on a manifold. Then, theory and algorithms from optimization on manifolds can be used.

Optimization on manifolds is about minimizing a cost function over a smooth manifold, such as spheres, low-rank matrices, orthonormal frames, rotations, etc. We will present the basic framework as well as parts of the more general convergence theory, including recent complexity results. (Toolbox: http://www.manopt.org)

Select parts are joint work with P.-A. Absil, A. Bandeira, C. Cartis and V. Voroninski.

Abstract 10: **Spotlight: QuickeNing: A Generic Quasi-Newton Algorithm for Faster Gradient-Based Optimization in OPT 2016: Optimization for Machine Learning**, 02:45 PM

We propose a technique to accelerate gradient-based optimization algorithms by giving them the ability to exploit L-BFGS heuristics. Our scheme is (i) generic and can be applied to a large class of first-order algorithms; (ii) it is compatible with composite objectives, meaning that it may provide exactly sparse solutions when a sparsity-inducing regularization is involved; (iii) it admits a linear convergence rate for strongly-convex problems; (iv) it is easy to use and it does not require any line search. Our work is inspired in part by the Catalyst meta-algorithm, which accelerates gradient-based techniques in the sense of Nesterov; here, we adopt a different strategy based on L-BFGS rules to learn and exploit the local curvature. In most practical cases, we observe significant improvements over Catalyst for solving large-scale high-dimensional machine learning problems.

Abstract 13: **Invited Talk: Oracle Complexity of Second-Order Methods for Finite-Sum Problems (Ohad Shamir, Weizmann Institute of Science) in OPT 2016: Optimization for Machine Learning**, 04:30 PM

Finite-sum optimization problems are ubiquitous in machine learning, and are commonly solved using first-order methods which rely on gradient computations. Recently, there has been growing interest in *second-order* methods, which rely on both gradients and Hessians. In principle, second-order methods can require much fewer iterations than first-order methods, and hold the promise for more efficient algorithms. Although computing and manipulating Hessians is prohibitive for high-dimensional problems in general, the Hessians of individual functions in finite-sum problems can often be efficiently computed, e.g. because they possess a low-rank structure. Can second-order information indeed be used to solve such problems more efficiently? In this talk, I'll provide evidence that the answer -- perhaps surprisingly -- is negative, at least in terms of worst-case guarantees. However, I'll also discuss what additional assumptions and algorithmic approaches might potentially circumvent this negative result.
Joint work with Yossi Arjevani.

Abstract 14: **Spotlight: Reliably Learning the ReLU in Polynomial Time in OPT 2016: Optimization for Machine Learning**, 05:15 PM

We give the first dimension-efficient algorithms for learning Rectified Linear Units (ReLUs), which are functions of the form max(0, w.x) with w a unit vector (2-norm equal to 1). Our algorithm works in the challenging Reliable Agnostic learning model of Kalai, Kanade, and Mansour where the learner is given access to a distribution D on labeled examples but the labeling may be arbitrary. We construct a hypothesis that simultaneously minimizes the false-positive rate and the l_p loss (for p=1 or p >=2) of inputs given positive labels by D.

It runs in polynomial-time (in n) with respect to {\em any} distribution on S^{n-1} (the unit sphere in n dimensions) and for any error parameter \epsilon = \Omega(1/ \log n). These results are in contrast to known efficient algorithms for reliably learning linear threshold functions, where epsilon must be Omega(1) and strong assumptions are required on the marginal distribution. We can compose our results to obtain the first set of efficient algorithms for learning constant-depth networks of ReLUs.

Our techniques combine kernel methods and polynomial approximations with a ``dual-loss'' approach to convex programming. As a byproduct we also obtain the first set of efficient algorithms for ``convex piecewise-linear fitting,'' and the first efficient algorithms for agnostically learning low-weight multivariate polynomials on the unit sphere.

Neural Abstract Machines & Program Induction

Matko Bošnjak, Nando de Freitas, Tejas D Kulkarni, Arvind Neelakantan, Scott E Reed, Sebastian Riedel, Tim Rocktäschel

Room 113, Sat Dec 10, 08:00 AM

Machine intelligence capable of learning complex procedural behavior, inducing (latent) programs, and reasoning with these programs is a key to solving artificial intelligence. The problems of learning procedural behavior and program induction have been studied from different perspectives in many computer science fields such as program synthesis, probabilistic programming, inductive logic programming, reinforcement learning, and recently in deep learning. However, despite the common goal, there seems to be little communication and collaboration between the different fields focused on this problem.

Recently, there have been a lot of success stories in the deep learning community related to learning neural networks capable of using trainable memory abstractions. This has led to the development of neural networks with differentiable data structures such as Neural Turing Machines, Memory Networks, Neural Stacks, and Hierarchical Attentive Memory, among others. Simultaneously, neural program induction models like Neural Program Interpreters and Neural Programmer have created a lot of excitement in the field, promising induction of algorithmic behavior, and enabling inclusion of programming languages in the processes of execution and induction, while staying end-to-end trainable. Trainable program induction models have the potential to make a substantial impact in many problems involving long-term memory, reasoning, and procedural execution, such as question answering, dialog, and robotics.

The aim of the NAMPI workshop is to bring researchers and practitioners from both academia and industry, in the areas of deep learning, program synthesis, probabilistic programming, inductive programming and reinforcement learning, together to exchange ideas on the future of program induction with a special focus on neural network models and abstract machines. Through this workshop we look to identify common challenges, exchange ideas among and lessons learned from the different fields, as well as establish a (set of) standard evaluation benchmark(s) for approaches that learn with abstraction and/or reason with induced programs.

Areas of interest for discussion and submissions include, but are not limited to (in alphabetical order):
- Applications
- Compositionality in Representation Learning
- Differentiable Memory
- Differentiable Data Structures
- Function and (sub-)Program Compositionality
- Inductive Logic Programming
- Knowledge Representation in Neural Abstract Structures
- Large-scale Program Induction
- Meta-Learning and Self-improving
- Neural Abstract Machines
- Program Induction: Datasets, Tasks, and Evaluation
- Program Synthesis
- Probabilistic Programming
- Reinforcement Learning for Program Induction
- Semantic Parsing

Schedule

08:50 AM	**Introduction**
09:00 AM	**Stephen Muggleton - What use is Abstraction in Deep Program Induction?**
09:30 AM	**Daniel Tarlow - In Search of Strong Generalization: Building Structured Models in the Age of Neural Networks**
10:00 AM	**Charles Sutton - Learning Program Representation: Symbols to Semantics**
10:30 AM	**Coffee Break**
11:00 AM	**Doina Precup - From temporal abstraction to programs**
11:30 AM	**Rob Fergus - Learning to Compose by Delegation**
12:00 PM	**Percy Liang - How Can We Write Large Programs without Thinking?**
12:30 PM	**Lunch**
02:00 PM	**Martin Vechev - Program Synthesis and Machine Learning**
02:30 PM	**Ed Grefenstette - Limitations of RNNs: a computational perspective**
03:00 PM	**Coffee Break & Poster Session**
04:00 PM	**Jürgen Schmidhuber - Learning how to Learn Learning Algorithms: Recursive Self-Improvement**
04:30 PM	**Joshua Tenenbaum & Kevin Ellis - Bayesian program learning: Prospects for building more human-like AI systems**
05:00 PM	**Alex Graves - Learning When to Halt With Adaptive Computation Time**
05:30 PM	**Debate with Percy Liang, Jürgen Schmidhuber, Joshua Tenenbaum and Martin Vechev**
06:30 PM	**Closing word**

Towards an Artificial Intelligence for Data Science

Charles Sutton, James Geddes, Zoubin Ghahramani, Padhraic Smyth, Chris Williams

Room 114, Sat Dec 10, 08:00 AM

Machine learning methods have applied beyond their origins in artificial intelligence to a wide variety of data analysis problems in fields such as science, health care, technology, and commerce. Previous research in machine learning, perhaps motivated by its roots in AI, has primarily aimed at fully-automated approaches for prediction problems. But predictive analytics is only one step in the larger pipeline of data science, which includes data wrangling, data cleaning, exploratory visualization, data integration, model criticism and revision, and presentation of results to domain experts.

An emerging strand of work aims to address all of these challenges in one stroke is by automating a greater portion of the full data science pipeline. This workshop will bring together experts in machine learning, data mining, databases and statistics to discuss the challenges that arise in the full end-to-end process of collecting data, analysing data, and making decisions and building new methods that support, whether in an automated or semi-automated way, more of the full process of analysing real data.

Considering the full process of data science raises interesting questions for discussion, such as: What aspects of data analysis might potentially be automated and what aspects seem more difficult? Statistical model building often emphasizes interpretability and human understanding, while machine learning often emphasizes predictive modeling --- are ML methods truly suitable for supporting the full data analysis pipeline? Do recent advances in ML offer help here? Finally, are there low hanging fruit, i.e., how much time is wasted on routine tasks in scientific data analysis that could be automated?

Specific topics of interest include: data cleaning, exploratory data analysis, semi-supervised learning, active learning, interactive machine learning, model criticism, automated and semi-automated model construction, usable machine learning, interpretable prediction methods and automatic methods to explain predictions. We are especially interested in contributions that take a broader perspective, i.e., that aim toward supporting the process of data science more holistically.

Schedule

09:10 AM	**Automated Data Cleaning via Multi-View Anomaly Detection**	*Dietterich*
09:50 AM	**Automatic Discovery of the Statistical Types of Variables in a Dataset**	*Valera, Ghahramani*
10:10 AM	**Poster spotlights**	
11:00 AM	**Invited talk, Christian Steinruecken**	*Steinruecken*
11:40 AM	**Probabilistic structure discovery in time series data**	*Janz, Paige, Rainforth, van de Meent*
12:00 PM	**Poster session**	
02:00 PM	**Invited talk, Carlos Guestrin**	*Guestrin*
02:40 PM	**An Overview of the DARPA Data Driven Discovery of Models (D3M) Program**	*Lippmann, Campbell*
03:30 PM	**Invited talk, Frank Hutter**	*Hutter*
04:10 PM	**Data Analytics as Data: A Semantic Workflow Approach**	*Bennett*
04:30 PM	**General-Purpose Inductive Programming for Data Wrangling Automation**	

Abstracts (7):

Abstract 1: **Automated Data Cleaning via Multi-View Anomaly Detection in Towards an Artificial Intelligence for Data Science**, *Dietterich* 09:10 AM

One of the first steps in the data analysis pipeline is data cleaning: detecting data from failed sensors. This talk will discuss the application of anomaly detection algorithms to find and remove bad readings from weather station data. We will review our previous work on DBN time series models and our current work on applying non-parametric anomaly detection algorithms as part of our SENSOR-DX multi-view anomaly detection architecture. A major challenge in evaluating these algorithms is to obtain ground truth, because real sensor data tends to be labeled conservatively by domain experts.

