TÉCNICO LISBOA

MSc in Computer Science and Engineering

Learning and Decision Making 2016-2017

Homework 5. Reinforcement learning

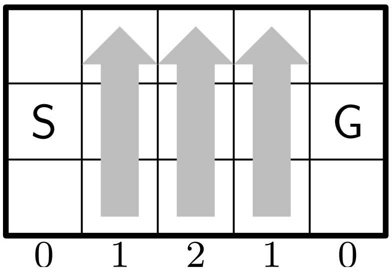

Figure 1: Windy gridworld.

Consider a boat navigating the gridworld depicted in Fig. 1. The cell marked with G corresponds to the goal state. There is a crosswind upward through the middle of the grid, in the cells marked by the gray arrows.

The boat has available the standard four actions—up, down, left and right. In the region affected by the wind, however, the resulting next state is shifted upward as a consequence of the crosswind, the strength of which is different between the two columns. The strength of the wind is given below each column, and corresponds to the number of cells that the movement is shifted upward. For example, if the boat is one cell to the right of the goal, then the action left takes you to the cell just above the goal.

The agent pays a cost of 1 in every step before reaching the goal.

Exercise 1.

(a) The problem described above can be modeled as an MDP. Identify the state space, $\mathcal{X}$, and the action space, $\mathcal{A}$, and write down the cost function for the MDP (you need not specify the transition probabilities). Consider throughout $\gamma = 0.95$.

Suppose that the boat starts in the state marked with S and executes the action "Right" twice.

(b) Indicate the transition information resulting from the two actions of the agent (state, action, cost, next state).

(c) Suppose that the agent is following the Q-learning algorithm, with the Q-function initialized as an all-zeros function. Indicate the Q-values after the two Q-learning updates with step-size $\alpha = 0.1$, resulting from the transitions in (b).