Doctopic: Algorithms, automata, complexity and games ARTICLE IN PRESS TCS:11433

Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

On minimaxity of Follow the Leader strategy in the stochastic setting ☆

Wojciech Kotłowski [1]

Institute of Computing Science, Poznań University of Technology, Piotrowo 2, 60-965 Poznań, Poland

ARTICLE INFO

Article history:
Received 30 January 2017
Received in revised form 15 September 2017
Accepted 9 October 2017
Available online xxxx

Keywords:
Online learning
Prediction with expert advice
Follow the Leader
Risk
Regret
Pseudo-regret
Stochastic data
Minimax algorithm
Exponential families

ABSTRACT

We consider the setting of prediction with expert advice with an additional assumption that each expert generates its losses i.i.d. according to some distribution. We first identify a class of "admissible" strategies, which we call permutation invariant, and show that every strategy outside this class will perform not better than some permutation invariant strategy. We then show that when the losses are binary, a simple Follow the Leader (FL) algorithm is the minimax strategy for this game, where minimaxity is simultaneously achieved for the expected regret, the pseudo-regret, and the excess risk. Furthermore, FL has also the smallest regret, pseudo-regret, and excess risk over all permutation invariant prediction strategies, simultaneously for all distributions over binary losses. We generalize these minimax results to the case in which each expert generates its losses from a distribution belonging to a one-dimensional exponential family, as well as to the case of loss vectors generated jointly from a multinomial distribution. We also show that when the losses are in the interval $[0, 1]$ and the learner competes against all distributions over $[0, 1]$, FL remains minimax only when an additional trick called "loss binarization" is applied.

1. Introduction

In the game of prediction with expert advice [2,3], the learner sequentially decides on one of K experts to follow, and suffers loss associated with the chosen expert. The difference between the learner's cumulative loss and the cumulative loss of the best expert is called *regret*. The goal is to minimize the regret in the worst case over all possible loss sequences. A prediction strategy which achieves this goal (i.e., minimizes the worst-case regret) is called *minimax*. While algorithms such as Weighted Majority/Hedge [4–6] or Follow the Perturbed Leader [7] guarantee the optimal worst-case regret in the asymptotic sense (i.e., their regrets grow at the optimal rate), there is no known exact solution to the minimax problem in the general setting. Still, it is possible to derive minimax algorithms for some special variants of this game: when the losses follow from evaluating binary predictions on binary labels [2,3], for binary losses with fixed loss budget [8], and when $K = 2$ [9].

☆ A preliminary version of this paper appeared at 27th International Conference on Algorithmic Learning Theory (ALT 2016) [1]. In this journal version we extended the background discussion and added two new sections on exponential family model (Section 4) and dependent experts (Section 6).

E-mail address: wkotlowski@cs.put.poznan.pl.

[1] The author was supported by the Polish National Science Centre under grant No. 2016/22/E/ST6/00299.

https://doi.org/10.1016/j.tcs.2017.12.030
0304-3975/© 2017 Published by Elsevier B.V.

Interestingly, all these minimax algorithms share a similar strategy of playing against a maximin adversary which assigns losses uniformly at random. They also have the *equalization* property: all data sequences lead to the same value of the regret. While this property makes them robust against the worst-case sequence, it also makes them over-conservative, preventing them from exploiting the case, when the actual data is not adversarially generated.

In this paper, we drop the analysis of worst-case performance entirely, and explore the minimax principle in a more constrained setting, in which the adversary is assumed to be *stochastic*. In particular, we associate with each expert k a fixed distribution P_k over loss values, and assume the observed losses of expert k are generated independently from P_k. The goal is then to determine the minimax algorithm under these stochastic assumptions. The motivation behind studying such a setting is in its practical usefulness: the data encountered in practice are rarely adversarial and can often be modeled as generated from a fixed (yet unknown) distribution (for instance, selecting the best classifier from a set of already trained candidates based on data gathered in an online manner).

We immediately face two difficulties here. First, due to stochastic nature of the adversary, it is no longer possible to follow standard approaches of minimax analysis, such as backward induction [2,3] or sequential minimax duality [10,9], and we need to resort to a different technique. We define the notion of *permutation invariance* of prediction strategies. This let us identify a class of "admissible" strategies (which we call permutation invariant), and show that every strategy outside this class will perform not better than some permutation invariant strategy. Secondly, while the regret is a single, commonly used performance metric in the worst-case setting, the situation is different in the stochastic case. We know at least three potentially useful metrics in the stochastic setting: the *expected regret*, the *pseudo-regret*, and the *excess risk* [11], and it is not clear, which of them should be used to define the minimax strategy.