Abstract 3: **Poster spotlights in Towards an Artificial Intelligence for Data Science**, 10:10 AM

Isabel Valera and Zoubin Ghahramani.
Automatic Discovery of the Statistical Types of Variables in a Dataset

David Janz, Brooks Paige, Tom Rainforth, Jan-Willem van de Meent and Frank Wood
Probabilistic structure discovery in time series data

Richard Lippmann, William Campbell and Joseph Campbell
An Overview of the DARPA Data Driven Discovery of Models (D3M) Program

Kristin Bennett, John Erickson, Hannah De Los Santos, Evan Patton, John Sheehan and Deborah McGuinness
Data Analytics as Data: A Semantic Workflow Approach

Lidia Contreras-Ochando, Fernando Martínez-Plumed, Cesar Ferri, Jose Hernandez-Orallo and Maria Jose Ramirez-Quintana
General-Purpose Inductive Programming for Data Wrangling Automation

Lev Faivishevsky and Amitai Armon.
Using Downhill Simplex Method for Optimizing Machine Learning Training Running Time

Lidia Contreras-Ochando, Fernando Martínez-Plumed, Cesar Ferri, Jose Hernandez-Orallo and Maria Jose Ramirez-Quintana. Logging Data Scientists: Collecting Evidence for Data Science Automation

Lin Li, William Campbell, Cagri Dagli and Joseph Campbell
Making Sense of Unstructured Text Data

Cornelia Caragea
Identifying Descriptive Keyphrases from Scholarly Big Data

Simao Eduardo and Charles Sutton
Data Cleaning using Probabilistic Models of Integrity Constraints

Udayan Khurana, Fatemeh Nargesian, Horst Samulowitz, Elias Khalil and Deepak Turaga
Automating Feature Engineering

Zhao Xu and Lorenzo von Ritter
Poster Adaptive Streaming Anomaly Analysis

Abstract 4: **Invited talk, Christian Steinruecken in Towards an Artificial Intelligence for Data Science**, *Steinruecken* 11:00 AM

Christian Steinruecken, University of Cambridge

Abstract 5: **Probabilistic structure discovery in time series data in Towards an Artificial Intelligence for Data Science**, *Janz, Paige, Rainforth, van de Meent* 11:40 AM

Existing methods for structure discovery in time series data construct interpretable, compositional kernels for Gaussian process regression models. While the learned Gaussian process model provides posterior mean and variance estimates, typically the structure is learned via a greedy optimization procedure. This restricts the space of possible solutions and leads to over-confident uncertainty estimates. We introduce a fully Bayesian approach, inferring a full posterior over structures, which more reliably captures the uncertainty of the model.

Abstract 8: **An Overview of the DARPA Data Driven Discovery of Models (D3M) Program in Towards an Artificial Intelligence for Data Science**, *Lippmann, Campbell* 02:40 PM

Richard Lippmann, William Campbell, Joseph Campbell

A new DARPA program called Data Driven Discovery of Models (D3M) aims to develop automated model discovery systems that can be used by researchers with specific subject matter expertise to create empirical models of real, complex processes. Two major goals of this program are to allow experts to create empirical models without the need for data scientists and to increase the productivity of data scientists via automation. Automated model discovery systems developed will be tested on real-world problems that progressively get harder during the course of the program. Toward the end of the program, problems will be both unsolved and underspecified in terms of data and desired outcomes. The program will emphasize creating and leveraging open source technology and architecture. Our presentation reviews the goals and structure of this program which will begin early in 2017. Although the deadline for submitting proposals has past, we welcome suggestions concerning challenge tasks, evaluations, or new open-source data sets to be included for system development and evaluation that would supplement data currently being curated from many sources.

Abstract 10: **Data Analytics as Data: A Semantic Workflow Approach in Towards an Artificial Intelligence for Data Science**, *Bennett* 04:10 PM

Kristin Bennett, John Erickson, Hannah De Los Santos, Evan Patton, John Sheehan, Deborah McGuinness

By treating the end-to-end data science workflow as data itself and through the conceptual modeling of the goals and functional intent of the data analyst, the entire process of data analytics becomes open and accessible to the powerful tools of artificial intelligence, machine learning, statistics, and data mining. We examine the fundamental questions and capabilities that must be addressed to realize cap- turing and reasoning over workflows as well as interpreting and contextualizing their results. Our approach focuses on capturing key components of complete workflow processes, making explicit the “deep” semantics of the workflow plan; the analysis performed; the structure and sub-components of the workflow; and intermediate and final data products. Our goal is to provide sufficient detail to facilitate practical workflow and work product integration, interpretation, reuse, reproducibility, recommendation, and search. The structure for this workflow-as- data view is formalized by an extensible, reusable ontology that we are creating that applies to all aspects of the workflow representation and reasoning process. We report on our exploration and reuse of existing methods, tools and ontologies as well as our semantic analytics contributions to real world projects addressing childhood health challenges.

Abstract 11: **General-Purpose Inductive Programming for Data Wrangling Automation in Towards an Artificial Intelligence for Data Science**, 04:30 PM

Lidia Contreras-Ochando, Fernando Martínez-Plumed, Cesar Ferri, Jose Hernandez-Orallo, Maria Jose Ramirez-Quintana

Data acquisition, integration, transformation, cleansing and other highly tedious tasks take a large proportion of data science projects. These routine tasks are tedious basically because they are repetitive and, hence, automatable. As a consequence, progress in the automation of this process can lead to a dramatic reduction of the cost and duration of data science projects. Recently, Inductive Programming (IP) has shown a large potential as a paradigm for addressing this automation. This short paper elaborates on the recent success of induction using domain-specific languages (DSLs) for the automation of data wrangling process and advocating for the use of inductive programming over general-purpose declarative languages (GPDLs) using domain-specific background knowledge (DSBKs).

The Future of Gradient-Based Machine Learning Software

Alex Wiltschko, Zachary DeVito, Frederic Bastien, Pascal Lamblin

Room 115, Sat Dec 10, 08:00 AM

The calculation of gradients and other forms of derivatives is a core part of machine learning, computer vision, and physical simulation. But the manual creation of derivatives is prone to error and requires a high "mental overhead" for practitioners in these fields. However, the process of taking derivatives is actually the highly mechanical application of the chain rule and can be computed using formal techniques such as automatic or symbolic differentiation. A family of "autodiff" approaches exist, each with their own particular strengths and tradeoffs.

In the ideal case, automatically generated derivatives should be competitive with manually generated ones and run at near-peak performance on modern hardware, but the most expressive systems for autodiff which can handle arbitrary, Turing-complete programs, are unsuited for performance-critical applications, such as large-scale machine learning or physical simulation. Alternatively, the most performant systems are not designed for use outside of their designated application space, e.g. graphics or neural networks. This workshop will bring together developers and researchers of state-of-the-art solutions to generating derivatives automatically and discuss ways in which these solutions can be evolved to be both more expressive and achieve higher performance. Topics for discussion will include:

- Whether it is feasible to create a single differentiable programming language, or if we will always have separate solutions for different fields such as vision and ML.
- What are the primitive data types of a differentiable language? N-dimensional arrays are useful for many machine learning applications, but other domains make use of graph types and sparse matrices.
- What are the challenges in elevating an expressive autodiff implementation from just a “prototyping language” to one used directly in performance-critical industrial settings?
- A shared representation of programs like LLVM IR has transformed programming language and compiler research. Is there any benefit to a common representation of differentiable programs that would enable shared tooling amongst autodiff libraries and implementations?

Schedule

09:00 AM	**Opening Remarks**	
09:10 AM	**Barak A. Pearlmutter – Automatic Differentiation: History and Headroom**	*Pearlmutter*
09:50 AM	**David Duvenaud – No more mini-languages: The power of autodiffing full-featured Python**	*Duvenaud*
10:30 AM	**Coffee Break 1**	
11:00 AM	**Jeff Dean – TensorFlow: Future Directions for Simplifying Large-Scale Machine Learning**	*Dean*
11:40 AM	**Lunch break**	
02:30 PM	**Yoshua Bengio – Credit assignment: beyond backpropagation**	*Bengio*
03:10 PM	**Matthew Johnson – Autodiff writes your exponential family inference code**	*Johnson*
03:50 PM	**Coffee Break 2**	
04:30 PM	**Jeffrey M. Siskind – The tension between convenience and performance in automatic differentiation**	*Siskind*
05:10 PM	**Panel Discussion**	

Machine Learning Systems

Aparna Lakshmiratan, Li Erran Li, Siddhartha Sen, Sarah Bird, Hussein Mehanna

Room 116, Sat Dec 10, 08:00 AM

A new area is emerging at the intersection of machine learning (ML) and systems design. This birth is driven by the explosive growth of diverse applications of ML in production, the continued growth in data volume, and the complexity of large-scale learning systems. Addressing the challenges in this intersection demands a combination of the right abstractions -- for algorithms, data structures, and interfaces -- as well as scalable systems capable of addressing real world learning problems.

Designing systems for machine learning presents new challenges and opportunities over the design of traditional data processing systems. For example, what is the right abstraction for data consistency in the context of parallel, stochastic learning algorithms? What guarantees of fault tolerance are needed during distributed learning? The statistical nature of machine learning offers an opportunity for more efficient systems but requires revisiting many of the challenges addressed by the systems and database communities over the past few decades. Machine learning focused developments in distributed learning platforms, programming languages, data structures, general purpose GPU programming, and a wide variety of other domains have had and will continue to have a large impact in both academia and industry.

As the relationship between the machine learning and systems communities has grown stronger, new research in using machine learning tools to solve classic systems challenges has also grown. Specifically, as we develop larger and more complex systems and networks for storing, analyzing, serving, and interacting with data, machine learning offers promise for modeling system dynamics, detecting issues, and making intelligent, data-driven decisions within our systems. Machine learning techniques have begun to play critical roles in scheduling, system tuning, and network analysis. Through working with systems and databases researchers to solve systems challenges, machine learning researchers can both improve their own learning systems as well impact the systems community and infrastructure at large.

The goal of this workshop is to bring together experts working at the crossroads of ML, system design and software engineering to explore the challenges faced when building practical large-scale machine learning systems. In particular, we aim to elicit new connections among these diverse fields, identify tools, best practices and design principles. The workshop will cover ML and AI platforms and algorithm toolkits (Caffe, Torch, TensorFlow, MXNet and parameter server, Theano, etc), as well as dive into the reality of applying ML and AI in industry with challenges of data and organization scale (with invited speakers from companies like Google, Microsoft, Facebook, Amazon, Netflix, Uber and Twitter).

The workshop will have a mix of invited speakers and reviewed papers with talks, posters and panel discussions to facilitate the flow of new ideas as well as best practices which can benefit those looking to implement large ML systems in academia or industry.

