MIE1615: Markov Decision Processes

Department of Mechanical and Industrial Engineering, University of Toronto

Reference: "Markov Decision Processes - Discrete Stochastic Dynamic Programming", Martin L. Puterman, Wiley, 1994.
"Markov Decision Processes in Artificial Intelligence", Olivier Sigaud and Olivier Buffet, Wiley, 2008.
"Approximate Dynamic Programming", Warren Powell, Wiley, 2007.
"Neuro-Dynamic Programming", Dimitri Bertsekas and John Tsitsiklis, Athena Scientific, 1996.

Instructor: Chi-Guhn Lee, BA8110, 946-7867, cglee@mie.utoronto.ca.

Office Hours: By appointment.

Course Description: This is a course to introduce the students to theories of Markov decision processes. Emphasis will be on the rigorous mathematical treatment of the theory of Markov decision processes. Topics will include MDP finite horizon, MDP with infinite horizon, and some of the recent development of solution method.

Course Assessment: **Assignments (20%):** There will be two assignments.
Exams (40%): There will be one two-hour exam in class on Mar/11.
Project (40%): Each group will be asked to give two 25-minute lectures on two papers: one from each of the two lists; Also, students are asked to read two additional papers from each list.

Important Dates: **Feb/24:** Team formation and paper selection (2 papers/team)
Mar/2: Exam in class
Mar/30 and Apr/6: No lecture, student presentation only.

Policies: Late submission of assignments will NOT be accepted.
Requests for regrading of assignments or exams will be considered only within one week from the time the graded work is returned in class.

Topics Covered: **Introduction, examples and structural properties** (3 weeks)

Finite Horizon Markov Decision Processes (2 weeks)

- Policies and optimality criteria
- Policy evaluation
- The principle of optimality
- Optimality of deterministic Markov policies
- Backward induction

Infinite Horizon Markov Decision Processes (2 weeks)

- Preliminary mathematics
- Value of a Policy
- Expected Total Discounted Reward
- Policy evaluation
- Optimality Equations
- Value Iteration
- Policy Iteration

Function Approximation (1.5 weeks)

- Linear Approximation
- Approximate Policy Iteration

Policy Gradient Method (1.5 weeks)

- Gradient Descent
- Policy Gradient Theorem
- Gradient Estimmation
- Actor Critic Methods