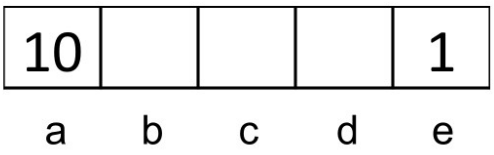
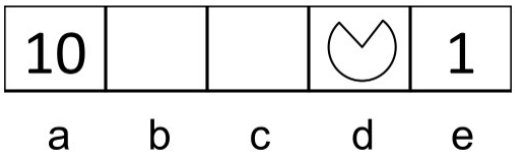
1. Consider a gridworld MDP as shown below. The available actions in each state are to move to the neighboring grid squares. Left and Right actions are successful 80% of the time. When not successful, the agent stays in place. From state a , there is also an exit action available, which results in going to the terminal state and collecting a reward of 10. Similarly, in state e, the reward for the exit action is 1. Exit actions are successful 100% of the time.



Fill in all of the following quantities.

a) $T(c, \text{Right}, d) =$
b) $T(c, \text{Right}, e) =$
c) $T(c, \text{Right}, c) =$
d) $T(c, \text{Right}, b) =$
e) $T(c, \text{Right}, a) =$
f) $T(c, \text{Right}, \text{terminal state}) =$
g) $T(a, \text{Exit}, \text{terminal state}) =$
h) $T(a, \text{Right}, b) =$
i) $T(a, \text{Right}, a) =$
j) $R(a, \text{Right}, a) =$
k) $R(a, \text{Exit}, \text{terminal state}) =$
l) $R(c, \text{Right}, d) =$

2. Consider the same gridworld MDP as in the previous quiz. The available actions in each state are to move to the neighboring grid squares. Left and Right actions are successful 80% of the time. When not successful, the agent stays in place. From state a, there is also an exit action available, which results in going to the terminal state and collecting a reward of 10. Similarly, in state e, the reward for the exit action is 1. Exit actions are successful 100% of the time. Note that the agent is now in state d, as shown below.

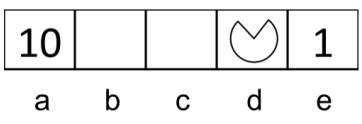


a) Let the discount factor $\gamma = 0.1$. What is the optimal action in the current state (d)?
b) Now let the discount factor $\gamma = 0.9999$. What is the optimal action in the current state (d)?

3.

a) $V_0(d) =$

Consider the same gridworld MDP as in the previous quiz, except that now Left and Right actions are 100% successful. Specifically, the available actions in each state are to move to the neighboring grid squares. From state a, there is also an exit action available, which results in going to the terminal state and collecting a reward of 10. Similarly, in state e, the reward for the exit action is 1. Exit actions are successful 100% of the time.



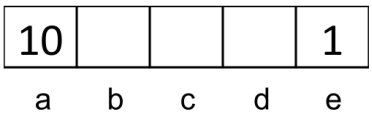
Let the discount factor $\gamma = 1$. Fill in the following quantities.

b) $V_1(d) =$
c) $V_2(d) =$

d) $V_3(d) =$
e) $V_4(d) =$
f) $V_5(d) =$

4.

Consider the same gridworld as in the previous quiz (where Left and Right actions are successful 100% of the time). Specifically, the available actions in each state are to move to the neighboring grid squares. From state a, there is also an exit action available, which results in going to the terminal state and collecting a reward of 10. Similarly, in state e, the reward for the exit action is 1. Exit actions are successful 100% of the time.



Let the discount factor $\gamma = 0.2$. Fill in the following quantities.

a) $V^*(a) = V_\infty(a) =$
b) $V^*(b) = V_\infty(b) =$
c) $V^*(c) = V_\infty(c) =$
d) $V^*(d) = V_\infty(d) =$
e) $V^*(e) = V_\infty(e) =$