Focal points for discussions and solicited submissions include but are not limited to:
- Systems for online and batch learning algorithms
- Systems for out-of-core machine learning
- Implementation studies of large-scale distributed learning algorithms --- challenges faced and lessons learned
- Database systems for Big Learning --- models and algorithms implemented, properties (fault tolerance, consistency, scalability, etc.), strengths and limitations
- Programming languages for machine learning
- Data driven systems --- learning for job scheduling, configuration tuning, straggler mitigation, network configuration, and security
- Systems for interactive machine learning
- Systems for serving machine learning models at scale

Schedule

08:45 AM	**Opening Remarks**	
09:00 AM	**Invited Talk: You've been using asynchrony wrong your whole life! (Chris Re, Stanford)**	*Ré*

09:20 AM	**Contributed Talk: Hemingway: Modeling Distributed Optimization Algorithms**	
09:40 AM	**Invited Talk: Paleo: A Performance Model for Deep Neural Networks (Ameet Talwalkar, UCLA)**	*Talwalkar*
10:00 AM	**Poster Previews**	
11:30 AM	**Invited Talk: Scaling Machine Learning Using TensorFlow (Jeff Dean, Google Brain)**	*Dean*
11:50 AM	**Contributed Talk: Demitasse: SPMD Programing Implementation of Deep Neural Network Library for Mobile Devices**	
12:10 PM	**Lunch**	
01:30 PM	**ML System Updates from Caffe (Andrew Tulloch), Clipper (Daniel Crankshaw), Decision Service (Siddhartha Sen), MxNET (Tianqi Chen), Torch (Soumith Chintala), and VW (John Langford)**	
02:50 PM	**Invited Talk: Optimizing Large-Scale Machine Learning Pipelines with KeystoneML (Tomer Kaftan, UW)**	*Kaftan*
03:10 PM	**Invited Talk: Optimizing Machine Learning and Deep Learning (John Canny, UC Berkeley & Google Research)**	*Canny*
03:30 PM	**Posters & Coffee**	
04:30 PM	**Contributed Talk: Yggdrasil: An Optimized System for Training Deep Decision Trees at Scale**	
04:50 PM	**Contributed Talk: TensorForest: Scalable Random Forests on TensorFlow**	
05:10 PM	**Closing Remarks**	

Bayesian Optimization: Black-box Optimization and Beyond

Roberto Calandra, Bobak Shahriari, Javier Gonzalez, Frank Hutter, Ryan P Adams

Room 117, Sat Dec 10, 08:00 AM

Bayesian optimization has emerged as an exciting subfield of machine learning that is concerned with the global optimization of expensive, noisy, black-box functions using probabilistic methods. Systems implementing Bayesian optimization techniques have been successfully used to solve difficult problems in a diverse set of applications. Many recent advances in the methodologies and theory underlying Bayesian optimization have extended the framework to new applications and provided greater insights into the behaviour of these algorithms. Bayesian optimization is now increasingly being used in industrial settings, providing new and interesting challenges that require new algorithms and theoretical insights. Classically, Bayesian optimization has been used purely for expensive single-objective black-box optimization. However, with the increased complexity of tasks and applications, this paradigm is proving to be too restricted. Hence, this year's theme for the workshop will be "black-box optimization and beyond". Among the recent trends that push beyond BO we can briefly enumerate: - Adapting BO to not-so-expensive evaluations. - "Open the black-box" and move away from viewing the model as a way of simply fitting a response surface, and towards modelling for the purpose of discovering and understanding the underlying process. For instance, this so-called grey-box modelling approach could be valuable in robotic applications for optimizing the controller, while simultaneously providing insight into the mechanical properties of the robotic system. - "Meta-learning", where a higher level of learning is used on top of BO in order to control the optimization process and make it more efficient. Examples of such meta-learning include learning curve prediction, Freeze-thaw Bayesian optimization, online batch selection, multi-task and multi-fidelity learning. - Multi-objective optimization where not a single objective, but multiple conflicting objectives are considered (e.g., prediction accuracy vs training time). The target audience for this workshop consists of both industrial and academic practitioners of Bayesian optimization as well as researchers working on theoretical and practical advances in probabilistic optimization. We expect that this pairing of theoretical and applied knowledge will lead to an interesting exchange of ideas and stimulate an open discussion about the long term goals and challenges of the Bayesian optimization community. A further goal is to encourage collaboration between the diverse set of researchers involved in Bayesian optimization. This includes not only interchange between industrial and academic researchers, but also between the many different subfields of machine learning which make use of Bayesian optimization or its components. We are also reaching out to the wider optimization and engineering communities for involvement.

Schedule

09:00 AM	**Introduction and opening remarks**
09:10 AM	**Invited Talk 1: Roman Garnett**
09:40 AM	**Contributed Talk 1: TBA**
09:55 AM	**Contributed Talk 2: TBA**
10:10 AM	**Poster Spotlights 1**
10:30 AM	**Coffe Break**

11:00 AM	**Invited Talk 2: Joshua Knowles**	
11:30 AM	**Poster Spotlights 2**	
11:45 AM	**Poster Session 1**	
12:00 PM	**Lunch Break**	
02:00 PM	**Invited Talk 3: Jasper Snoek**	
02:30 PM	**Poster Session 2**	
03:00 PM	**Coffee Break**	
03:30 PM	**Poster Session 3**	
04:00 PM	**Invited Talk 4: Marc Toussaint**	
04:30 PM	**Invited Talk 5: Katharina Eggensperger**	
05:00 PM	**Panel discussion — Black-box Optimization & Beyond**	

Adaptive and Scalable Nonparametric Methods in Machine Learning

Aaditya Ramdas, Arthur Gretton, Bharath K. Sriperumbudur, Han Liu, John Lafferty, Samory Kpotufe, Zoltán Szabó

Room 120 + 121, Sat Dec 10, 08:00 AM

Large amounts of high-dimensional data are routinely acquired in scientific fields ranging from biology, genomics and health sciences to astronomy and economics due to improvements in engineering and data acquisition techniques. Nonparametric methods allow for better modelling of complex systems underlying data generating processes compared to traditionally used linear and parametric models. From statistical point of view, scientists have enough data to reliably fit nonparametric models. However, from computational point of view, nonparametric methods often do not scale well to big data problems.

The aim of this workshop is to bring together practitioners, who are interested in developing and applying nonparametric methods in their domains, and theoreticians, who are interested in providing sound methodology. We hope to effectively communicate advances in development of computational tools for fitting nonparametric models and discuss challenging future directions that prevent applications of nonparametric methods to big data problems.

We encourage submissions on a variety of topics, including but not limited to:
- Randomized procedures for fitting nonparametric models. For example, sketching, random projections, core set selection, etc.
- Nonparametric probabilistic graphical models
- Scalable nonparametric methods
- Multiple kernel learning
- Random feature expansion
- Novel applications of nonparametric methods
- Bayesian nonparametric methods
- Nonparametric network models

This workshop is a fourth in a series of NIPS workshops on modern nonparametric methods in machine learning. Previous workshops focused on time/accuracy tradeoffs, high dimensionality and dimension reduction strategies, and automating the learning pipeline.

Schedule

08:30 AM	**Richard Samworth. Adaptation in log-concave density estimation**	*Samworth*
09:00 AM	**Ming Yuan. Functional nuclear norm and low rank function estimation.**	*Yuan*
09:30 AM	**Mladen Kolar. Post-Regularization Inference for Dynamic Nonparanormal Graphical Models.**	*Kolar*
11:00 AM	**Debarghya Ghoshdastidar, Ulrike von Luxburg. Do Nonparametric Two-sample Tests work for Small Sample Size? A Study on Random Graphs.**	
11:20 AM	**Diana Cai, Trevor Campbell, Tamara Broderick. Paintboxes and probability functions for edge-exchangeable graphs.**	
11:40 AM	**Alessandro Rudi, Raffaello Camoriano, Lorenzo Rosasco. Generalization Properties of Learning with Random Features.**	
12:00 PM	**Makoto Yamada, Yuta Umezu, Kenji Fukumizu, Ichiro Takeuchi. Post Selection Inference with Kernels.**	
12:20 PM	**Yunpeng Pan, Xinyan Yan, Evangelos Theodorou, Byron Boots. Solving the Linear Bellman Equation via Kernel Embeddings and Stochastic Gradient Descent.**	
12:40 PM	**Lunch break**	
02:30 PM	**Francis Bach. Harder, Better, Faster, Stronger Convergence Rates for Least-Squares Regression.**	*Bach*

03:00 PM	**Richard (Fangjian) Guo. Boosting Variational Inference.**	*Guo*
03:30 PM	**Break**	
03:45 PM	**Olga Klopp. Network models and sparse graphon estimation.**	*Klopp*
04:15 PM	**Emily Fox. Sparse Graphs via Exchangeable Random Measures.**	*Fox*
04:45 PM	**Coffee break + posters**	

Abstracts (11):

Abstract 1: **Richard Samworth. Adaptation in log-concave density estimation in Adaptive and Scalable Nonparametric Methods in Machine Learning**, *Samworth* 08:30 AM

The log-concave maximum likelihood estimator of a density on the real line based on a sample of size n is known to attain the minimax optimal rate of convergence of $O(n^{-4/5})$ with respect to, e.g., squared Hellinger distance. In this talk, we show that it also enjoys attractive adaptation properties, in the sense that it achieves a faster rate of convergence when the logarithm of the true density is k-affine (i.e. made up of k affine pieces), provided k is not too large. Our results use two different techniques: the first relies on a new Marshall's inequality for log-concave density estimation, and reveals that when the true density is close to log-linear on its support, the log-concave maximum likelihood estimator can achieve the parametric rate of convergence in total variation distance. Our second approach depends on local bracketing entropy methods, and allows us to prove a sharp oracle inequality, which implies in particular that the rate of convergence with respect to various global loss functions, including Kullback--Leibler divergence, is $O(kn^{-1} \log^{5/4} n)$ when the true density is log-concave and its logarithm is close to k-affine.

Abstract 2: **Ming Yuan. Functional nuclear norm and low rank function estimation. in Adaptive and Scalable Nonparametric Methods in Machine Learning**, *Yuan* 09:00 AM

The problem of low rank estimation naturally arises in a number of functional or relational data analysis settings, for example when dealing with spatio-temporal data or link prediction with attributes. We consider a unified framework for these problems and devise a novel penalty function to exploit the low rank structure in such contexts. The resulting empirical risk minimization estimator can be shown to be optimal under fairly general conditions.

Abstract 3: **Mladen Kolar. Post-Regularization Inference for Dynamic Nonparanormal Graphical Models. in Adaptive and Scalable Nonparametric Methods in Machine Learning**, *Kolar* 09:30 AM

We propose a novel class of dynamic nonparanormal graphical models, which allows us to model high dimensional heavy-tailed systems and the evolution of their latent network structures. Under this model we develop statistical tests for presence of edges both locally at a fixed index value and globally over a range of values. The tests are developed for a high-dimensional regime, are robust to model selection mistakes and do not require commonly assumed minimum signal strength. The testing procedures are based on a high dimensional, debiasing-free moment estimator, which uses a novel kernel smoothed Kendall's tau correlation matrix as an input statistic. The estimator consistently estimates the latent inverse Pearson correlation matrix uniformly in both index variable and kernel bandwidth. Its rate of convergence is shown to be minimax optimal. Thorough numerical simulations and an application to a neural imaging dataset support the usefulness of our method.

Joint work with Junwei Lu and Han Liu.

