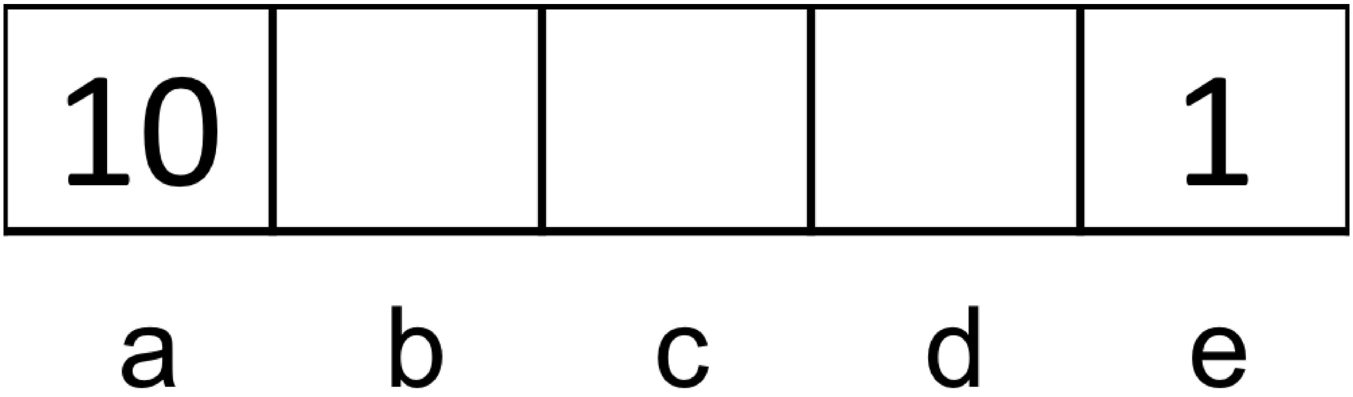
hw4_mdps_q2_value_iteration_convergence_values

Question 2: Value Iteration Convergence Values

0.0/5.0 points (graded)

Consider the gridworld 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.

$$V^*(a) = V_\infty(a) =$$

10

Answer: 10

The optimal action from a is to take the exit action.
Call t the terminal state.

$$V^*(a) = R(a, exit, t) + \gamma V^*(t) = 10$$

$$V^*(b) = V_\infty(b) =$$

2

Answer: 2

From state b, it is quite clear that you should move toward the closer, larger reward at state a.

$$V^*(b) = R(b, left, a) + \gamma V^*(a) = 2$$

$$V^*(c) = V_\infty(c) =$$

0.4

Answer: 0.4

From state c, you are equally close to both rewards, so the optimal action is to move toward the larger reward in state a.

$$V^*(c) = R(c, left, b) + \gamma V^*(b) = .4$$

$$V^*(d) = V_\infty(d) =$$

0.2

Answer: 0.2

It is not immediately obvious which way we should go from state d, so we must do some calculations first.

$$V^*(d) = max(R(d, left, c) + \gamma V^*(c), R(d, right, e) + \gamma V^*(e)) = max(.08, .2) = .2$$

Notice that from d, we prefer the closer, smaller reward to the farther, larger reward. This is because our discount factor (0.2) is low enough for us to prefer the closer reward. If our discount factor was higher, we might prefer the farther reward instead.

$$V^*(e) = V_\infty(e) =$$

1

Answer: 1

In state e, we have a similar situation as state d, where we could go for the closer, smaller reward or the farther, larger reward. However, because we know the correct action from state d is "Right", we know that state e will also prefer the closer reward.

$$V^*(e) = R(e, exit, t) + \gamma V^*(t) = 1$$

Submit