CSC 449 Advanced Topics in Artificial Intelligence

Deep Reinforcement Learning
Exam 1
Fall, 2022

Name: _____ ID#: _____ Score: _____

1. (30 pts) Write the Bellman equation and describe each term in it.

2. (30 pts) Assume that you are performing value iteration on a standard gridworld problem, The immediate reward for taking any action in any state is -1, except for the RIGHT or DOWN actions in state 15. The immediate reward for those actions is zero. Taking an action that would move off the grid will cause the state to remain the same.
On iteration i, the policy, $\pi(s)$, and the current value function estimate, $V_\pi^i(s)$, are shown below.

$\pi(s)$

State	$p(L)$	$p(U)$	$p(R)$	$p(D)$
0	0.2	0.1	0.4	0.3
1	0.25	0.2	0.45	0.1
2	0.4	0.25	0.15	0.2
3	0.2	0.3	0.4	0.1
4	0.4	0.4	0.1	0.1
5	0.2	0.4	0.3	0.1
6	0.4	0.25	0.3	0.05
7	0.2	0.1	0.4	0.3
8	0.4	0.3	0.2	0.1
9	0.3	0.2	0.4	0.1
10	0.35	0.1	0.3	0.25
11	0.25	0.1	0.3	0.35
12	0.25	0.4	0.05	0.3
13	0.1	0.1	0.3	0.5
14	0.25	0.4	0.05	0.3
15	0.3	0.25	0.1	0.35

$V_\pi^i(s)$

0 −5.85	1 −5.84	2 −5.82	3 −5.78
4 −5.84	5 −5.81	6 −5.71	7 −5.58
8 −5.82	9 −5.71	10 −5.43	11 −4.95
12 −5.78	13 −5.58	14 −4.95	15 −3.30

a) Assuming a discount factor, γ, of 0.1, calculate the new value for state nine, $V_\pi^{i+1}(s_9)$.

b) Based on the value function above, $V_\pi^i(s)$, what appears to be the best deterministic policy for state one, $\mu(s_1)$.

3. (10 pts) How do you determine:
   a) When to stop the value iteration step while performing Generalized Policy Iteration?

   b) When to stop performing Generalized Policy Iteration?

4. (30 pts) Write the algorithm for either Q-Learning or Sarsa, and indicate which one you have provided.