CS774 Reinforcement Learning

Lecture #1: Introduction

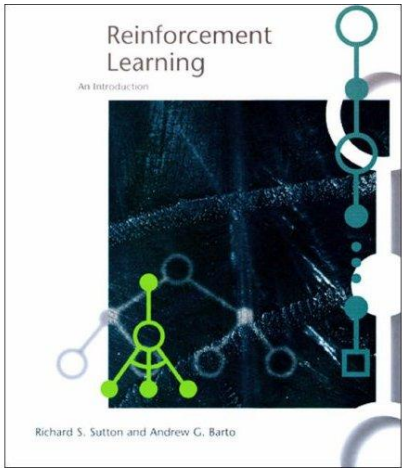
Kee-Eung Kim
KAIST EECS Department
Computer Science Division

KAIST
Korea Advanced Institute of Science and Technology
한국과학기술원

Course Overview

☐ Primary textbook (first half of the semester)

- Sutton & Barto,
  "Reinforcement Learning: an Introduction",
  MIT Press



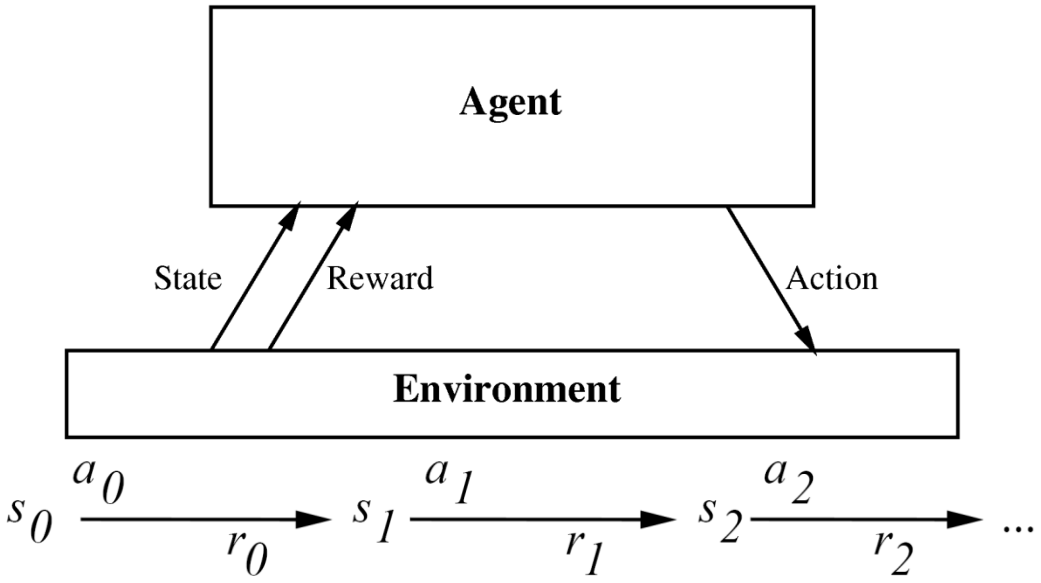
☐ Supplementary papers (second half of the semester)

- Policy methods, POMDPs and point-based methods, PSR,
  Multi-agent RL and applications

☐ Grading

- Class participation: 30%
- Midterm Exam: 50%
- Final Project Presentation: 20%

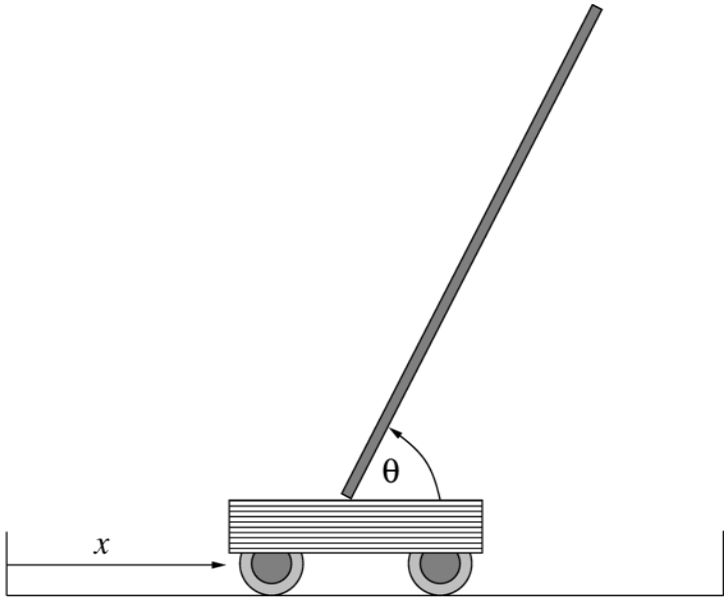
Reinforcement Learning



- An approach to Artificial Intelligence
- Learning from interaction
  - Learning about, from, and while interacting with an external environment; explore vs. exploration
- Goal-oriented learning
  - Learning what to do—how to map situations to actions—so as to maximize a numerical reward signal

KAIST
Korea Advanced Institute of Science and Technology
한국과학기술원

Example: Cart-Pole



☐ State: $s_t = \langle x_t, \dot{x}_t, \theta_t, \dot{\theta}_t \rangle$

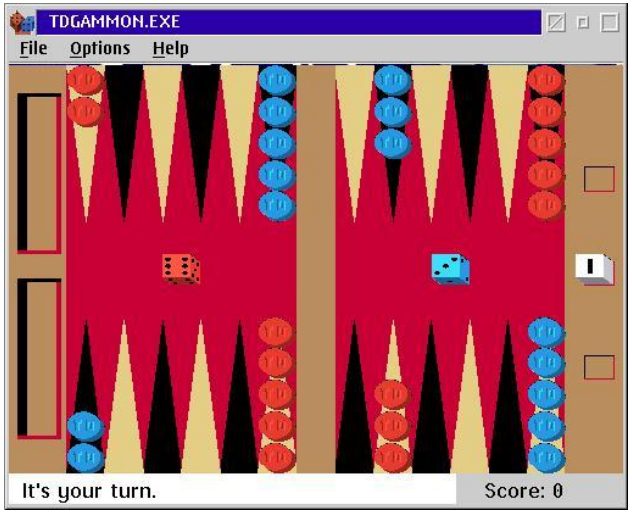
☐ Reward:

- -1 when the pole falls down or cart reaches the end of the rail
- 0 otherwise

☐ Action: jerk-left, jerk-right

Example: Backgammon

☐ TD-Gammon: learn to play Backgammon



☐ Reward

- +100 if win
- -100 if lose
- 0 for all other states

☐ Trained by playing 1.5 million games against itself

☐ Now approximately equal to best human player

☐ Shipped with IBM OS/2

Comparison to Supervised Learning

☐ Supervised learning: *"learning with teacher"*

- Target function to be learned: $h : \vec{X} \rightarrow Y$
- Input-output pairs from the function to be learned is required: $(\vec{x}_1, y_1), (\vec{x}_2, y_2), \ldots, (\vec{x}_N, y_N)$
- Compute function f that approximates the target function h

☐ Consider

- Robot learning to dock on battery charger
- Learning to choose actions to optimize factory output
- Learning to play backgammon

☐ Reinforcement learning: *"learning with critic"*

- Doesn't tell us what to do - how well we have been doing in the past
- Feedback is scarce, and it comes late

Assignments

☐ Read chapter 1 & 2 before the next class