MSc Project: Life-long Online Learning

Supervisor: Wouter M. Koolen
Keywords: Sequential Decision Making, Online Learning, Expert Setting

Background: A fundamental task in sequential decision making is Prediction with Expert Advice, also known as Decision Theoretic Online Learning or the Hedge Setting (Freund and Schapire, 1997). Here a decision maker plays a game of T rounds, picking one of K actions every round, and incurring a bounded loss. The objective is the regret compared the best fixed action in hindsight. The minimax regret is known to scale as $\sqrt{T \ln K}$. Many variations and extensions of the classic Hedge algorithm have been developed, featuring refined notions of time (de Rooij et al., 2014), and adaptivity to the complexity of the action set (Koolen and van Erven, 2015), etc.

Academic Content: This project explores what can be achieved in a long sequence of M Hedge interactions. It could be considered an investigation into "life-long-learning". Of course, one may play the M games with overall worst-case summed regret at most

$$M\sqrt{T \ln K}$$

The question raised here is whether one can adapt to the frequency distribution of the games' best actions.

1. We first ask what one can achieve if the distribution of winning actions were known. Say action k is the best in fraction p_k of the games. In this sense $\vec{p}$ could be considered a *prior distribution* on actions. Can one play the M games to achieve regret bounded by something of the form

$$M\sqrt{T \operatorname{cost}(\vec{p})}$$

and what would be the right cost function? A first, intuitive suggestion could be the Shannon entropy, but this is shown admissible but suboptimal by Koolen (2013). There are likely deep connections to the concept of mixability introduced by Vovk (1998).

2. We then ask what one can achieve if the fraction p is not known a priori. How can one still adapt to it, and how can one quantify the cost for learning p?

3. Finally, we ask the same question in the multi-scale extension of the setting, where recent adaptivity breakthroughs were made by P´erez-Ortiz and Koolen (2022).

The project is primarily expected to develop new theory: results include algorithms with performance guarantees as well as lower bounds. The project may benefit significantly from exploratory programming skills.

References

De Rooij, S., T. van Erven, P. Gr¨unwald, and W. M. Koolen (Apr. 2014). “Follow the Leader If You Can, Hedge If You Must”. In: Journal of Machine Learning Research 15, pp. 1281 1316.

Freund, Y. and R. E. Schapire (1997). “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”. In: Journal of Computer and System Sciences 55, pp. 119–139.

Koolen, W. M. (Dec. 2013). “The Pareto Regret Frontier”. In: Advances in Neural Information Processing Systems (NeurIPS) 26, pp. 863–871.

Koolen, W. M. and T. van Erven (June 2015). “Second-order Quantile Methods for Experts and Combinatorial Games”. In: Proceedings of the 28th Annual Conference on Learning Theory (COLT), pp. 1155–1175.

P´erez-Ortiz, M. and W. M. Koolen (Dec. 2022). “Luckiness in Multiscale Online Learning”. In: Advances in Neural Information Processing Systems (NeurIPS) 35. Accepted.

Vovk, V. (1998). “A Game of Prediction with Expert Advice”. In: Journal of Computer and System Sciences 56.2, pp. 153–173.