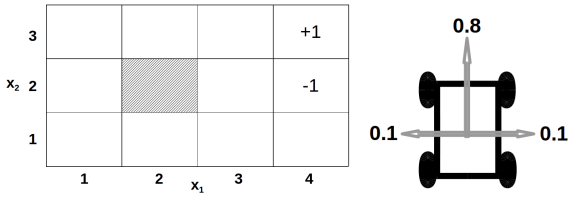
[Lab] Reinforcement Learning

Jae Yun JUN KIM*

Lab due: Before today lab session
Evaluation: Code and explanation about the code (in groups of 2 or 3 people (preferably 3))
Remark:

- Only groups of two or three people accepted (preferably three).
- Before you leave today lab session, you need to show the lab task results to the professor.
- No late lab will be accepted. There is no make-up session.
- No plagiarism. If plagiarism happens, both the "lender" and the "borrower" will have a zero.
- Code yourself from scratch. No lab/homework will be considered if any ML library is used.
- Do thoroughly all the demanded tasks.
- Study the theory for the questions.

For this lab session, you are asked to plan the motion of a 2D mobile robot using the Markov Decision Process formalism. Consider the following 2D map for the autonomous navigation of a mobile robot



This map consists of 12 cells. The dashed cell at (x_1, x_2)=$(2, 2)$ represents an obstacle to be avoided. The cell with reward "+1" at (x_1, x_2)=$(4, 3)$ is a desired absorbing cell (the goal), while the cell with reward "-1" at (x_1, x_2)=$(4, 2)$ is an undesired absorbing cell (e.g., a pit). On the other hand, the mobile robot can take four actions: A=$\{N, S, E, W\}$, where N, S, E, W represent north, south, east and west, respectively. If A=N, then the mobile robot behaves following transition probability distribution indicated in the above figure. This is also true for the rest of actions. Further, the reward function is defined as follows

$$R = \begin{cases} +1 & (x_1, x_2)=(4, 3) \\ -1 & (x_1, x_2)=(4, 2) \\ -0.02 & \text{otherwise} \end{cases}$$

Finally, assign the discount factor (γ) to be 0.99.

1 Pre-lab tasks (Due before the current lab session)

1) For all states, find the optimal value function $V^*(s)$ and the optimal policy function $\pi^*(s)$ using the *value iteration* algorithm.

2 Lab tasks (Due before the end of today lab session)

1) For all states, find the optimal action-value function $Q^*(s, a)$ and the optimal policy function $\pi^*(s)$ using the *Q-learning* algorithm.

*ECE Paris Graduate School of Engineering, 37 quai de Grenelle 75015 Paris, France; jae-yun.jun-kim@ece.fr