SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

50.007 Machine Learning, Fall 2021

Homework 6

For practice only. No submission is required.

Consider the following Markov decision process (MDP). It has states $\{0, 1, 2, 3, 4\}$ with 4 as the starting state. In every state, you can take one of two possible actions: walk (W) or jump (J). The Walk action decreases the state by one. The Jump action has probability 0.5 of decreasing the state by two, and probability 0.5 of leaving the state unchanged. Actions will not decrease the state below zero: you will remain in state 0 no matter which action you take (i.e., state 0 is a terminal state). Jumping in state 1 leads to state 0 with probability 0.5 and state 1 with probability 0.5. This definition leads to the following transition functions:

- For states $k \geq 1,\ T(k, W, k-1) = 1$
- For states $k \geq 2,\ T(k, J, k-2) = T(k, J, k) = 0.5$
- For state $k = 1,\ T(k, J, k-1) = T(k, J, k) = 0.5$
- For state $k = 0,\ T(k, J, k) = T(k, W, k) = 1$

The reward gained when taking an action is the distance travelled squared, i.e., $R(s, a, s') = (s - s')^2$. The discount factor is $\gamma = 0.5$.

1. Suppose we initialize $Q_0^*(s, a) = 0$ for all $s \in \{0, 1, 2, 3, 4\}$ and $a \in \{J, W\}$. Evaluate the Q-values $Q_1^*(s, a)$ after exactly one iteration of the Q-Value Iteration Algorithm. Write your answers in the table below.

	$s = 0$	$s = 1$	$s = 2$	$s = 3$	$s = 4$
J					
W					

2. What is the policy that we would derive from $Q_1^*(s, a)$? Answer by filling in the action that should be taken at each state in the table below.

$s = 1$	$s = 2$	$s = 3$	$s = 4$

3. What are the values $V_1^*(s)$ corresponding to $Q_1^*(s, a)$?

$s = 0$	$s = 1$	$s = 2$	$s = 3$	$s = 4$

4. Will the policy change after the second iteration? If your answer is"yes", briefly describe how.