Example Project Ideas, Topics, etc
DATA 37200: Learning, Decisions, and Limits (Winter'25)

Note: it is likely we will update this document with additional references and example topics, in which case we will update the copy on the website.

Suggested Project Topics (biased due to the instructor's knowledge, experience, etc)

Importantly, we would like to reiterate that we strongly encourage you to identify your own topics for the project. The suggested topics below is supposed to serve as examples to stimulate your thinking and also as your second resort, if you were not able to find your own project topics. Moreover, please feel free to discuss with the instructors/TA for suggestions/comments on your ideas and for additional sources of references.

NOTE: papers below are arbitrarily chosen and given to give EXAMPLES of recent research efforts in a theoretical direction and hopefully serve as helpful jumping off points into the broader literature. The examples given are not at all intended to be fully "representative" of all the work in each area, and we are unfortunately not aware of all important works in the massive literature related to these topics nor are we experts in all areas of research. For whatever specific direction you choose to work on, you need to search and try to understand the literature yourself. (Some basic tips: try looking at papers cited or cited by other papers using tools like google scholar, etc.)

1. Reinforcement learning, computational complexity, and computational-statistical gaps. See for example the recent paper `https://arxiv.org/pdf/2404.03774` and several previous works discussed in related work.

2. Partially Observable MDPs. See for example the recent papers `https://arxiv.org/abs/2206.03446`, `https://arxiv.org/abs/2307.02884`, and lots of previous work, for example as discussed there.

3. Various types of robustness questions in RL, Control Theory, etc. As phrased this includes a extremely massive literature. See `https://skoge.folk.ntnu.no/puublications_others/1978_doyle_margins-lqg-there-are-none.pdf` for an interesting classic paper in control theory. (See also H^∞ control, etc.)

4. RL theory with function approximation: statistical and computational considerations. Another massive literature. See e.g. `https://arxiv.org/pdf/2406.11640`, `https://proceedings.mlr.press/v162/zhang22aa/zhang22aa.pdf`, `https://www.jmlr.org/papers/volume25/22-0687/22-0687.pdf`, `https://arxiv.org/pdf/2112.14195`, `https://arxiv.org/abs/2102.02049`, for a few recent papers with some pointers to some subset of the literature...

5. Multiagent RL, another massive literature... https://arxiv.org/pdf/2006.12007, ...

6. Pessimism and Offline RL... https://arxiv.org/abs/2205.10671, https://arxiv.org/abs/2012.15085, https://yuxinchen2020.github.io/publications/Pessimistic_Qlearning_ICML.pdf, https://arxiv.org/abs/2106.06926, ...

7. Imitation learning, learning from demonstrations, etc. Some arbitrarily chosen recent works: https://arxiv.org/pdf/2307.14619, https://arxiv.org/abs/2407.15007v1, https://arxiv.org/abs/2312.00054, ...

8. Thompson sampling for bandits and related ideas in RL. See e.g. https://arxiv.org/abs/1611.06534, https://arxiv.org/pdf/1906.02870, https://arxiv.org/pdf/2308.07843.

9. ϵ-Greedy strategies in bandits and RL. E.g. https://arxiv.org/pdf/2206.09421, lecture notes https://arxiv.org/pdf/2312.16730, ...

10. Decision-Estimation coefficient, information-theoretic ideas in RL theory. Relevant lecture notes https://arxiv.org/pdf/2312.16730, an example recent work https://arxiv.org/abs/2209.11745, ...

11. Connections between martingales and online learning. E.g. https://arxiv.org/pdf/1803.07617, ... (I believe Vladimir Vovk's work may be quite relevant to this topic, so maybe looking at his books would be worthwhile.)

12. Actor-critic, policy gradient, etc related theory... https://epubs.siam.org/doi/full/10.1137/23M1560215, https://arxiv.org/pdf/2110.11280, ...

13. Online learning and RL in linear control, related topics. There is a lot of work in this area involving phrases like "system level synthesis". See e.g. https://arxiv.org/abs/1710.01688 and papers citing this one.

