Department of Mechanical and Mechatronics Engineering - University of Waterloo

ME780
Computational Intelligence
Quiz 3 – Fall 2021

INSTRUCTOR: William W. Melek

ASSIGNED: November 22th, 2021 **Due Date: Dec 6th 2021**

(15%) Consider the mini-world problem in the figure below. It has three terminal states; (2,6) with maximum reward and (4,5), (4,6) with maximum penalty. Your robot is required to navigate this world efficiently. You will design a navigation policy based on reinforcement learning to allow the robot to reach the goal state while maximizing its reward and avoiding the terminal states. At the initial state, the robot can start any learning trial only from positions (1,1) or (3,2). You have to attempt at least 20 trials (no duplicates) to develop your navigation policy using your reinforcement learning approach. The robot can move in any direction in the mini-world to transition from one state to the next. You will be required to solve this problem using the following two methods covered in class:

(a) Adaptive dynamic programming (ADP) using value iteration to optimize the value of each state in the world below. The initial value for each state can be set to 0. Use the final value of each state to develop the optimal policy for the robot to navigate this world.
(b) Q-learning (active learning) using value iterations on actions the robot can take at each state. If the optimal policy developed in this step is different from the one you obtained in step (a), you need to explain the reason(s) for these discrepancies.

You can choose your programming language of choice to solve this problem. You can use directional arrows to represent your optimal policy and the direction of travel from any given state in the mini-world to the terminal state with the maximum reward.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(1,2)	(2,2)	(3,2)	(2,4)	(2,5)	+1 (2,6)
(3,2)	(2,3)	(3,3)	(3,4)	(3,5)	(3,6)
(4,2)	(2,4)	(3,4)	(4,4)	-1 (4,5)	-1 (4,6)