Topics in Reinforcement Learning

Time: 3:30pm - 4:45pm

Instructor: Yuhua Zhu
Office: MS 8935
Office Hours: Tue, 4:45pm - 5:45pm
Course website: `https://www.yuhuazhu.org/stats219`

Course Outline

1. Multi-armed bandits

- Upper Confidence Bound & its theoretical guarantee
- Thompson Sampling, Bayesian Optimal Policy

2. Markov Decision Process

- Bellman equation, Temporal Difference & its theoretical guarantees, TD(λ)
- Value-based algorithms: SARSA, Q-learning & its theoretical guarantees
- Policy-based algorithms: Policy Gradient & its theoretical guarantee, actor-critic, TRPO, PPO
- Optimization-based algorithms: Double sampling problem, Primal-dual, BFF

3. Continuous-time RL

- Stochastic Optimal Control, Hamilton-Jacobi-Bellman Equation
- Linear quadratic regulator

Student Evaluation

1. Course Project: 90%**:**

Milestone 1: mid-term presentation (10%)
Milestone 2: end-term presentation (30%)
Milestone 3: end-term report (50%).

2. Student Participation: 10%

1. Course Project: Groups of **two students** will work on a research problem that is relevant to the course.

- The first component of the project is a **10-minute mid-term presentation** on a paper of your choice and is relevant to the course.
  - There are some reference papers that you can choose from. Feel free to choose other papers you are interested in, but please send your choice of paper to the instructor for approval in advance.
    Reference papers: https://ucla.box.com/s/s7mtt2zip58qr4k2zpkemhpfa77umy7u
  - Please sign up via a google sheet (will be created later) **by Feb 5th**.
  - It has to be an in-person slides presentation, which is expected to happen on Feb 11 and Feb 13. During the presentation, each group will receive questions from the instructor and the rest of the class.
- The second component of the project is a **15-minute end-term presentation** on a research project.
  - The research problem should be aligned with your choice of paper for the mid-term presentation.
  - It has to be an in-person slides presentation, which is expected to happen during the last two weeks of the quarter. During the presentation, each group will receive questions from the instructor and the rest of the class.
- The last component of the project will be an **end-of-term report**. It has to be written in a machine-learning conference format (e.g., NeurIPS), and has a **5-page** limit (reference excluded). The final report is due at **11:59 pm, Wednesday, Mar 21, 2025,** by uploading to Gradescope.

2. Student Participation: The students are expected to actively participate in the course with questions and suggestions, and are expected to ask questions during other teams' presentations.

Key References

[1] "***Bandit Algorithm***"
by Tor Lattimore and Csaba Szepesvári
Source: `https://tor-lattimore.com/downloads/book/book.pdf`

[2] "***Algorithms for Reinforcement Learning***"
by Csaba Szepesvári
Source: `https://sites.ualberta.ca/~szepesva/papers/RLAlgsInMDPs.pdf`

[3] "***Stochastic Optimal Control: The Discrete-Time Case***"
by Dimitri P. Bertsekas and Steven E. Shreve
Source: `https://web.mit.edu/dimitrib/www/SOC_1978.pdf`