SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

01.112 Machine Learning, Fall 2017
Homework 6 (Practice)

In this homework, we would like to look at some simple MDP problem.

Consider the Markov decision process (MDP) associated with a simplified version of the robot that we showed during class that plays the guessing game with us. It has 4 states {*Uncertain*, *Certain*, *Lose*, *Win*}, with *Win* being the final state. The following table summarizes the actions that the robot can take at each state, and the transition probabilities from one state to another after taking a certain action. For example, when the robot is at the *Certain* state (it is certain about the answer), there are two possible actions to take: ask (A) and guess (G). If it takes the action G, then there is a probability 0.7 that the robot will arrive at the *Win* state (and wins the game).

s	a	s'	$T(s, a, s')$
Uncertain	A	*Uncertain*	0.9
Uncertain	A	*Certain*	0.1
Uncertain	G	✗ *Lose*	0.9
Uncertain	G	*Win*	0.1
Certain	A	*Certain*	1.0
Certain	G	*Lose*	0.3
Certain	G	*Win*	0.7
Lose	A	*Uncertain*	0.8
Lose	A	*Certain*	0.2
Lose	G	✗ *Lose*	0.8
Lose	G	*Win*	0.2
Win	A	*Win*	1.0
Win	G	*Win*	1.0

The reward function $R(s, a, s') = R(s')$ is defined as:

s'	*Uncertain*	*Certain*	*Lose*	*Win*
$R(s')$	0.0	1.0	-2.0	2.0

The discount factor is $\gamma = 0.5$.

1. Suppose we initialize $Q_0^*(s, a) = 0$ for all $s \in S$ and $a \in \{A, G\}$. Evaluate the Q-values $Q_1^*(s, a)$ after exactly one iteration of the Q-Value Iteration Algorithm (assume we perform synchronized updates. In other words, we always use Q_0 values when we calculate Q_1 values). Write your answers in the table below.

	$s =$*Uncertain*	$s =$*Certain*	$s =$*Lose*	$s =$*Win*
A	0.9×0.0 +0.1×1.0 = 0.1 ✓	1.0×1.0 = 1.0 ✓	0.8×0.0 +0.2×1.0 = 0.2 ✓	1.0×2.0 = 2.0
G	0.9×-2.0 +0.1×2.0 = -1.8 + 0.2 = -1.6	0.3×-2.0 +0.7×2.0 = -0.6+1.4 = 0.8	0.8×-2.0 +0.2×2.0 = -1.6 + 0.4 = -1.2	1.0×2.0 = 2.0 ✓

1

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 (in case of draw, the action G is preferred).

	$s =$Uncertain	$s =$Certain	$s =$Lose	$s =$Win
	A	A	A	G

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

	$s =$Uncertain	$s =$Certain	$s =$Lose	$s =$Win
	0.9×0.0 +0.1×1.0 = 0.1	1.0×1.0 = 1.0	0.8×0.0 +0.2×1.0 = 0.2	1.0×2.0 = 2.0

4. Evaluate the Q-values $Q_2^*(s, a)$ after exactly two iterations of the Q-Value Iteration Algorithm. Write your answers in the table below.

	$s =$Uncertain	$s =$Certain	$s =$Lose	$s =$Win
A	0.9 [0.0 + 0.5 (0.1)] + 0.1 [1.0 +0.5(1.0)] = 0.045 + 0.15 = 1.95	1.0[1.0 + 0.5(1.0)] = 1.5	0.8[0.0+0.5(0.1)] +0.2[1.0 + 0.5(1.0)] = 0.04 + 0.3 = 0.34 ✓	1.0[2.0+0.5 (2.0)] = 3.0
G	0.9[-2.0+ 0.5 (0.2)] + 0.1[2.0 + 0.5 (2.0)] = 0.9[-1.9] + 0.1[3.0] = -1.41	0.3[-2.0+0.5(0.2)] +0.7[2.0+0.5(2.0)] = 0.3[-1.9] + 0.7[3.0] = 1.53 ✓	0.8[-2.0+0.5(0.2)] +0.2[2.0 + 0.5(2.0)] =0.8[-1.9]+0.2[3.0] = -0.92 ✓	1.0[2.0+0.5(2.0)] = 3.0 ✓

5. What is the policy that we would derive from $Q_2^*(s, a)$? Answer by filling in the action that should be taken at each state in the table below (in case of draw, the action G is preferred).

	$s =$Uncertain	$s =$Certain	$s =$Lose	$s =$Win
	A	G	A	G

6. What are the values $V_2^*(s)$ corresponding to $Q_2^*(s, a)$?

	$s =$Uncertain	$s =$Certain	$s =$Lose	$s =$Win
	1.95	1.53	0.34	3.0