Abstract 4: **Debarghya Ghoshdastidar, Ulrike von Luxburg. Do Nonparametric Two-sample Tests work for Small Sample Size? A Study on Random Graphs. in Adaptive and Scalable Nonparametric Methods in Machine Learning**, 11:00 AM

We consider the problem of two-sample hypothesis testing for inhomogeneous unweighted random graphs, where one has access to only a very small number of samples from each model. Standard tests cannot be guaranteed to perform well in this setting due to the small sample size. We present a nonparametric test based on comparison of the adjacency matrices of the graphs, and prove that the test is consistent for increasing sample size as well as when the graph size increases with sample size held fixed. Numerical simulations exhibit the practical significance of the test.

Abstract 6: **Alessandro Rudi, Raffaello Camoriano, Lorenzo Rosasco. Generalization Properties of Learning with Random Features. in Adaptive and Scalable Nonparametric Methods in Machine Learning**, 11:40 AM

We study the generalization properties of regularized learning with random features in the statistical learning theory framework. We show that optimal learning errors can be achieved with a number of features smaller than the number of examples.

Abstract 7: **Makoto Yamada, Yuta Umezu, Kenji Fukumizu, Ichiro Takeuchi. Post Selection Inference with Kernels. in Adaptive and Scalable Nonparametric Methods in Machine Learning**, 12:00 PM

We propose a novel kernel based post selection inference (PSI) algorithm, which can not only handle non-linearity in data but also structured output such as multi-dimensional and multi-label outputs. Specifically, we develop a PSI algorithm for independence measures, and propose the Hilbert-Schmidt Independence Criterion (HSIC) based PSI algorithm (hsicInf). We apply the hsicInf algorithm to a real-world data, and show that hsicInf can successfully identify important features.

Abstract 8: **Yunpeng Pan, Xinyan Yan, Evangelos Theodorou, Byron Boots. Solving the Linear Bellman Equation via Kernel Embeddings and Stochastic Gradient Descent. in Adaptive and Scalable Nonparametric Methods in Machine Learning**, 12:20 PM

We introduce a data-efficient approach for solving the linear Bellman equation, which corresponds to a class of Markov decision processes (MDPs) and stochastic optimal control (SOC) problems. We show that this class of control problem can be reformulated as a stochastic composition optimization problem, which can be further reformulated as a saddle point problem and solved via dual kernel embeddings. Our method is model-free and using only one sample per state transition from stochastic dynamical systems. Different from related work such as

Z-learning based on temporal-difference learning, our method is an on-line algorithm exploiting stochastic optimization. Numerical results are provided, showing that our method outperforms the Z-learning algorithm.

Abstract 10: **Francis Bach. Harder, Better, Faster, Stronger Convergence Rates for Least-Squares Regression. in Adaptive and Scalable Nonparametric Methods in Machine Learning**, *Bach* 02:30 PM

We consider the optimization of a quadratic objective function whose gradients are only accessible through a stochastic oracle that returns the gradient at any given point plus a zero-mean finite variance random error. We present the first algorithm that achieves jointly the optimal prediction error rates for least-squares regression, both in terms of forgetting of initial conditions in O(1/n^2), and in terms of dependence on the noise and dimension d of the problem, as O(d/n). Our new algorithm is based on averaged accelerated regularized gradient descent, and may also be analyzed through finer assumptions on initial conditions and the Hessian matrix, leading to dimension-free quantities that may still be small while the "optimal " terms above are large. In order to characterize the tightness of these new bounds, we consider an application to non-parametric regression and use the known lower bounds on the statistical performance (without computational limits), which happen to match our bounds obtained from a single pass on the data and thus show optimality of our algorithm in a wide variety of particular trade-offs between bias and variance. [joint work with Aymeric Dieuleveut and Nicolas Flammarion]

Abstract 11: **Richard (Fangjian) Guo. Boosting Variational Inference. in Adaptive and Scalable Nonparametric Methods in Machine Learning**, *Guo* 03:00 PM

Modern Bayesian inference typically requires some form of posterior approximation, and mean-field variational inference (MFVI) is an increasingly popular choice due to its speed. But MFVI is inaccurate in several aspects, including an inability to capture multimodality in the posterior and underestimation of the posterior covariance. These issues arise since MFVI considers approximations to the posterior only in a family of factorized parametric distributions. We instead consider a much more flexible approximating family consisting of all possible mixtures of a parametric base distribution (e.g., Gaussians) without constraining the number of mixture components. In order to efficiently find a high-quality posterior approximation within this family, we borrow ideas from gradient boosting and propose the boosting variational inference (BVI) method, which iteratively improves the current approximation by mixing it with a new component from the base distribution family. We develop practical algorithms for BVI and demonstrate their performance on both real and simulated data. Joint work with Xiangyu Wang, Kai Fan, Tamara Broderick and David Dunson.

Abstract 13: **Olga Klopp. Network models and sparse graphon estimation. in Adaptive and Scalable Nonparametric Methods in Machine Learning**, *Klopp* 03:45 PM

Inhomogeneous random graph models encompass many network models such as stochastic block models and latent position models. We consider the problem of statistical estimation of the matrix of connection probabilities based on the observations of the adjacency matrix of the network and derive optimal rates of convergence for this problem. Our results cover the important setting of sparse networks. We also establish upper bounds on the minimax risk for graphon estimation when the probability matrix is sampled according to a graphon model.

Abstract 14: **Emily Fox. Sparse Graphs via Exchangeable Random Measures. in Adaptive and Scalable Nonparametric Methods in Machine Learning**, *Fox* 04:15 PM

Statistical network modeling has focused on representing the graph as a discrete structure, namely the adjacency matrix. Assuming exchangeability of this array, the Aldous-Hoover theorem informs us that the graph is necessarily either dense or empty. We instead consider representing the graph as a point process on the positive quadrant. We then propose a graph construction leveraging completely random measures (CRMs) that leads to an exchangeable point process representation of graphs ranging from dense to sparse and exhibiting power-law degree distributions. We show how these properties are simply tuned by three hyperparameters. The resulting model lends itself to an efficient MCMC scheme from which we can infer these network attributes. We demonstrate our methods on a series of real-world networks with up to hundreds of thousands of nodes and millions of edges. We also discuss some recent advances in this area and open challenges. Joint work with Francois Caron.

Computing with Spikes

Sander M Bohte, Thomas Nowotny, Cristina Savin, Davide Zambrano

Room 122 + 123, Sat Dec 10, 08:00 AM

Despite remarkable computational success, artificial neural networks ignore the spiking nature of neural communication that is fundamental for biological neuronal networks. Understanding how spiking neurons process information and learn remains an essential challenge. It concerns not only neuroscientists studying brain function, but also neuromorphic engineers developing low-power computing architectures, or machine learning researchers devising new biologically-inspired learning algorithms. Unfortunately, despite a joint interest in spike-based computation, the interactions between these subfields remains limited. The workshop aims to bring them together and to foster the exchange between them by focusing on recent developments in efficient neural coding and spiking neurons' computation. The discussion will center around critical questions in the field, such as "what are the underlying paradigms?" "what are the fundamental constraints?", and "what are the measures for progress?", that benefit from varied perspectives. The workshop will combine invited talks reviewing the state-of-the-art and short contributed presentations; it will conclude with a panel discussion.

Schedule

08:50 AM	**Workshop opening**	
09:00 AM	**Reward-based self-configuration of networks of spiking neurons**	*Maass*
09:30 AM	**Robotic Vision with Dynamic Vision Sensors**	*Delbruck*
10:00 AM	**Spotlight Presentations I**	
10:00 AM	**Theory and Tools for the Conversion of Analog to Spiking Convolutional Neural Networks**	*Rueckauer*

10:05 AM	**Fast and Efficient Asynchronous Neural Computation in Deep Adaptive Spiking Neural Networks**	*Zambrano*
10:10 AM	**A wake-sleep algorithm for recurrent, spiking neural networks**	
10:15 AM	**Deep counter networks for asynchronous event-based processing**	*Binas*
10:20 AM	**Spike-based reinforcement learning for temporal stimulus segmentation and decision making**	*La Camera*
10:30 AM	**Coffee break and Posters**	
11:00 AM	**Deep Learning for Neuromorphic Computing**	*Merolla*
11:30 AM	**Spotlight Presentations II**	
11:30 AM	**Deep Spiking Networks**	*O'Connor*
11:35 AM	**Optimization-based computation with spiking neurons**	*Verzi*
11:40 AM	**Towards deep learning with spiking neurons in energy based models with contrastive Hebbian plasticity**	*Brea*
11:45 AM	**Can we be formal in assessing the strengths and weaknesses of neural architectures? A case study using a spiking cross-correlation algorithm**	*Severa*
11:50 AM	**Poster Session I**	
12:30 PM	**Lunch**	
02:00 PM	**tba**	*Denève*
02:30 PM	**Programming with spikes: The Nengo framework for efficient and adaptive large-scale spiking systems**	*Stewart*
03:30 PM	**SpiNNaker: a platform for computing with spikes**	*Plana*
04:00 PM	**Spike-based probabilistic computation**	*Savin*
04:30 PM	**Panel Discussion**	

Abstracts (8):

Abstract 2: **Reward-based self-configuration of networks of spiking neurons in Computing with Spikes**, *Maass* 09:00 AM

It is very difficult to construct by hand recurrent networks of noisy spiking neurons that are able to carry out nontrivial computational tasks. Obviously evolution has found a different strategy for that. Therefore we have analyzed the power of reward-based learning for configuring the connections and parameters (synaptic weights) of such a network. More specifically, we have considered a model where stochastic local plasticity rules drive the network to search for highly rewarded network configurations. On the abstract level, the resulting paradigm provides an interesting alternative to classical policy learning through gradient ascent: A continuous policy search through stochastic sampling from a posterior distribution that integrates structural constraints with reward expectations.

Abstract 4: **Spotlight Presentations I in Computing with Spikes**, 10:00 AM

Theory and Tools for the Conversion of Analog to Spiking Convolutional Neural Networks
Bodo Rueckauer, Iulia-Alexandra Lungu, Yuhuang Hu, and Michael Pfeiffer

Fast and Efficient Asynchronous Neural Computation in Deep Adaptive Spiking Neural Networks
Davide Zambrano and Sander Bohte

A wake-sleep algorithm for recurrent, spiking neural networks
Johannes Thiele, Peter Diehl and Matthew Cook

Deep counter networks for asynchronous event-based processing
Jonathan Binas, Giacomo Indiveri and Michael Pfeiffer

Spike-based reinforcement learning for temporal stimulus segmentation and decision making
Luisa Le Donne, Luca Mazzucato, Robert Urbanczik, Walter Senn and Giancarlo La Camera

Abstract 7: **A wake-sleep algorithm for recurrent, spiking neural networks in Computing with Spikes**, 10:10 AM

Johannes Thiele, Peter Diehl and Matthew Cook

Abstract 11: **Deep Learning for Neuromorphic Computing in Computing with Spikes**, *Merolla* 11:00 AM

Deep learning has made great strides in the last few years. For example, it is now possible to train networks with millions of neurons--using gradient-based learning methods--to classify images at near human performance. One exciting possibility is to run these networks on energy-efficient neuromorphic hardware, such as IBM's TrueNorth chip. However, these specialized architectures impose constraints that are not typically considered in deep learning; for example to achieve energy efficiency, TrueNorth uses low precision synapses, spiking neurons, and restricted fan-in. In this talk, I will describe our recent work that modifies deep learning to be compatible with typical neuromorphic constraints. Using this approach, we demonstrate near state-of-the-art accuracy on 8 datasets, while running between 1,200 and 2,600 frames per second and using between 25mW and 275mW on TrueNorth.