14. Connections with state space models. State space models are an alternative to transformers partialy inspired by linear control. There is a ton of work on them recently. See e.g. https://arxiv.org/pdf/2303.03982 for an empirical work using state space models in an MDP context.

15. Connections between empirical bayes and online/sequential decision making. https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-36/issue-3/Sequential-Co 10.1214/aoms/1177700060.full ("Sequential Compound Estimators ", an old paper in annals of mathematical statistics). If you are interested in this topic, you would probably be interested to talk to Professor Ignatiadis who suggested this paper.

16. Alternatives models for online decision making besides regret (e.g. competitive ratio). This includes topics like the "k-server problem" and so far is mostly very theoretical with interesting connections to online learning. (For any theoretical topic, it would be interesting to see if it could be used/modified to serve some practical application.) See e.g. https://arxiv.org/pdf/1905.11968, https://arxiv.org/abs/1711.01085, https://arxiv.org/abs/2211.05753.

17. Recent research relates to mirror descent, natural gradient descent, etc e.g. https://arxiv.org/pdf/2004.01025, https://par.nsf.gov/servlets/purl/10232280, https://arxiv.org/abs/2007.06558.

18. Variance reduction in RL. See e.g. `https://www.jmlr.org/papers/volume5/greensmith04a/greensmith04a.pdf` for an older paper, papers citing this one, etc.

19. Reward shaping in RL. See e.g. `https://people.eecs.berkeley.edu/~russell/papers/icml99-shaping.pdf` for an old paper.

20. RL algorithms for fine-tuning LLMs, with RLHF [OWJ+22] and DPO [RSM+24] as main representatives, and numerous papers citing these two.

21. Adversarial bandits [ACBFS02], which has an extensive literature as well – easiest way will be look at highly-cited follow-up papers that cited [ACBFS02].

22. Multi-agent bandit learning, and convergence to equilibrium.

    There is an extensive literature on this frontier for zero-sum games [PP16, MJS19], congestion games [CXFD22], dominance-elimination solvable games [WXY22, WKBJ22], monotone games [COZ22]. However, many problems remain open. For example, the regret bound of [CXFD22] leaves significant room for improvement. An interesting open problem from [WXY22] is whether there exists a no-regret learning algorithm that provably converges to rationalizable equilibrium in two-player games.

23. Decisions + Languages, with applications to solving real games such as *Diplomacy* [FBB+22], or by letting language agents to debate in order to arrive at better reasoning capabilities

    [DVJ+24]

References

[ACBFS02] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.

[COZ22] Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. Finite-time last-iterate convergence for learning in multi-player games. *Advances in Neural Information Processing Systems*, 35:33904–33919, 2022.

[CXFD22] Qiwen Cui, Zhihan Xiong, Maryam Fazel, and Simon S Du. Learning in congestion games with bandit feedback. *arXiv preprint arXiv:2206.01880*, 2022.

[DVJ+24] Tim R Davidson, Veniamin Veselovsky, Martin Josifoski, Maxime Peyrard, Antoine Bosselut, Michal Kosinski, and Robert West. Evaluating language model agency through negotiations. *arXiv preprint arXiv:2401.04536*, 2024.

[FBB+22] Meta Fundamental AI Research Diplomacy Team (FAIR)†, Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074, 2022.

[MJS19] Eric V Mazumdar, Michael I Jordan, and S Shankar Sastry. On finding local nash equilibria (and only local nash equilibria) in zero-sum games. *arXiv preprint arXiv:1901.00838*, 2019.

[OWJ+22] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

[PP16] Christos Papadimitriou and Georgios Piliouras. From nash equilibria to chain recurrent sets: Solution concepts and topology. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 227–235, 2016.

[RSM+24] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36, 2024.

[WKBJ22] Yuanhao Wang, Dingwen Kong, Yu Bai, and Chi Jin. Learning rationalizable equilibria in multiplayer games. 2022.

[WXY22] Jibang Wu, Haifeng Xu, and Fan Yao. Multi-agent learning for iterative dominance elimination: Formal barriers and new algorithms. In Po-Ling Loh and Maxim Raginsky, editors, *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 543–543. PMLR, 02–05 Jul 2022.