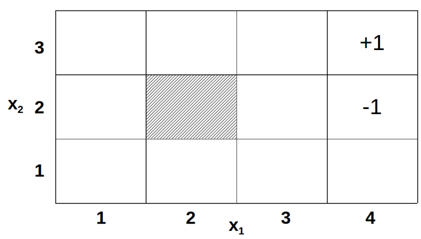
[Prelab] Reinforcement Learning

Jae Yun JUN KIM*

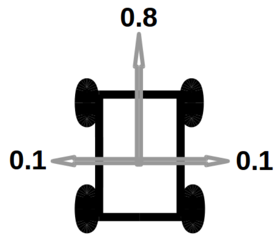
<u>**Lab due**</u>: Before the next lab session
<u>**Evaluation**</u>: Code and explanation about the code in groups of only two or three people
<u>**Remark**</u>:

- Only groups of two or three people accepted (preferably three).
- No plagiarism. If plagiarism happens, both the "lender" and the "borrower" will have a zero.
- Code yourself from scratch **following the theory given in class**. No prelab/lab will be considered if any ML library is used.
- Do thoroughly all the demanded tasks.
- Study the theory for the questions.
- There is no prelab/lab make-up session.

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



(a) 2D map



(b) 2D mobile robot

Figure 1: 2D map and 2D 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 Figure 1(b). This is also true for the rest of actions. Further, the reward function is defined as follows

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

Finally, assign the discount factor (γ) to be 0.9.

1 Tasks

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

*ECE Paris Graduate School of Engineering, 10 Rue Sextius Michel 75015 Paris, France; jae-yun.jun-kim@ece.fr