Abstract 12: **Spotlight Presentations II in Computing with Spikes**, 11:30 AM

Deep Spiking Networks
Peter O'Connor and Max Welling

Optimization-based computation with spiking neurons
Stephen Verzi, Craig Vineyard, Eric Vugrin, Meghan Galiardi, Conrad James and James Aimone

Towards deep learning with spiking neurons in energy based models with contrastive Hebbian plasticity
Thomas Mesnard, Wulfram Gerstner and Johanni Brea

Can we be formal in assessing the strengths and weaknesses of neural architectures? A case study using a spiking cross-correlation algorithm
William Severa, Kristofor Carlson, Ojas Parekh, Craig Vineyard and James Aimone

Abstract 17: **Poster Session I in Computing with Spikes**, 11:50 AM

Storage capacity of spatio-temporal patterns in LIF spiking networks: mixed rate and phase coding
Antonio de Candia and Siliva Scarpetta,

Theory and Tools for the Conversion of Analog to Spiking Convolutional Neural Networks
Bodo Rueckauer, Iulia-Alexandra Lungu, Yuhuang Hu, and Michael Pfeiffer

Somatic inhibition controls dendritic selectivity in a 2 sparse coding network of spiking neurons.
Damien Drix

Fast and Efficient Asynchronous Neural Computation in Deep Adaptive Spiking Neural Networks
Davide Zambrano and Sander Bohte

Spiking memristor logic gates are a type of time-variant perceptron.
Ella Gale.

A wake-sleep algorithm for recurrent, spiking neural networks
Johannes Thiele, Peter Diehl and Matthew Cook

Deep counter networks for asynchronous event-based processing
Jonathan Binas, Giacomo Indiveri and Michael Pfeiffer

Spike-based reinforcement learning for temporal stimulus segmentation and decision making
Luisa Le Donne, Luca Mazzucato, Robert Urbanczik, Walter Senn and Giancarlo La Camera

Deep Spiking Networks
Peter O'Connor and Max Welling

Working Memory in Adaptive Spiking Neural Networks
Roeland Nusselder, Davide Zambrano and Sander Bohte

An Efficient Approach to Boosting Performance of Deep Spiking Network Training
Seongsik Park, Sung-gil Lee, Huynha Nam and Sungroh Yoon.

Optimization-based computation with spiking neurons
Stephen Verzi, Craig Vineyard, Eric Vugrin, Meghan Galiardi, Conrad James and James Aimone

Learning binary or real-valued time-series via spike-timing dependent plasticity
Takayuki Osogami

Towards deep learning with spiking neurons in energy based models with contrastive Hebbian plasticity
Thomas Mesnard, Wulfram Gerstner and Johanni Brea

Can we be formal in assessing the strengths and weaknesses of neural architectures? A case study using a spiking cross-correlation algorithm
William Severa, Kristofor Carlson, Ojas Parekh, Craig Vineyard and James Aimone

Nonnegative autoencoder with simplified random neural network
Yonghua Yin and Erol Gelenbe

Abstract 20: **Programming with spikes: The Nengo framework for efficient and adaptive large-scale spiking systems in Computing with Spikes**, *Stewart* 02:30 PM

Given the rapidly growing interest in neuromorphics and spike-based computation, there are a wide range of techniques, software frameworks, and hardware implementations that explore these ideas. We have been integrating some of these approaches into a common software toolkit, Nengo, which provides a high-level programming interface for the specification of spike-based neural networks, and then compiles these models to target different hardware, including CPUs, GPUs, digital neuromorphics, and analog neuromorphics. We will discuss some of the challenges involved in compiling to such a wide range of hardware, and show examples of efficiency gains both for neuroscientific modelling of large-scale biological systems and for modern machine-learning algorithms such as deep networks.

Abstract 21: **SpiNNaker: a platform for computing with spikes in Computing with Spikes**, *Plana* 03:30 PM

Luis Plana: The SpiNNaker machine supports large-scale spiking neural networks that operate in biological real time with up to hundreds of million of neurons and hundreds of billions of synapses. So far demonstrations of the machine's capabilities have been modest in scale, such as small-scale cortical microcolumn models and a stochastic spiking network that solves “diobolical” Sudoku problems, but the platform is now openly available under the auspices of the EU Flagship Human Brain Project, and we look forward to much larger, more challenging demonstrations over the next year or two!

Constructive Machine Learning

Fabrizio Costa, Thomas Gärtner, Andrea Passerini, Francois Pachet

Room 127 + 128, Sat Dec 10, 08:00 AM

In many real-world applications, machine learning algorithms are employed as a tool in a "constructive process". These processes are similar to the general knowledge-discovery process but have a more specific goal: the construction of one-or-more domain elements with particular properties. In this workshop we want to bring together domain experts employing machine learning tools in constructive processes and machine learners investigating novel approaches or theories concerning constructive processes as a whole. Interesting applications include but are not limited to: image synthesis, drug and protein design, computational cooking, generation of art (paintings, music, poetry). Interesting approaches include but are not limited to: deep generative learning, active approaches to structured output learning, transfer or multi-task learning of generative models, active search or online optimization over relational domains, and learning with constraints.

Many of the applications of constructive machine learning, including the ones mentioned above, are primarily considered in their respective application domain research area but are hardly present at machine learning conferences. By bringing together domain experts and machine learners working on constructive ML, we hope to bridge this gap between the communities.

Schedule

08:30 AM	**Introduction**	*Costa, Passerini, Gärtner, Pachet*
08:45 AM	**Artificially-intelligent drug design**	*Schneider*
09:15 AM	**Chef Watson: Computational Creativity Applied To Recipes**	*Pinel*
09:45 AM	**Efficient optimization for probably submodular constraints in CRFs**	*Berman*
10:00 AM	**A constructive approach for graph concepts with long range dependencies**	*Mautner, Costa*
10:15 AM	**Constructive Layout Synthesis via Coactive Learning**	*Dragone, Passerini*
11:00 AM	**Multiplicative and Fine-grained Gating for Reading Comprehension**	*Salakhutdinov*
11:30 AM	**Magenta**	
12:00 PM	**Chord2Vec: Learning Musical Chord Embeddings**	*Walder*
12:00 PM	**A Machine Learning Approach to Support Music Creation by Musically Untrained People**	*Kitahara*
12:00 PM	**Collaborative creativity with Monte-Carlo Tree Search and Convolutional Neural Networks**	*Akten*
12:00 PM	**Modelling human appreciation of machine generated What-if ideas**	*Žnidarši■, Kranjc*
01:30 PM	**Narrated Reality**	*Goodwin*
02:00 PM	**Computational Creativity**	*Colton*
02:30 PM	**Fast Patch-based Style Transfer of Arbitrary Style**	*Chen, Schmidt*
02:45 PM	**Out-of-class novelty generation: an experimental foundation**	*Kégl*
03:30 PM	**Structured Prediction with Logged Bandit Feedback**	*Joachims*
04:00 PM	**Automatic Chemical Design using Variational Autoencoders**	*Hernández-Lobato*
04:30 PM	**Generating Class-conditional Images with Gradient-based Inference**	*Duvenaud*
04:30 PM	**C-RNN-GAN: Generative adversarial training of sequence models with continuous data**	*Mogren*
04:30 PM	**Optimal Teaching for Online Perceptrons**	*Zhang, Zhu*
04:40 PM	**Poster Session**	
05:10 PM	**Panel Discussion**	*Schneider, Goodwin, Colton, Salakhutdinov, Joachims, Pinel*

Abstracts (5):

Abstract 2: **Artificially-intelligent drug design in Constructive Machine Learning**, *Schneider* 08:45 AM

Future success in pharmaceutical research will fundamentally rely on the combination of advanced synthetic and analytical technologies that are embedded in a theoretical framework that provides a rationale for the interplay between chemical structure and biological effect. A driving role in this setting falls on leading edge concepts in computer-assisted molecular design and machine learning, by providing access to a virtually infinite source of novel tool compounds and lead structures, and guiding experimental screening campaigns. We will discuss representations of molecular structure, predictive models of structure-activity relationships using constructive machine learning, automated molecular de novo design, and showcase prospective applications. Emphasis will be put on the automated construction of potent and selective new chemical entities. As we are currently witnessing strong renewed interest in bioactive natural products we will present applications of this approach to natural-product inspired molecular design.

Abstract 3: **Chef Watson: Computational Creativity Applied To Recipes in Constructive Machine Learning**, *Pinel* 09:15 AM

Can computers be creative? Meet Chef Watson. Aimed at adventurous cooks, Chef Watson is a cognitive computing application revolutionizing how people combine ingredients to create unique dishes with novel flavors.
Compared to artifacts in expressive or performance domains, work

products resulting from scientific creativity (including culinary recipes) seem much more conducive to data-driven assessment. One can apply computationally intensive techniques to generate many possible combinations and use automated assessors to evaluate each of them. Assembly work plans for the selected novel products can subsequently be inferred from existing records.
Chef Watson applies this approach to the culinary world. After gathering data and creating a knowledge base of recipes, ingredients, and flavor compounds, the system generates ingredient combinations that satisfy user inputs such as the choice of a key ingredient, desired dish, and dietary constraints. Once a combination has been selected with the help of novelty and quality evaluators, the system further generates ingredient proportions and recipe steps. Using several variations of this approach, the system can generate new wildly creative recipes, or merely adapt existing recipes to personal preferences.

Abstract 13: **Narrated Reality in Constructive Machine Learning**, *Goodwin* 01:30 PM

Can machine intelligence enable new forms and interfaces for written language, or does it merely reveal an "uncanny valley" of text? Join Ross Goodwin as he discusses his work with neural networks for creative applications, including expressive image captioning, narration devices for your home and car, and a film (Sunspring) created from a computer generated screenplay.

Abstract 14: **Computational Creativity in Constructive Machine Learning**, *Colton* 02:00 PM

In Computational Creativity research, we study how to engineer software which
can take on creative responsibilities in arts and science projects. At the heart of most creative systems is a generative engine, and constructive machine learning has the potential to drive forward Computational Creativity
research with new generative processes and the production of new cultural
artifacts such as paintings and musical compositions. In an effort to help the
emerging field of constructive machine learning to fast-track to having cultural (as well as scientific) impact, in the talk, I'll describe some of the practical projects I've been involved with and what lessons I've learned
about the value of creative software in society at large. I'll describe some foundational philosophical issues that have arise in the field over recent years, and discuss how we've addressed these issues to make scientific progress, but also to lay the groundwork for creative software to have an important and lasting impact in certain cultural spheres.

