Dear Reviewer #1:

> the result is far from being unexpected.

We believe that some readers may find unexpected results in our work. For example, the authors of the literature [8] asked the following question in their conclusion: "For the problem of online ranking,... In particular, is the optimal regret $\Theta(n^2\sqrt{T})$ in this setting?". To this question, our Theorem 1 gives the following answer: "No, the optimal regret is $\Theta(n^{5/2}\sqrt{T})$".

> its significance is unclear because neither the results nor the techniques are particularly surprising in the context of previous works.

We consider that our analysis includes nontrivial techniques, especially in the construction of hard instances. As presented in lines 110-118, we carefully control the parameters of distributions to obtain desired lower bounds.

Dear Reviewer #2:

> I think the writing could be improved. In particular, I found reading this paper confusing because there were several different results, and the introduction kept skipping between results and proof ideas/intuition and descriptions of new problems and so on. I think for this paper it might be better to first clearly list out all the problem types that will be considered, then list out all the new results, and then finally say a bit of proof summary/preview.

Thanks for your suggestion. We will modify the manuscript so that all results and techniques are clarified.

> Theorem 2 says "mutiple"

Thanks for pointing out the typo. We will fix it in the revised version.

> The definition of $\hat{\ell}$ isn't really given, it's just implicit that $\hat{\ell}$ corresponds to $\hat{a}$. Better to be explicit.

In (8), we define a *class* of distributions D_{a^*} *for all* $a^* \in \{0,1\}^d$, and $\hat{\ell}$ is defined to be a random vector following the distribution $D_{\hat{a}}$ given by (8). We will describe this explicitly in the revised manuscript. Thanks for your suggestion.

> As the authors suggest, it would be great to get the correct rate up to an absolute constant.

We are now tackling this problem, but have not found the answer yet. We believe that our lower bound is tight and that one can improve the upper bound to find the correct rate. To achieve this, however, a novel idea seems to be needed as the upper bound has not been improved since 2009.

Dear Reviewer #3:

> there seems to be a gap in the analysis: In the derivation sequence between line 248 and line 249, I do not understand the last equality (after the inequality). It does not look to be true to me, since it is true only when the first time of the LHS of the equation, (P'(i)-P(i))/P(i) is set to zero, but this is not the case.
> I expect the authors could fill the gap I mentioned above, and explain how the analysis should proceed, then I would be happy to raise my score.

Our analysis proceeds as follows:

$$-\sum_{i=0}^{k} P(i)\left(\frac{P'(i)-P(i)}{P(i)} - 2\left(\frac{P'(i)-P(i)}{P(i)}\right)^2\right) = -\sum_{i=0}^{k} P(i)\frac{P'(i)-P(i)}{P(i)} + 2\sum_{i=0}^{k} P(i)\left(\frac{P'(i)-P(i)}{P(i)}\right)^2$$
$$= -\sum_{i=0}^{k}(P'(i)-P(i)) + 2\sum_{i=0}^{k}\frac{(P'(i)-P(i))^2}{P(i)} = -\sum_{i=0}^{k} P'(i) + \sum_{i=0}^{k} P(i) + 2\sum_{i=0}^{k}\frac{(P'(i)-P(i))^2}{P(i)}.$$

Since P and P' are probability mass functions over $\{0,1,\ldots,k\}$, we have $\sum_{i=0}^{k} P(i) = \sum_{i=0}^{k} P'(i) = 1$. Hence, we have $-\sum_{i=0}^{k} P'(i) + \sum_{i=0}^{k} P(i) = -1+1 = 0$. By substituting this to the above displayed equation, we obtain $-\sum_{i=0}^{k} P(i)\left(\frac{P'(i)-P(i)}{P(i)} - 2\left(\frac{P'(i)-P(i)}{P(i)}\right)^2\right) = 2\sum_{i=0}^{k}\frac{(P'(i)-P(i))^2}{P(i)}$. We hope the above discussion fills the gap you mentioned. In the revised version, we will clarify how this equality is derived.