Programming Session: Exercise

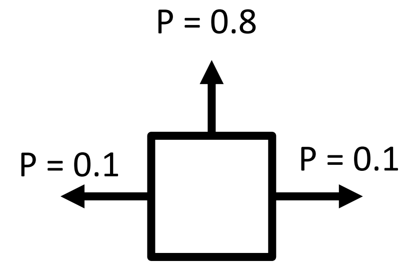
Exercise 1: Policy Iteration

0	0	0	1
0		0	-100
0	0	0	0

Rewards r(s)

$\gamma = 0.9$

The agent moves in the selected direction with probability 0.8 and in the perpendicular directions with prob. 0.1. If the agent bumps the wall, it stays in the same cell.



Implement policy iteration to find the optimal policy for the problem presented above. Initial values V(s) = 0. Policy initialized randomly. Threshold 0.01. r(s,a) = r(s).

Present the maximum values at each state in a matrix form coinciding with the grid (3x4).

Specify the action for each state from the final policy in a matrix form (3x4). Use the notation 1 for action up, 2 for down, 3 for left, and 4 for right.

Policy Evaluation

Provide policy π

Init $V_0(s) = 0$

Repeat

For each $s \in S$

$$V_{k+1}(s) = r + \gamma \sum_{s'} p(s' \mid s, \pi(s)) V_k(s')$$

Until $\left| v_{k+1} - v_k \right|_\infty < threshold$

Return $v_{k+1} \approx v^\pi$

Policy Improvement

$$\pi'(s) = \arg\max_a \left[r(s,a) + \gamma \sum_{s'} p(s' \mid s, a) V^\pi(s') \right]$$

Exercise 2: Action-value Iteration

Implement action-value iteration to find the optimal policy for the problem presented in Exercise 1. Initial values Q(s,a) = 0.

Present the Q-values for each action as well as the the maximum Q-value for each state in seprate matrices (total 4 matrices of 3x4).

Specify the action for each state from the final policy in a matrix form (3x4). Use the notation 1 for action up, 2 for down, 3 for left, and 4 for right.

Action-value Iteration

Init $Q_0(s,a) = 0$

Repeat

For each $s, a \in S, A$

$$Q_{k+1}(s,a) = r(s,a) + \gamma \sum_{s'} p(s' \mid s,a) \max_{a'} Q_k(s',a')$$

Until $\left| \mathbf{q}_{k+1} - \mathbf{q}_k \right|_\infty < threshold$

Return $\pi(s) = \arg\max_a Q(s,a)$

$\pi(s) \approx \pi^*(s)$

Programming Session

Implement the code (preferably) in Matlab. Save all the implemented files in a folder `L1_surnames`. Implement a script called `test.m` that executes the implemented functions and presents the requested matrices.

Send the folder compressed (.zip) by email to alejandro.agostini@tum.de with the subject RLRWS20_L1_surnames. **Deadline: next lecture**.