Abstract 17: **Structured Prediction with Logged Bandit Feedback in Constructive Machine Learning**, *Joachims* 03:30 PM

Conventional supervised learning algorithms require training data that includes 'optimal' labels. Unfortunately, such optimal labels may be difficult to annotate or even define for many constructive ML tasks. For example, what is the optimal layout of a personalized newspaper for a particular user on a given day? While the optimal layout may be unattainable as training data, it may be easy to infer the quality of a particular layout that was presented to the user (e.g., from behavioral signals). This means that we may easily get bandit feedback for learning, but not full-information feedback. In fact, such bandit-style log data is one of the most ubiquitous forms of data available, as it can be recorded from a variety of systems (e.g., search engines, recommender systems, ad placement) at little cost.

Machine Learning for Education

Richard Baraniuk, Jiquan Ngiam, Christoph Studer, Phillip Grimaldi, Andrew Lan

Room 129 + 130, Sat Dec 10, 08:00 AM

In recent years, we have seen a rise in the amount of education data available through the digitization of education. Schools are starting to use technology in classrooms to create personalized learning experiences. Massive open online courses (MOOCs) have attracted millions of learners and present an opportunity for us to apply and develop machine learning methods towards improving student learning outcomes, leveraging the data collected.

However, development in student data analysis remains limited, and education largely follows a one-size-fits-all approach today. We have an opportunity to have a significant impact in revolutionizing the way (human) learning can work.

The goal of this workshop is to foster discussion and spur research between machine learning experts and researchers in education fields that can solve fundamental problems in education.

For this year's workshop, we are highlighting the following areas of interest:

-- Assessments and grading
Assessments are core in adaptive learning, formative learning, and summative evaluation. However, the creation and grading of quality assessments remains a difficult task for instructors. Machine learning methods can be applied to self-, peer-, auto-grading paradigms to both improve the quality of assessments and reduce the burden on instructors and students. These methods can also leverage the multimodal nature of learner data (i.e., textual/programming/mathematical open-form responses, demographic information, student interaction in discussion forums, video and audio recording of the class), posing challenges of how to effectively and efficiently fuse these different forms of data so that we can better understand learners.

-- Content augmentation and understanding:
Learning content is rich and multimodal (e.g., programming code, video, text, audio). There has been a growth of online educational resources in the past years, and we have an opportunity to leverage them further. Recent advances in natural language understanding can be applied to understand learning materials better and connect different sources together to create better learning experiences. This can help learners by providing them with more relevant resources and instructors in the creation of content.

-- Personalized learning and active interventions:
Personalized learning through custom feedback and interventions can make learning much more efficient, especially when we cater to the individual's background, goals, state of understanding, and learning context. Methods such as Markov Decision Processes and Multi-armed Bandits are applicable in these context.

-- Human-interpretability:
In education applications, transparency and interpretability is important as it can help learners better understand their learning state. Interpretability can provide instructors with insights to better guide their activities with students. It can also help education researchers better understand the foundations of human learning. This can also be especially critical where models are deployed in processes that grade students, as evaluation needs to demonstrate a degree of fairness.

This workshop will lead to new research directions in machine learning-driven educational research and also inspire the development of novel machine learning algorithms and theories that can extend to a large number of other applications that study human data.

Schedule

08:30 AM	**Opening remarks**
08:40 AM	**Phil Grimaldi, OpenStax/Rice University -- BLAh: Boolean Logic Analysis for Graded Student Response Data**
09:00 AM	**Steve Ritter, Carnegie Learning -- Eliminating testing through continuous assessment**
09:25 AM	**Pieter Abbeel, UC Berkeley -- Gradescope -- AI for Grading**
09:50 AM	**Mihaela van der Schaar, UCLA -- A Machine Learning Approach to Personalizing Education: Improving Individual Learning through Tracking and Course Recommendation**
10:40 AM	**Zhenghao Chen, Coursera -- Machine Learning Challenges and Opportunities in MOOCs**
11:00 AM	**Lise Getoor, UC Santa Cruz -- Understanding Engagement and Sentiment in MOOCs using Probabilistic Soft Logic (PSL)**
11:25 AM	**Kangwook Lee, KAIST -- Machine Learning Approaches for Learning Analytics: Collaborative Filtering Or Regression With Experts?**
11:50 AM	**Poster spotlight**
12:00 PM	**Lunch break**
02:00 PM	**Poster session**
02:30 PM	**Anna Rafferty, Carleton College -- Using Computational Methods to Improve Feedback for Learners**
02:55 PM	**Michael Mozer, CU Boulder -- Estimating student proficiency: Deep learning is not the panacea**
03:20 PM	**Yan Karklin, Knewton -- Modeling skill interactions with multilayer item response functions**
03:45 PM	**Coffee break**
04:10 PM	**Utkarsh Upadhyay, MPI-SWS -- On Crowdlearning: How do People Learn in the Wild?**
04:35 PM	**Christopher Brinton, Zoomi -- Beyond Assessment Scores: How Behavior Can Give Insight into Knowledge Transfer**
05:00 PM	**Emma Brunskill, CMU -- Using Old Data To Yield Better Personalized Tutoring Systems**
05:25 PM	**Panel discussion and closing remarks**

Connectomics II: Opportunities and Challenges for Machine Learning

Viren Jain, Srini C Turaga

Room 131 + 132, Sat Dec 10, 08:00 AM

The "wiring diagram" of essentially all nervous systems remains unknown due to the extreme difficulty of measuring detailed patterns of synaptic connectivity of entire neural circuits. At this point, the major bottleneck is in the analysis of tera or peta-voxel 3d electron microscopy image data in which neuronal processes need to be traced and synapses localized in order for connectivity information to be inferred. This presents an opportunity for machine learning and machine perception to have a fundamental impact on advances in neurobiology. However, it also presents a major challenge, as existing machine learning methods fall short of solving the problem.
The goal of this workshop is to bring together researchers in machine learning and neuroscience to discuss progress and remaining challenges in this exciting and rapidly growing field. We aim to attract machine learning and computer vision specialists interested in learning about a new problem, as well as computational neuroscientists at NIPS who may be interested in modeling connectivity data. We will discuss the release of public datasets and competitions that may facilitate further activity in this area. We expect the workshop to result in a significant increase in

the scope of ideas and people engaged in this field.

Schedule

08:30 AM	**Viren Jain (Google)**	
08:40 AM	**Keynote: Terry Sejnowski (Salk Institute)**	
09:00 AM	**Nir Shavit (MIT)**	

"What If?" Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems

Ricardo Silva, John Shawe-Taylor, Adith Swaminathan, Thorsten Joachims

Room 133 + 134, Sat Dec 10, 08:00 AM

One of the promises of Big Data is its potential to answer “what if?” questions in digital, natural and social systems. Whether we speak of genetic interactions in a cell, passengers commuting in railways and roads, recommender systems matching users to ads, or understanding contagion in social networks, such systems are composed of many interacting components that suggest that learning to control them or understanding the effect of shocks to a system is not an easy task. What if some railways are closed, what will passengers do? What if we incentivize a member of a social network to propagate an idea, how influential can they be? What if some genes in a cell are knocked-out, which phenotypes can we expect?

Such questions need to be addressed via a combination of experimental and observational data, and require a careful approach to modelling heterogeneous datasets and structural assumptions concerning the causal relations among the components of the system. The workshop is aimed at bringing together research expertise from a variety of communities in machine learning, statistics, engineering, and the social, medical and natural sciences. It is an opportunity for methods for causal inference, reinforcement learning and game theory to be cross-fertilized with more traditional research in statistics and the real-world constraints found in practical applications. Ultimately, this can lead to new research platforms to aid the assessment of policies, shocks and experimental design methods in the discovery of breakthroughs in a variety of domains.

Schedule

08:20 AM	**Welcome**	
08:30 AM	**The Data-Fusion Problem: Causal Inference and Reinforcement Learning**	*Bareinboim*
09:15 AM	**A Contextual Research Program**	*Langford*
10:00 AM	**Poster Session I**	
11:00 AM	**Optimal and Adaptive Off-policy Evaluation in Contextual Bandits**	*Wang*
11:30 AM	**Joint Causal Inference on Observational and Experimental Datasets**	*Magliacane*
01:45 PM	**Extracting Templates from Media Event Sequences**	*Grobelnik*
02:30 PM	**Estimating What-if Outcomes for Targeting Interventions in a Clinical Setting**	*Saria*
03:00 PM	**Poster Session II**	
04:00 PM	**Long-term Causal Effects in Policy Analysis**	*Toulis*
04:30 PM	**Causal Inference for Recommendation Systems**	*Blei*
05:15 PM	**Panel & Closing**	

Abstracts (8):

Abstract 3: **A Contextual Research Program in "What If?" Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems**, *Langford* 09:15 AM

A theory of contextual interventions has developed and matured to the point where contextual bandits can be routinely deployed to solve appropriate problems. A more general theory of contextual interventions in complex settings appears desirable and is under development leading to developments in two new areas:

1) Sequential decision making around deviations from existing solutions
2) Global exploration strategies for arbitrary contexts.

Abstract 4: **Poster Session I in "What If?" Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems**, 10:00 AM

The first of two sessions. Each session will include all posters.

Abstract 5: **Optimal and Adaptive Off-policy Evaluation in Contextual Bandits in "What If?" Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems**, *Wang* 11:00 AM

We consider the problem of off-policy evaluation—estimating the value of a target policy using data collected by another policy—under the contextual bandit model. We establish a minimax lower bound on the mean squared error (MSE), and show that it is matched up to constant factors by the inverse propensity scoring (IPS) estimator. Since in the multi-armed bandit problem the IPS is suboptimal, our result highlights the difficulty of the contextual setting with non-degenerate context distributions. We further consider improvements on this minimax MSE bound, given access to a reward model. We show that the existing doubly robust approach, which utilizes such a reward model, may continue to suffer from high variance even when the reward model is perfect. We propose a new estimator called SWITCH which more effectively uses the reward model and achieves a superior bias-variance tradeoff compared with prior work. We prove an upper bound on its MSE and demonstrate its benefits empirically on a diverse collection of datasets, often seeing orders of magnitude improvements over a number of baselines.

Abstract 6: **Joint Causal Inference on Observational and Experimental Datasets in "What If?" Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems**, *Magliacane* 11:30 AM

We introduce joint causal inference, a powerful formulation of causal discovery over multiple datasets in which we jointly learn both the causal structure and targets of interventions from independence test results. While offering many advantages, joint causal inference induces faithfulness violations due to deterministic relations, so we extend a recently proposed constraint-based method to deal with this type of violations. A preliminary evaluation shows the benefits of joint causal inference.

Abstract 7: **Extracting Templates from Media Event Sequences in "What If?" Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems**, *Grobelnik* 01:45 PM

In this preliminary research we'll present early results on extracting repeatable probabilistic templates from global media-event sequences. Such patterns could hint on some weak forms of causality in the global social dynamics. As a basis, we are using the evolving graph of interlinked events generated by the "Event Registry" system (eventregistry.org), where each event is represented as an object composed from three main components: social, topical and temporal. In the analysis we will show early results on the structure of the problem and the spectrum of extracted templates from simple to hard ones.