Fortunately, it turns out that there exists a single strategy which is minimax with respect to all three metrics simultaneously. In the case of *binary* losses, which take out values from $\{0, 1\}$, this strategy turns out to be the *Follow the Leader* (FL) algorithm, which chooses an expert with the smallest cumulative loss at a given trial (with ties broken randomly). Interestingly, FL is known to perform poorly in the worst-case, as its worst-case regret will grow linearly with T [3]. On the contrary, in the stochastic setting with binary losses, FL has the smallest regret, pseudo-regret, and excess risk over all permutation invariant prediction strategies, *simultaneously for all distributions over binary losses!* We later show that all these minimax properties of FL strategy generalize to the case, in which each expert generates its losses from a distribution belonging to a one-dimensional exponential family (e.g., Gaussian, Bernoulli, Poisson, gamma, geometric, etc.), and the previously considered case of losses from $\{0, 1\}$ becomes a special case of the Bernoulli family.

Furthermore, we also show that the optimality of FL strategy breaks down in the case of losses in the range $[0, 1]$, in which each expert generates losses from an *arbitrary* distribution over $[0, 1]$. Here, FL is provably suboptimal. However, by applying *binarization trick* to the losses [12], i.e. randomly setting them to $\{0, 1\}$ such that the expectation matches the actual loss, and using FL on the binarized sequence (which results in the *binarized FL* strategy), we obtain the minimax strategy in this setup.

We finally consider the case of dependent experts, i.e. when the losses are i.i.d. between trials, but not necessarily between experts. While the general case turns out to be hard to approach, and our methods based on permutation invariance fail, we are able to analyze the simplest variant of dependent experts, where the loss vectors follow multinomial distribution, i.e. only a single expert gets loss in a given trial. We show that the FL strategy retains the minimax properties analogous to those given for binary losses and independent experts.

We note that when the excess risk is used as a performance metric, our setup falls into the framework of statistical decision theory [13,14], and the question we pose can be reduced to the problem of finding the minimax decision rule for a properly constructed loss function, which matches the excess risk on expectation. In principle, one could try to solve our problem by using the complete class theorem and search for the minimax rule within the class of (generalized) Bayesian decision rules. We initially followed this approach, but it turned out to be futile, as the classes of distributions we are considering are large (e.g., all distributions in the range $[0, 1]$), and exploring prior distributions over such classes becomes very difficult. On the other hand, the analysis presented in this paper is relatively simple, and works not only for the excess risk, but also for the expected regret and the pseudo-regret. To the best of our knowledge, both the results and the analysis presented here are novel.

We also note that there has recently been much work dedicated to combine almost optimal worst-case performance with good performance on "easy" (e.g., stochastic) sequences [15,12,16–18]. These methods, however, are motivated from different principles than the minimax principle, and their analysis is tangential to the topic of this work.

Follow the Leader strategy in the stochastic setting has already been analyzed extensively in the past. It is known that when the losses of experts are generated i.i.d., FL performs very well in terms of the expected regret [15,12,19]. Furthermore, the asymptotically optimal Upper-Confidence-Bound (UCB) algorithm used in the stochastic multi-armed bandit setting [3,20] would reduce to FL in our ("full information") setup, as confidence intervals maintained by UCB for each expert would all be of the same size. If one uses excess risk as a performance metric, the setup considered here reduces to a simple scenario of learning in the finite hypothesis class in statistical learning theory, where it is known that Empirical Risk Minimization (equivalent to FL strategy) achieves $O(\sqrt{\log K/T})$ excess risk, which can be shown to be tight [21]. This immediately (by summing over trials) gives $O(\sqrt{T \log K})$ bound on the pseudo-regret of FL, and the same bound on the regret of FL, by using the fact that the difference between the pseudo-regret and the expected regret is independent of prediction strategy and lower than the expected regret of any online learning algorithm (which is, again, of order $O(\sqrt{T \log K})$). All these bounds hold even for dependent experts. In fact, since the tight lower bound $\Omega(\sqrt{T \log K})$ on the regret in the adversarial expert