Khoa Tran
02/18/2021
CSE 415

A5 Report

0a.

$$V_k^*(s) = \max_a Q_k^*(s,a)$$

$$Q_k^*(s,a) = \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma V^*(s')\right]$$

0b.

$$V_{k+1}^*(s) = \max_a Q_{k+1}^*(s,a)$$

$$Q_{k+1}^*(s,a) = \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma V_k^*(s')\right]$$

1.
a. 4 iterations of VI to make 1/3 of the states turn green
b. 8 iterations of VI to turn all the states to 100
c. This policy is not effective as most of the policy arrows points towards places with no change of state

2.
a.8 iterations of VI are needed for the start state to have a nonzero value
b. This is a good policy because most of the policy arrows points towards valid states that results in a change of state, which is closer to the goal state.
c. 56 iterations of VI are needed to converge
d. The policy has not changed because the initial action that is chosen is based on the highest utility value, which remains the same between 8 and 56 iterations.

3.
a. The start state has a value of 0.82. With a smaller discount, the algorithm focuses on immediate rewards. As a result, the policy shows that it prefers R=10 for the goal state and has a path towards it
b. The start state has a value of 36.9. With a higher discount, the algorithm focuses on future rewards. As a result, the policy shows that it prefers the R=100 goal state as there aren't any path towards the R=10 goal state.

4.
a.

Simulation	1	2	3	4	5	6	7	8	9	10
Off plan	Y	N	N	Y	Y	N	Y	Y	N	N
Goal state	N	Y	Y	N	N	Y	Y	N	Y	Y

If no goal, how many steps left	3	0	0	2	5	0	0	1	0	0

d. Top triangle wasn't visited

5.

a. It is not necessary to have convergence in an optimal policy because the optimal policy can be found before convergence is resulted. After a certain number of iterations, the Vk values changes minimally, which doesn't change the policy. As a result, convergence is not needed for an optimal policy.

b. It would be very important for tall the states to be visited multiple times in an algorithm that does not compute Value Iteration and has to explore all the possible states. In order to understand the relationship between the states, the algorithm has to visit each state multiple times in order to compute the optimal policy.