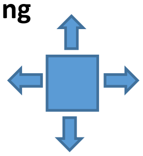
Reinforcement Learning

Homework 3: Deadline November 1, 2021

The environment is modeled by a grid of 5x5 cell size as shown in figure below, but you are welcome to try a larger grid size (e.g., 10 x 10). The cells of the grid correspond to the states of the environment. Assume that the robot has four actions (up, down, right, left) to select at each time/iteration. You would need to define the reward for the robot to learn to find an optimal/sub-optimal way to go to the goal. Optimal here means less/minimum number of actions taken by the robot

Suggested reward (you are encouraged to define your own reward):

- Action that makes the robot tend to go out of the grid will get a reward of -1 (when the robot is in the border cells)
- Action that makes the robot reach the goal will get a reward of 100
- All other actions will get a reward of 0

Starting 				
				Goal

Using Q learning (Off-Policy Control) presented in lecture 10, slides# 10-11, to train the robot for this task. This Q learning technique is in Chapter 6 of the Sutton's book.

Requirement:

1. Plot the action selection of the initial learning episode and the last learning episode. Something is similar to this table:

Starting →	→			
	↓			
	→	→		
		↓	→	↓
				Goal ↓

2. Show the Q table of the last learning episode
3. Plot the reward of all learning episodes
4. Submit your homework with source code to Webcampus/Canvas