Artificial Intelligence

1)

a) -1 + 0.8 * (1 * 6) = 3.8
b) -1 + 0.8 * (1 * 8) = 5.4
c) The main difference between Q update rule and Q learning update is Q learning does not require knowledge of the transition function T

2)

a) (s, a) pairs transitioning to states 2, 3, or 4, leading to overestimated values for V(2),V(3), V(4). (s, a) pairs from states 2, 3, or 4 with action 'down', moving away from the target state 8.

Specific pairs include (1, right), (2, right), (2, down), (3, left), (3, right), (3, down), (4, left), (4, down), (6, up), and (7, up).

b) No additional categories are identified.

3)

When using Q-values, selecting a good action only requires the value itself, whereas using a value function also requires knowledge of the transition function.

4)

From the Q1
Q(3, right): 7*(1-0.2)+0.2*(-1+0.8*9)7*(1-0.2)+0.2*(-1+0.8*9)
Q(4, down): 9*(1-0.2)+0.2*109*(1-0.2)+0.2*10

5)

Initially increase epsilon to promote exploration, allowing the agent to explore more of the state space. As training progresses, decrease epsilon to shift focus towards exploitation, using the knowledge the agent has gained for better decision-making.

6)

a) pi(3, right) = e7 / e8 + e5 + e7 = e2 / e3 + e2 + 1 = 0.259

pi(3, left) = e5 / e8 + e5 + e7 = 1 / e3 + e2 + 1 = 0.035

b) Exploration guided by Q-values implies that choices are not entirely random but are influenced by the perceived values of actions, allowing for more informed exploration.

7)

a)

Probabilities for transitioning from state 3 with action 'Down' to states 6 and 7 are 0.099 and 0.901, respectively.

p(s=3,a=Down, s'=6)=0.099p(s=3,a=Down, s'=6)=0.099

p(s=3,a=Down, s'=7)=0.901p(s=3,a=Down, s'=7)=0.901

b)

The Q-value for action 'Down' at state 3 is calculated as

−1+0.099×Q(s=6,right)+0.901×Q(s=7,right), where Q(s=6,right)=7Q(s=6,right)=7 and Q(s=7,right)=8Q(s=7,right)=8.