Abstract 8: **Estimating What-if Outcomes for Targeting Interventions in a Clinical Setting in "What If?" Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems**, *Saria* 02:30 PM

Individuals have heterogeneous outcomes from interventions. In a clinical setting, estimating how patients will respond to different treatments is critical for targeted care. Clinicians constantly ask themselves, given a patient's history, what would happen to the patient's clinical trajectory if they were given one treatment versus another. However, in practice it is often unknown how the patient's signals will change in response to treatment until that treatment is actually administered. Even then, it is impossible to observe the counterfactual from real data, i.e., what would have happened to the patient if the doctor had made a different choice. In order to solve this causal question, we use the g-formula with proper assumptions to estimate physiologic trajectories and treatment responses from observed data. To demonstrate this we model blood pressure and heart rate for patients in the intensive care unit (ICU) and estimate their responses to six types of treatments that are used in their management. These two signals are among the most commonly used vital signs in the ICU and are critical for identifying life-threatening conditions like septic and hemorrhagic shock. To model the signal with treatment response from observed data, we use two different Bayesian non-parametric (BNP) methods to build the estimator. BNP are known to have an extremely flexible functional form, which helps to overcome the model mis-specification problem and makes the estimator more robust.

Abstract 10: **Long-term Causal Effects in Policy Analysis in "What If?" Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems**, *Toulis* 04:00 PM

In most causal problems we want to evaluate the long-term effects of policy changes but only have access to short-term experimental data. For example, for the long-term effects of minimum wage increase we may only have access to one-year worth of employment data. In this technical note we argue that such conceptual gap between what is to be estimated and what is in the data has not been adequately addressed. To make our criticism constructive we describe our approach in studying multiagent systems and the long-term effects of interventions in such systems. Central to our approach is behavioral game theory, where a behavioral model of how agents act conditional on their latent behaviors is combined with a temporal model of how behaviors evolve.

Abstract 11: **Causal Inference for Recommendation Systems in "What If?" Inference and Learning of Hypothetical and Counterfactual Interventions in Complex Systems**, *Blei* 04:30 PM

We develop a causal inference approach to recommender systems. Observational recommendation data contains two sources of information: which items each user decided to look at and which of those items each user liked. We assume these two types of information come from different models---the exposure data comes from a model by which users discover items to consider; the click data comes from a model by which users decide which items they like. Traditionally, recommender systems use the click data alone (or ratings data) to infer the user preferences. But this inference is biased by the exposure data, i.e., that users do not consider each item independently at random. We use causal inference to correct for this bias. On real-world data, we demonstrate that causal inference for recommender systems leads to improved generalization to new data.

(Joint work with Dawen Liang and Laurent Charlin)

Brains and Bits: Neuroscience meets Machine Learning

Eva L Dyer, Allie Fletcher, Jascha Sohl-Dickstein, Joshua T Vogelstein, Konrad Koerding, Jakob H Macke

Room 211, Sat Dec 10, 08:00 AM

The goal of this workshop is to bring together researchers from neuroscience, deep learning, machine learning, computer science theory, and statistics for a rich discussion about how computer science and neuroscience can inform one another as these two fields rapidly move forward. We invite high quality submissions and discussion on topics including, but not limited to, the following fundamental questions: a) shared approaches for analyzing biological and artificial neural systems, b) how insights and challenges from neuroscience can inspire progress in machine learning, and c) methods for interpreting the revolutionary large scale datasets produced by new experimental neuroscience techniques.

Experimental methods for measuring neural activity and structure have undergone recent revolutionary advances, including in high-density recording arrays, population calcium imaging, and large-scale reconstructions of anatomical circuitry. These developments promise unprecedented insights into the collective dynamics of neural populations and thereby the underpinnings of brain-like computation. However, these next-generation methods for measuring the brain's architecture and function produce high-dimensional, large scale, and complex datasets, raising challenges for analysis. What are the machine learning and analysis approaches that will be indispensable for analyzing these next-generation datasets? What are the computational bottlenecks and challenges that must be overcome?

In parallel to experimental progress in neuroscience, the rise of deep learning methods has shown that hard computational problems can be solved by machine learning algorithms that are inspired by biological neural networks, and built by cascading many nonlinear units. In contrast to the brain, artificial neural systems are fully observable, so that experimental data-collection constraints are not relevant. Nevertheless, it has proven challenging to develop a theoretical understanding of how neural networks solve tasks, and what features are critical to their performance. Thus, while deep networks differ from biological neural networks in many ways, they provide an interesting testing ground for evaluating strategies for understanding neural processing systems. Are there synergies between analysis methods for biological and artificial neural systems? Has the resurgence of deep learning resulted in new hypotheses or strategies for trying to understand biological neural networks? Conversely, can neuroscience provide inspiration for the next generation of machine-learning algorithms?

We welcome participants from a range of disciplines in statistics, applied physics, machine learning, and both theoretical and experimental neuroscience, with the goal of fostering interdisciplinary insights. We hope that active discussions among these groups can set in motion new collaborations and facilitate future breakthroughs on fundamental research problems.

Machine Learning in Computational Biology

Gerald Quon, Sara Mostafavi, James Y Zou, Barbara Engelhardt, Oliver Stegle, Nicolo Fusi

Room 212, Sat Dec 10, 08:00 AM

The field of computational biology has seen dramatic growth over the past few years. A wide range of high-throughput technologies developed in the last decade now enable us to measure parts of a biological system at various resolutions—at the genome, epigenome, transcriptome, and proteome levels. These technologies are now being used to collect data for an ever-increasingly diverse set of problems, ranging from classical problems such as predicting differentially regulated genes between time points and predicting subcellular localization of RNA and proteins, to models that explore complex mechanistic hypotheses bridging the gap between genetics and disease, population genetics and transcriptional regulation. Fully realizing the scientific and clinical potential of these data requires developing novel supervised and unsupervised learning methods that are scalable, can accommodate heterogeneity, are robust to systematic noise and confounding factors, and provide mechanistic insights.

The goals of this workshop are to i) present emerging problems and innovative machine learning techniques in computational biology, and ii) generate discussion on how to best model the intricacies of biological data and synthesize and interpret results in light of the current work in the field. We will invite several leaders at the intersection of computational biology and machine learning who will present current research problems in computational biology and lead these discussions based on their own research and experiences. We will also have the usual rigorous screening of contributed talks on novel learning approaches in computational biology. We encourage contributions describing either progress on new bioinformatics problems or work on established problems using methods that are substantially different from established alternatives. Deep learning, kernel methods, graphical models, feature selection, non-parametric models and other techniques applied to relevant bioinformatics problems would all be appropriate for the workshop. We will also encourage contributions to address new challenges in analyzing data generated from gene editing, single cell genomics and other novel technologies. The targeted audience are people with interest in machine learning and applications to relevant problems from the life sciences, including NIPS participants without any existing research link to computational biology. Many of the talks will be of interest to the broad machine learning community.

Schedule

08:35 AM	**Introduction**	
08:40 AM	**TBA**	*Marchini*
09:25 AM	**Multiple Output Regression with Latent Noise.**	*Gillberg*
09:45 AM	**Predicting Protein Folding by Ultra-Deep Learning.**	*Xu*
10:05 AM	**Dissecting the non-infinitesimal architecture of complex traits using group spike-and-slab priors.**	*Sarkar*
11:00 AM	**Poster Session**	
12:30 PM	**Lunch**	
01:30 PM	**Predicting the impact of rare regulatory variation.**	*Battle*
02:15 PM	**Modelling-based experiment retrieval: A case study with gene expression clustering.**	*Blomstedt*
02:35 PM	**Convolutional Kitchen Sinks for Transcription Factor Binding Site Prediction.**	*Shankar*
03:30 PM	**Modelling cell-cell interactions with spatial Gaussian processes.**	*Arnol*
03:50 PM	**Predicting off-target effects for CRISPR guide design.**	*Listgarten*
04:10 PM	**Beta Tucker decomposition for DNA methylation data.**	*Schein*
04:30 PM	**Deep Learning for Branch Point Selection in RNA Splicing.**	*Dean*
04:50 PM	**Applying Faster R-CNN for Object Detection on Malaria Images.**	*Hung*

05:10 PM	**Deep learning and new technologies in compbio.**
05:55 PM	**Closing**

Neurorobotics: A Chance for New Ideas, Algorithms and Approaches

Elmar Rueckert, Martin Riedmiller

VIP Room, Sat Dec 10, 08:00 AM

Workshop webpage: http://www.neurorobotic.eu

Modern robots are complex machines with many compliant actuators and various types of sensors including depth and vision cameras, tactile electrodes and dozens of proprioceptive sensors. The obvious challenges are to process these high dimensional input patterns, memorize low dimensional representations of them and to generate the desired motor commands to interact in dynamically changing environments. Similar challenges exist in brain machine interfaces (BMIs) where complex prostheses with perceptional feedback are controlled, or in motor neuroscience where in addition cognitive features need to be considered. Despite this broad research overlap the developments happened mainly in parallel and were not ported or exploited in the related domains. The main bottleneck for collaborative studies has been a lack of interaction between the core robotics, the machine learning and the neuroscience communities.

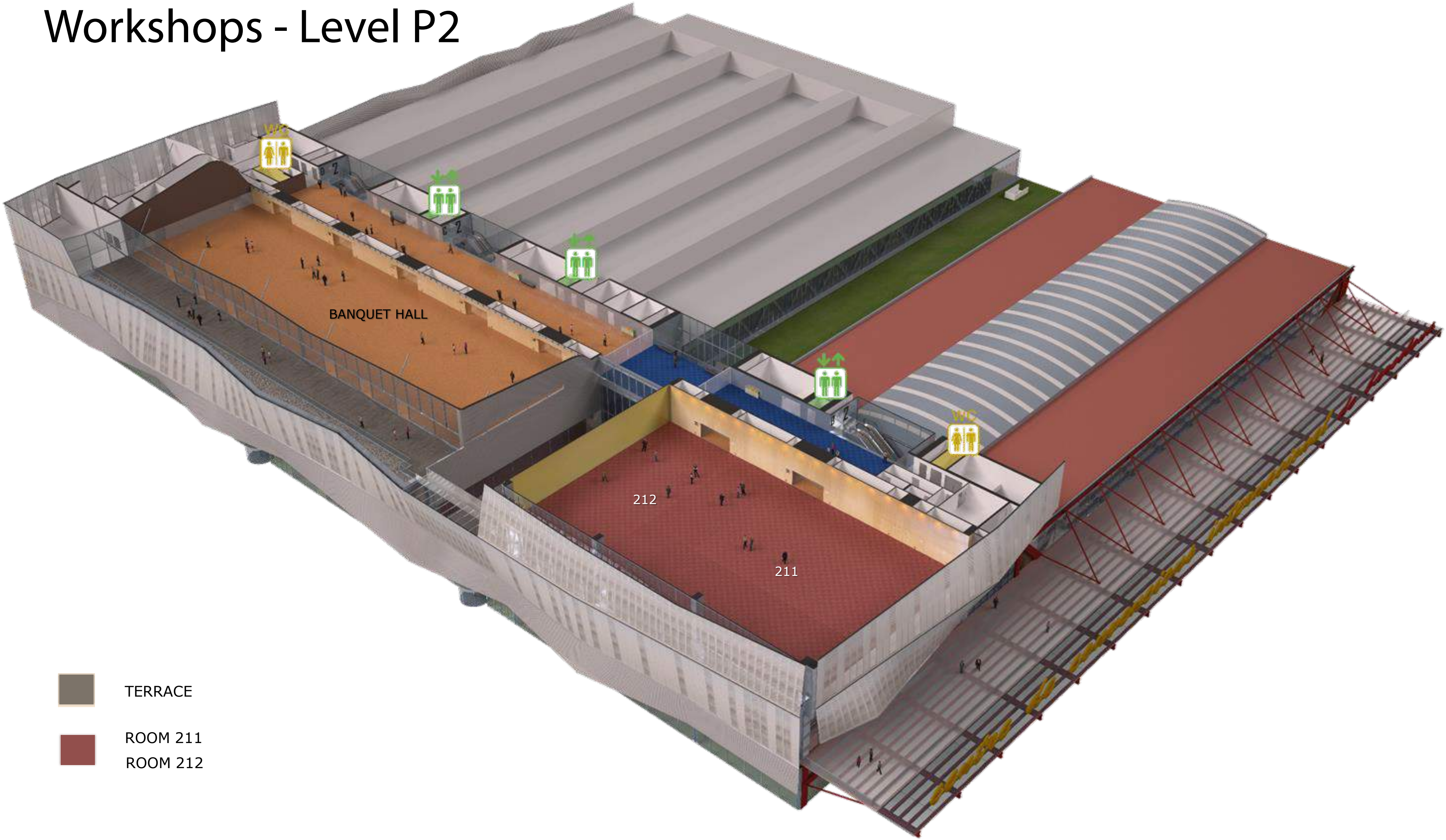
Why is it now just the right time for interactions?

