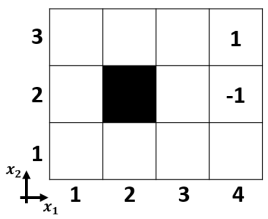
# [Quiz] Reinforcement Learning

Number: 1
Group:
Name:

1. Consider a problem to find the best policy function for the following map and conditions:

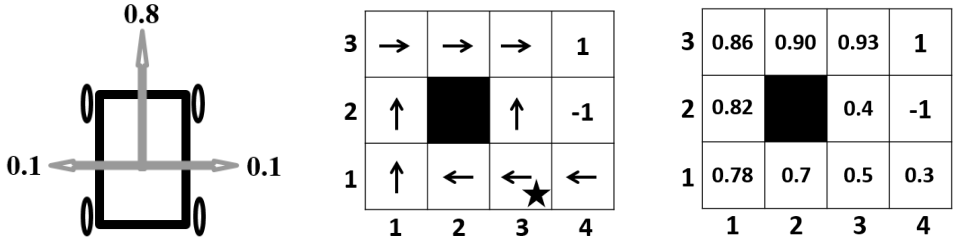



   - 11 states (i.e., 11 cells).
   - 4 actions: $A = \{N, S, E, W\}$, where N=North, S=South, E=East, W=West.
   - Rewards:

$$R = \begin{cases} 1, & \text{state} = (4,3) \\ -1, & \text{state} = (4,2) \\ -0.05, & \text{otherwise} \end{cases}$$

   - For simplicity, suppose that the discount factor $\gamma$ is 1 (i.e., no discount).
   - Finish the algorithm when the robot hits either the cell of $R$=1 or $R$=−1.

   Because the system is noisy, the action is modeled as shown in Figure 1(a). In addition, suppose that we have the following policy (Figure 1(b)) and value function (Figure 1(c)):




(a) Action model (b) Policy (c) Value function

Figure 1: Action model, policy and value function

   Using all the information given above, compute the new value function for the state $s$=(3, 1), where the state in question is indicated with a star in Figure 1(b).

2. Write down the Value Iteration (VI) algorithm. Describe it.