CE 273 – Markov Decision Processes (3 credits)

Instructor: Tarun Rambha
Email: tarunrambha@iisc.ac.in
Timings and Location: TTh 9:30 – 11:00, Google Meet
Course Website: civil.iisc.ac.in/~tarunr/ce273.html

Course Description:
This course provides an introduction to Markov decision processes (MDPs), which is class of problems that involve repeated decision making in stochastic environments. MDPs allow us to alter the evolution of stochastic systems by exercising control to achieve a certain objective. They have a wide range of applications within and outside transportation and operations research and play a major role in some of the state-of-the-art artificial intelligence systems. Applications in transportation such as shortest paths, pricing, signal control, demand estimation, and parking will also be discussed.

The course is designed to be self contained but a background in linear programming and a knowledge of undergraduate-level probability and statistics is assumed. We will also use some elementary concepts from functional analysis and linear algebra. Basic knowledge of stochastic processes is also very handy and we will devote the first few lectures to understand these ideas using some examples.

Reading Material:
The following books could be used as references. Additional papers and reading material will be shared during the lectures.

1. Kulkarni, V. G. (2016). Modeling and analysis of stochastic systems. CRC Press.
2. Bertsekas, D. P. (2012). Dynamic programming and optimal control (Vol. 1 & 2). Belmont, MA: Athena Scientific.
3. Puterman, M. L. (2014). Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons.
4. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

Assignments:
The course will have 5 individual assignments which may include small programming exercises. In addition there will be 5 form-based quizzes. All programming tasks should be carried out using Python. Some of the programming tasks may involve using Python libraries such as `mdptoolbox`. The assignments are to be typed and submitted electronically. The first two submissions must be in LaTeX, but the remaining assignments can be prepared using your preferred typesetting tool. You are encouraged to discuss the problems with your classmates but you must write your own solutions. **Plagiarism and copying is strictly prohibited and will be penalized**.[1]

Projects:
The course will, in addition, have one individual project and one paper presentation project. For the paper presentation project, I will upload a list of candidate papers but you are free to choose a paper outside this list after discussing with me. Specific details regarding the individual project will be provided later. The individual project will involve applying some of the concepts learnt on to play a popular computer game. It will be due at the end of the course.

[1] http://www.iisc.ac.in/about/student-corner/academic-integrity/

Examinations:
The course will have one comprehensive end-semester exam. Given the current situation, the end-semester is likely to be a take-home exam.

Grading:
Students will be graded on a 100-point scale. The weights for different components of the course are as follows: written assignments (25%), form-based quizzes (25%), paper presentation (10%), individual project (10%), and end-semester exam (30%). Scores for the paper presentation will be provided by your peers.

Lesson Schedule:
The following is a rough outline of the topics that will be covered and the number of lectures dedicated to it. Minor adjustments will be made as the semester progresses.

1. **Stochastic Processes and Finite Horizon MDPs** (6 lectures):
   Discrete Time Markov Chains (DTMCs) and applications, transition probabilities and occupancy times, classification of DTMCs, limiting behavior, finite horizon MDPs, backward induction, structural results, and applications.

2. **Infinite Horizon Discounted MDPs** (4 lectures):
   Banach spaces and contraction mappings, value iteration, policy iteration, linear programming methods, and applications.

3. **Infinite Horizon Total and Average Cost MDPs** (6 lectures):
   Existence of optimal policies, solution methods (value iteration and policy iteration), unichain and multichain models, and applications.

4. **Approximate Dynamic Programming and RL** (6 lectures):
   Roll-out methods, lookahead and Monte-Carlo Tree Search, model-free methods (TD(λ), Q-learning, SARSA), function approximation, and policy gradient.

5. **Additional topics** (remaining time):
   Dynamic discrete choice models, risk-sensitive MDPs, partially observable MDPs (POMDP), continuous-time models.