- Latest developments based on deep neural networks have advanced the capabilities of robotic systems by learning control policies directly from the high dimensional sensor readings.
- Many variants of networks have been recently developed including the integration of feedback through recurrent connections, the projection to different feature spaces, may be trained at different time scales and can be modulated through additional inputs.
- These variants can be the basis for new models and concepts in motor neuroscience, where simple feed forward structures were not sufficiently powerful.
- Robotic applications demonstrated the feasibility of such networks for real time control of complex systems, which can be exploited in BMIs.
- Modern robots and new sensor technologies require models that can integrate a huge amount of inputs of different dimension, at different rates and with different noise levels. The neuroscience communities face such challenges and develop sophisticated models that can be evaluated in robotic applications used as benchmarks.
- New learning rules can be tested on real systems in challenging environments.

Topics:

- Convolutional Networks and Real-time Robotic and Prosthetic applications
- Deep Learning for Robotics and Prosthetics
- End-to-End Robotics / Learning
- Feature Representations for Big Data
- Movement Representations, Movement Primitives and Muscle Synergies
- Neural Network Hardware Implementation, Neuromorphic Hardware
- Recurrent Networks and Reservoirs for Control of high dimensional systems
- Reinforcement Learning and Bayesian Optimization in Neural Networks from multiple reward sources
- Sampling Methods and Spiking Networks for Robotics
- Theoretical Learning Concepts, Synaptic Plasticity Rules for Neural Networks

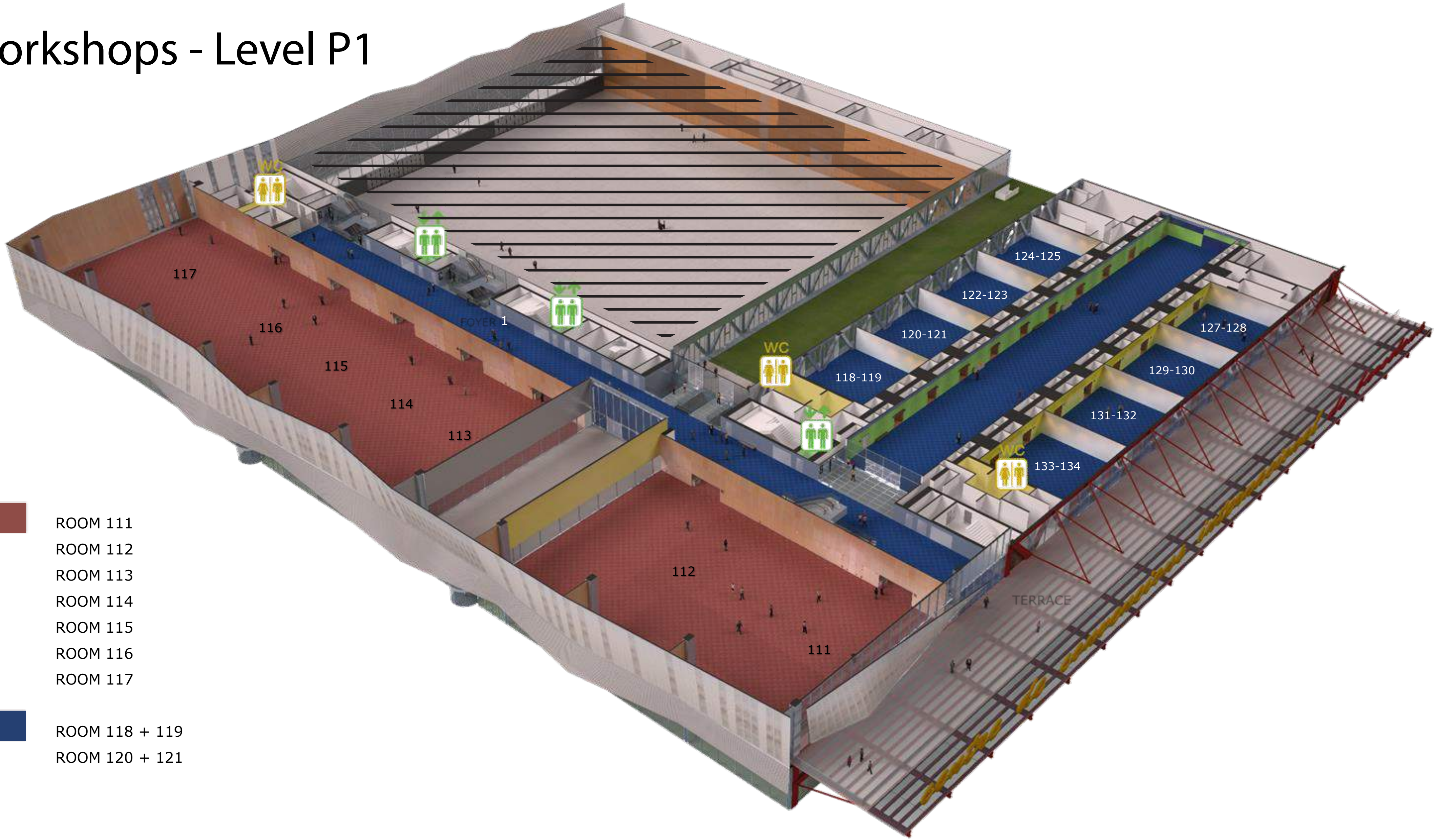
Workshops - Level P2



TERRACE

ROOM 211
ROOM 212

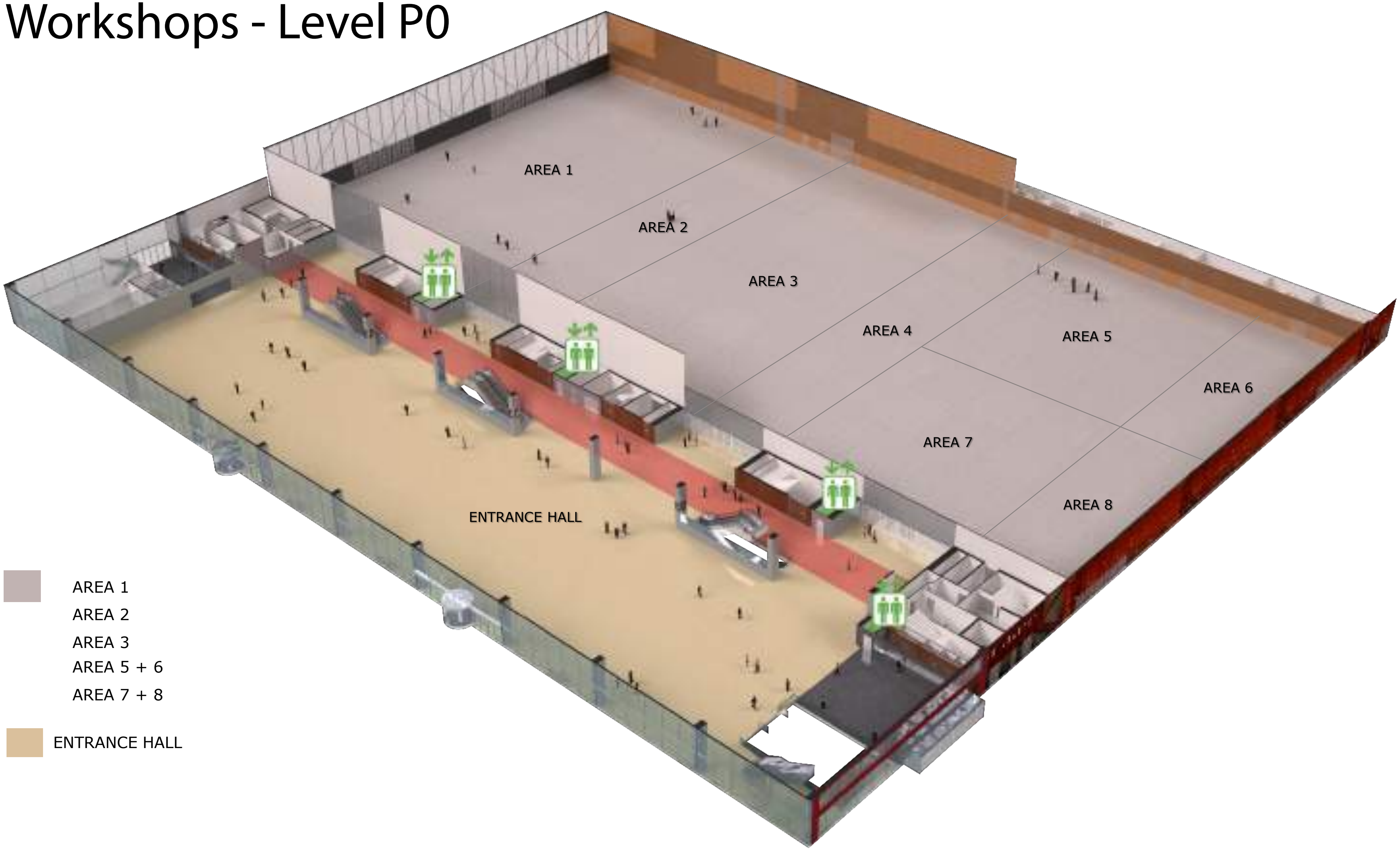
Workshops - Level P1



ROOM 111
ROOM 112
ROOM 113
ROOM 114
ROOM 115
ROOM 116
ROOM 117

ROOM 118 + 119
ROOM 120 + 121

Workshops - Level P0



AREA 1
AREA 2
AREA 3
AREA 5 + 6
AREA 7 + 8

ENTRANCE HALL