Proceedings of Machine Learning Research vol 134:1–1, 2021 34th Annual Conference on Learning Theory

Online Learning from Optimal Actions

Omar Besbes OB2105@GSB.COLUMBIA.EDU
Yuri Fonseca YFONSECA23@GSB.COLUMBIA.EDU
Columbia Business School

Ilan Lobel ILOBEL@STERN.NYU.EDU
NYU Stern School of Business

Editors: Mikhail Belkin and Samory Kpotufe

Abstract

The classical framework in online learning and online optimization is one in which the decision-maker takes an action in each period and then receives some feedback in the form of a payoff or loss. However, there are important situations where a machine learning algorithm might not have access to the loss caused by a decision, but it does have access after-the-fact to the decision an expert (or oracle) would have taken in the situation. This problem is called imitation learning. To formalize this problem, we consider an online contextual optimization problem where, in each period t, a decision-maker receives a feasible set $\mathcal{X}_t \subseteq \mathbb{R}^n$ and a context function $f_t : \mathbb{R}^n \to \mathbb{R}^d$, and must choose some $x_t \in \mathcal{X}_t$ in order to minimize $f_t(x_t)'c^\star$, where $c^\star \in C_0$ is an unknown cost vector and C_0 is an initial knowledge set within $\mathbb{R}^d$. The decision-maker can gain information about $c^\star$ over time by observing what would have been an optimal decision to take in period t: the expert action $x_t^\star$. We do not make distributional assumptions on the feasible sets, context functions or the true cost vector. Our objective is to minimize the worst-case regret over a time horizon T, which is defined as the difference between the losses we incurred with the losses incurred by the expert. Our analysis starts by considering the one-period problem. For the latter, we establish that the optimal regret is a function of what we call the uncertainty angle of the initial knowledge set C_0, $\alpha(C_0)$, which is a measure of how large a revolution cone do we need to contain C_0. The axis of such a revolution cone is what's called the circumcenter of C_0. Furthermore, our first theorem shows that the one-period policy that guarantees an optimal worst-case regret bound is to treat the circumcenter of C_0 as if its was the true cost. Next, we consider a version of the multi-period problem where the uncertainty angle of C_0 is bounded by $\pi/2$. We show that a naïve application of the circumcenter policy fails because this policy might, at the same time, incur regret and fail to learn any new information. This is due to the potentially complex geometry of the knowledge set. To address this issue, we regularize our knowledge sets by replacing them with ellipsoidal cones and we construct an algorithm that has regret bounded by $\mathcal{O}(d^2 \ln(T \tan \alpha(C_0)))$. The final part of our paper deals with the general case. The key idea is to establish that it is possible to always maintain a subspace such that the projection of our knowledge set onto it has a bounded uncertainty angle. We then extend our algorithm to have two types of updates: periods where we perform an ellipsoidal cone update, and periods where we add an extra dimension to our subspace. First, we need to robustify the ellipsoidal cone cuts to account for potential error introduced by the projection onto a subspace. Second, we need to construct a new ellipsoidal cone every time we increase the dimension of the subspace, and we need to argue that this new ellipsoidal cone has a bounded uncertainty angle. With this algorithm, we are able to obtain our universal $\mathcal{O}(d^4 \ln T)$ regret bound regardless of the initial knowledge set. [1]

[1] Full version is available at [arXiv reference, 2106.14015] under the title “Contextual Inverse Optimization: Offline and Online Learning.”