hw4_mdps_q6_policy_evaluation

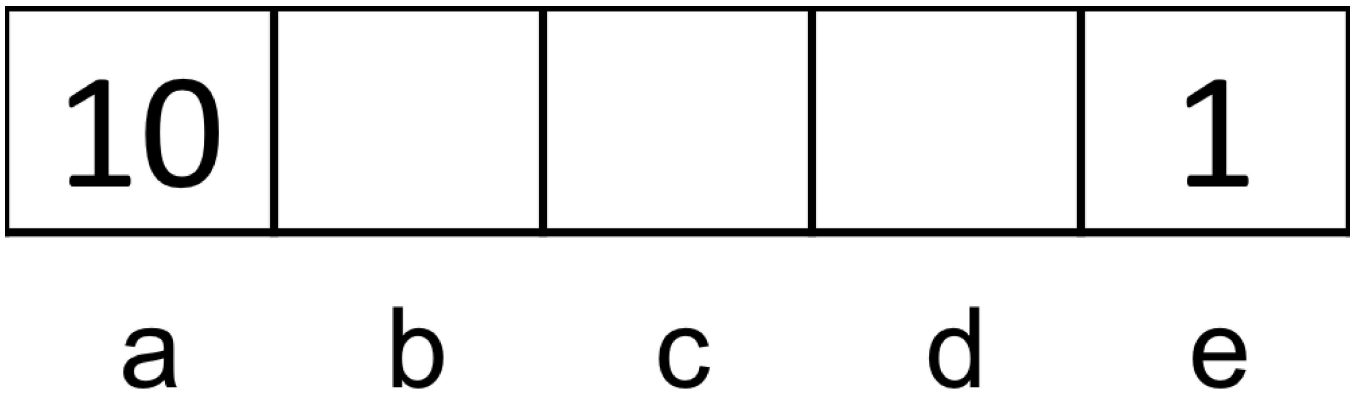
Question 6: Policy Evaluation

0.0/10.0 points (graded)

Consider the gridworld where Left and Right actions are successful 100% of the time.

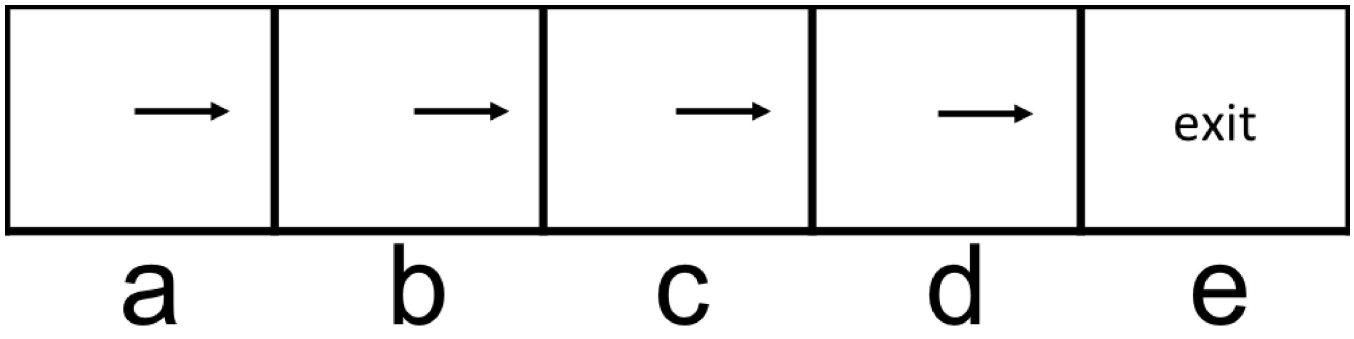
Specifically, the available actions in each state are to move to the neighboring grid squares. From state a, there is also an exit action available, which results in going to the terminal state and collecting a reward of 10. Similarly, in state e, the reward for the exit action is 1. Exit actions are successful 100% of the time.

The discount factor (γ) is 1.



Part 1

Consider the policy π_1 shown below, and evaluate the following quantities for this policy.



$V^{\pi_1}(a) =$

1 Answer: 1

$V^{\pi_1}(b) =$

1 Answer: 1

$V^{\pi_1}(c) =$

1 Answer: 1

$V^{\pi_1}(d) =$

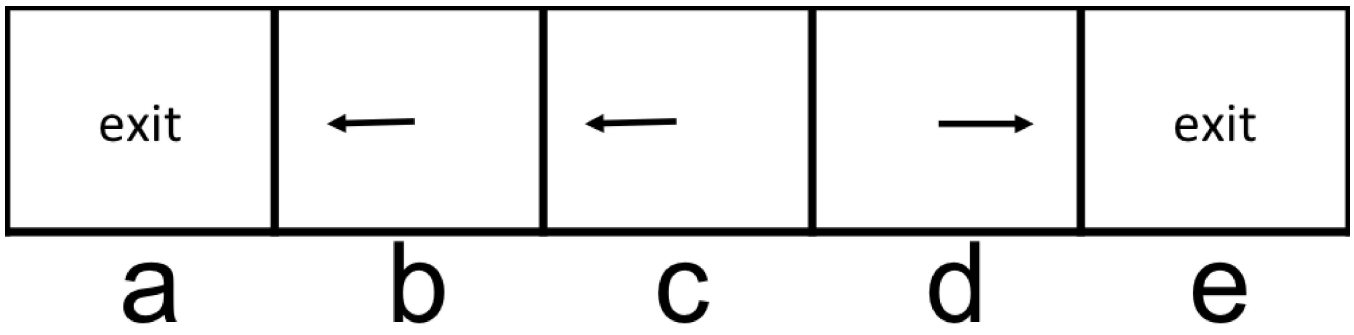
1 Answer: 1

$V^{\pi_1}(e) =$

1 Answer: 1

Part 2

Consider the policy π_2 shown below, and evaluate the following quantities for this policy.



$V^{\pi_2}(a) =$

10 Answer: 10

$V^{\pi_2}(b) =$

10 Answer: 10

$V^{\pi_2}(c) =$

10

Answer: 10

$V^{\pi_2}(d) =$

1

Answer: 1

$V^{\pi_2}(e) =$

1

Answer: 1

Because there is no discounting and the only reward you receive is for taking an exit action, the value of a state is determined only by which exit will be taken from that state according to the policy.

Part 1: Because you will exit at state e from every state, the value of every state will be 1.
Part 2: From states a, b, and c, you will exit from state a, so the value from these states is 10. From states d and e, you will exit from state e, so the value of these states is 1.

Submit

Answers are displayed within the problem