Sequential Decision Making

Time: TuTh 2pm - 3:20pm

Instructor: Yuhua Zhu
Location: AP&M 7218
Office: AP&M 5618
Office Hours: Thu 3:20pm - 4:20pm or by appointment

Course Outline

RL provides the key to enabling machines to make intelligent decisions, learn from their experiences, and optimize their actions in dynamic and uncertain environments. From recommendation systems tailored to individual preferences to self-driving cars navigating complex traffic scenarios, RL is at the heart of enabling these remarkable feats. This course will delve into the depth and breadth of RL, covering a wide range of topics from theoretical foundations to practical algorithmic implementations, including but not limited to multi-armed bandits, markov decision process, stochastic optimal control.

Topics include:

1. Multi-armed bandits:

- Upper Confidence Bound
- Regret analysis for UCB
- Thompson Sampling, Bayesian Optimal Policy

2. Markov Decision Process:

- Primal-dual formulation
- Temporal difference (TD)
- Convergence analysis of TD
- GTD, Q-learning, Policy gradient, Actor-Critic

3. Continuous-time RL

- Stochastic Optimal Control
- Hamilton-Jacobi-Bellman Equation

Student Evaluation

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

Milestone 1: end-term presentation (40%)
Milestone 2: end-term report (50%).

2. Student Participation: 10%

1. Course Project: Each student will work on a research problem that is relevant to the course.

- The first component of the project is a **20-minute end-term presentation** on a research project.
  - Please send your research problem to the instructor by the eighth week of the quarter for approval.
  - It has to be a slides presentation, which is expected to happen during the end of the quarter. During the presentation, each student will receive questions from the instructor and the rest of the class.
- The second 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 the end of the tenth week by uploading to Gradescope

3. 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`

[4] "***Stochastic Approximation and Recursive Algorithms and Applications***"
by Harold J. Kushner , G. George Yin
Source: `https://link.springer.com/book/10.1007/b97441`

Prerequisites

This course is ideal for graduate students who are interested in applying novel research concepts to their own research. Students are expected to be familiar with basic concepts in decition-making and have a solid background in linear algebra and probability.