CS6700 - Reinforcement Learning

Instructor:
Prof. Ravindran B.
Teaching Assistants:
Arjun Manoharan, Nikita Moghe, Rahul Ramesh, Pavan Ravishankar, Rohan Saphal, Manan Tomar
Course Venue & Timings: CS34, J slot, Monday 4:50 - 6:00pm, Wednesday 2:00 - 3:15pm, Thursday 3:25 - 4:40pm

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. Recently, the vast modeling capability of Deep Learning gave new life to the field and this is evident in breakthroughs such as Deepminds AlphaGo Zero and OpenAIs 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 as well as have an introduction to the relatively modern area of Deep Reinforcement Learning.

Course Content

- *Bandits*: value functions, multi-armed bandits, gradient bandit algorithms, contextual bandits.
- *The Reinforcement Learning problem*: evaluative feedback, non-associative learning, Rewards and returns, Markov Decision Processes, Value functions, optimality and approximation.
- *Dynamic programming*: value iteration, policy iteration, asynchronous DP, generalized policy iteration.
- *Monte-Carlo methods*: policy evaluation, rollouts, on policy and off policy learning, importance sampling, MCTS.
- *Temporal Difference learning*: TD prediction, Optimality of TD(0), SARSA, Q-learning
- *Eligibility traces*: n-step TD prediction, TD (λ), forward and backward views, Q (lambda), SARSA (λ), replacing traces and accumulating traces.
- *Function Approximation*: Value prediction, gradient descent methods, linear function approximation, ANN based function approximation (DQN) , generalised value function.
- *Policy Gradient methods*: non-associative learning REINFORCE algorithm, exact gradient methods, estimating gradients, deterministic policy gradient algorithms(DPG), DDPG, approximate policy gradient algorithms, A2C, A3C.
- *Planning and Learning*: Model based learning and planning, Dyna
- *Hierarchical RL*: MAXQ framework, Options framework, HAM framework, Option discovery algorithms, POMDP
- *Misc. Topics*: Exploration, Imitation Learning, Average Reward RL.

Schedule

Visit `https://goo.gl/HesgL6` for the Course schedule. Please check this sheet regularly for updates. A copy of it will also be available on Moodle.

Marking Scheme

Written Assignments	30
Project/Programming A.	40
Mid Sem	15
End Sem	15
Total	100

Written Assignments

The written assignments are compulsory for all the students. There will be 3-5 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.

Project/Programming Assignments

The students can opt for either a Project or Programming Assignments.

Programming Assignments

The programming assignments focus on the implementation of standard RL algorithms. The 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 along which will contain answers to follow-up questions in the assignment and analysis of the results.

Project

The students are selected through an evaluation of programming assignment and paper critique. Post the selection, the students can work in teams of two on a project suggested by a TA or formulate their own problem statement. The project serves as an introduction to research in Reinforcement Learning and gives hands-on experience on implementation/extension of recent RL literature. There will be a total of 5-7 different projects.

Course Requirements

You are *required* to attend all the lectures. If you miss any of them it is your responsibility to find out what went on during the classes and to collect any materials that may be 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 the material being discussed in the class.

Classroom Mode

Traditional classroom lectures in the scheduled slots. Also have a few additional video tutorials that augment concepts taught in the classroom.

Pre-requisites

Probability and Statistics, Machine Learning (CS4011 or equivalent)

Academic Honesty

Academic honesty is expected from each student participating in the course. NO sharing (willing, unwilling, knowing, unknowing) of reports/assignments between students, submission of downloaded material (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.

Text books and references

[1] Richard S. Sutton and Andrew G. Barto. Introduction to Reinforcement Learning. MIT Press, Cambridge, MA, 2nd edition, 2017.