CS 6700: Reinforcement Learning

Lectures: Wed (2:00 : 3:15 p.m.), Thu (3:30 : 4:45 p.m); Tutorial: Mon (5:00 : 5:50 p.m.)

Prof. Balaraman Ravindran, BSB 349; Phone: 4370
Email: *ravi@cse.iitm.ac.in, ravindran.b@gmail.com*
TAs: Vihaan Akshaay (*me17b171@smail.iitm.ac.in*)
Prashant Govindarajan (*bs17b011@smail.iitm.ac.in*)

Updated on February 6, 2022

1 Learning Outcomes

Reinforcement Learning is a technique used to make sequential decisions in stochastic environments. The generality of its definition allows us to approach key problems in domains ranging from robotics to advertising. The vast modeling capability of deep learning gave new life to the field, evident in breakthroughs such as Deepmind's AlphaGo Zero and OpenAI's DoTA 2 Bot. By the end of this course, the student will have an understanding of the core principles and major advances in the subject and will be introduced to the relatively modern area of deep reinforcement learning.

2 PreRequisites

Probability and Statistics, Machine Learning (CS4011 or equivalent)

3 Classroom Mode

Live online lectures (click to join) and tutorials (click to join) in scheduled slots. There will be two lectures and one tutorial per week.

4 Textbooks and References

- Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. 2^{nd} Edition, MIT press, 2018.

5 Course Requirements

You are required to attend all the lectures. If you miss any of them, it is your responsibility to update yourself with class progress and to collect any materials that may have been handed out. You are required to adhere to the consequently enlisted submission deadlines of reports/assignments. Class participation is strongly encouraged to demonstrate an appropriate level of understanding of class material.

6 Course Content

1. **Bandits:** Explore-exploit dilemma, Value functions, Multi-armed bandits, Contextual bandits.
2. **The Full Reinforcement Learning Problem:** Evaluative feedback, Non-associative learning, RL vs MPC, Rewards and returns, Markov Decision Processes, Value functions, Optimality and approximation
3. **Bellman Equations & Dynamic Programming:** Bellman Equation & Optimality, Value iteration, Policy iteration, Asynchronous DP, Generalized Policy Iteration
4. **Evaluation & Control**: TD learning, SARSA, Q-learning, Monte Carlo RL
5. **Function Approximation & Value-based Methods**: Maximization Bias, Double Q learning, Value function approximation, Gradient descent methods, Linear function approximation, LSTD, LSTDQ, LSPI, Fitted Q, Deep Q learning, Double DQN, Prioritized Experience Replay, Duelling Architectures, Expected SARSA
6. **Policy Gradient & Actor Critic Methods:** REINFORCE, PG Theorem, Actor Critic methods, Baselines, A2C, A3C, Deterministic PG and DDPG, SAC, Constrained Policy Optimization
7. **Hierarchical RL:** Intro to Hierarchies, SMDPs, Options framework and Option discovery, DDO, MAXQ framework, HAMs
8. **POMDPs:** Definitions, Belief States, Solution Methods: Q-MDPs, LSTMs, Direct Solutions, PSR

9. **Model-based RL:** Connections to Planning, Types of MBRL, RL with a Learnt Model, Dyna-style models, Latent variable models, Implicit MBRL

   The lecture schedule is available here. Please check this sheet regularly for updates

7 Marking Scheme

Written Assignments	24%
Programming Assignments	36%
Written Exam 1	14%
Written Exam 2	14%
Tutorials	12%
Total	100%

8 Written Assignments

The written assignments are compulsory for all the students. There will be 3 written assignments during the course. These assignments encourage students to read and appreciate the course content beyond the standard textbook chapters by referring to several technical papers and video tutorials.

9 Programming Assignments

The programming assignments focus on the implementation of standard RL algorithms. 3 programming assignments will be carried out in OpenAI Gym or related RL environments. The students are expected to submit well-documented code and a detailed report which will contain answers to follow-up questions in the assignment and an analysis of the results.

10 Important Dates

Exams: (14 marks x2)
Exam 1 - March 14; Exam 2 - May 4

Programming Assignments: (12 marks x3)
Assignment 1 - [Q Learning & SARSA]
Release: (Feb 21) Due: (March 7)
Assignment 2 - [DQN & Actor Critic]
Release: (March 11) Due: (March 23)
Assignment 3 - [HRL & Options]
Release: (April 1) Due: (April 18)

Written Assignments: (12 marks x2)
Assignment 1 -
Release: (25th Feb) Due: (11th March)
Assignment 2 -
Release: (11th April) Due: (27th April)

11 Academic Honesty

Academic honesty is expected from each student participating in the course. NO sharing (willing, unwilling, knowing, unknowing) of assignment code between students, submission of downloaded code (from the Internet, Campus LAN, or anywhere else) is allowed.

The project work done as a part of this course cannot be used as-is, to meet any other degree requirements. The project must NOT be copied/downloaded material from the Internet or elsewhere.

Academic violations will be handled by IITM Senate Discipline and Welfare (DISCO) Committee. Typically, the first violation instance will result in ZERO marks for the corresponding component of the Course Grade and a drop of one- penalty in overall course grade. The second instance of code copying will result in a 'U' Course Grade and/or other penalties. The DISCO Committee can also impose additional penalties.

Please protect your Moodle account password. Do not share it with ANYONE. Do not share your academic disk drive space on the Campus LAN.