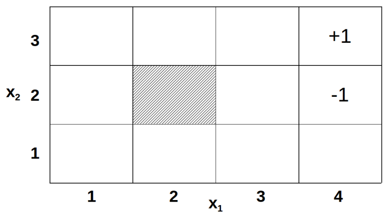
INF436 Machine Learning: Lab 7
Reinforcement learning

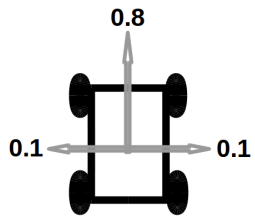
Jae Yun JUN KIM*

February 28, 2017

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 0.99.

Exercises

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

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

3) Compare the results obtained in 1) to those of 2).

*ECE Paris Graduate School of Engineering, 37 quai de Grenelle CS71520 75 725 Paris 15, France; jae-yun.jun-kim@ece.fr