Stochastic Linear Bandits An Empirical Study

Students:

Antoine L tbd@ens-paris-saclay.fr

Benjamin Deporte benjamin.deporte@ens-paris-saclay.fr

Lecturer: Claire Vernade, Contact me on Slack if anything looks weird, or find my email on my website

Stochastic Linear Bandits are a great way to model contextual sequential decision making problems. We saw them in class and you can read further in the Bandit book (Lattimore and Szepesvári, 2020). In this assignment, your task will be to implement them in order to discuss concretely their performance on simulated environments. The 'companion' notebook is attached in this Overleaf. it contains a boiler plate for the experiments and the main questions you will need to answer below. On my website you can find a sample notebook for K-armed bandits (not linear) that may help you get started.

Hard constraint: Your report can be at most 3 pages + a possible appendix if you decide to also address the Bonus part at the end of the notebook. Thus, if you need to include code snippets or figures, you will have to carefully choose which ones. You are welcome to (and you should!) delete anything I wrote here and below, as long as you re-introduce enough context for your answers, plots and discussions. The clarity of your presentation will be taken into account in the evaluation. Each additional page (> 3 for the main part, appendix and references do not count) will count -1 on your grade.

Problem 1: Linear Epsilon Greedy (/5)

- Complete the implementation of the Linear Bandit environment and the action generation function;

- Implement Linear Epsilon Greedy and test it on a simple problem of your choice. Describe your chosen benchmark, report your results and comment. Is it a strong baseline?

- Discuss the complexity of the matrix inversion step and propose a better, incremental update. Report the gain in runtimes that you observed as a function of d.

Problem 2: LinUCB and LinTS (/5)

- Implement LinUCB and LinTS.

- For Thompson Sampling, what is the posterior at time t?

- We want to know if there is a better algorithm out of all the ones you implemented. Propose an experiment (describe it) and report your results and conclusions.

References

Lattimore, T. and Szepesvari, C. (2017). The end of optimism? an asymptotic analysis of finite-armed linear bandits. In *Artificial Intelligence and Statistics*, pages 728–737. PMLR.

Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

A   Bonus section: The role of the action set

This last part allows you to study how certain actions sets can be hard for LinUCB. I propose some code to highlight this phenomenon, that was unveiled by Lattimore and Szepesvari (2017). Follow the instructions in the notebook and report your results and comments here. *This part can count up to (+2) points, meaning that the maximal grade is (12/10) and could compensate missed points on the Quiz or on the project.*