hw4_mdps_q7_policy_iteration

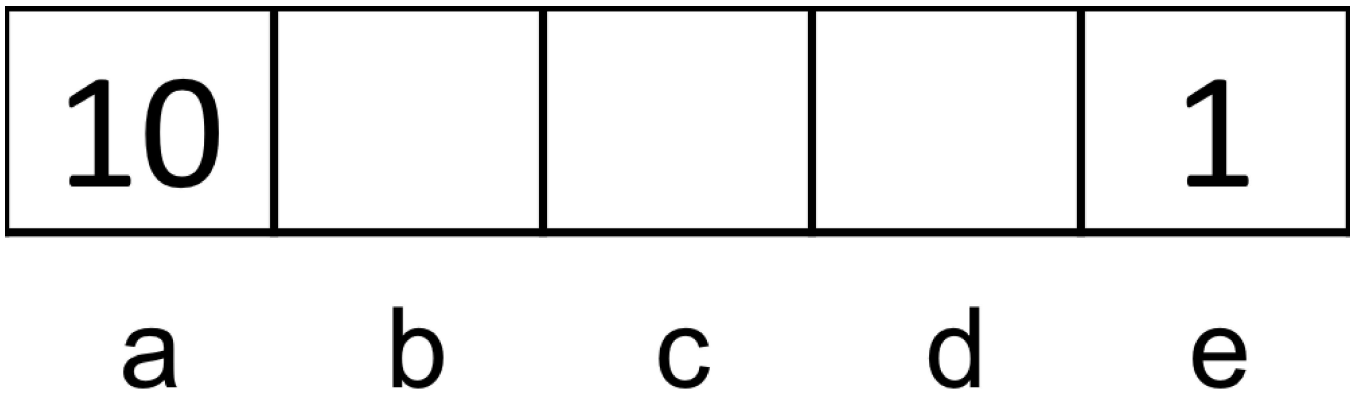
Question 7: Policy Iteration

5/5 points (ungraded)

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.

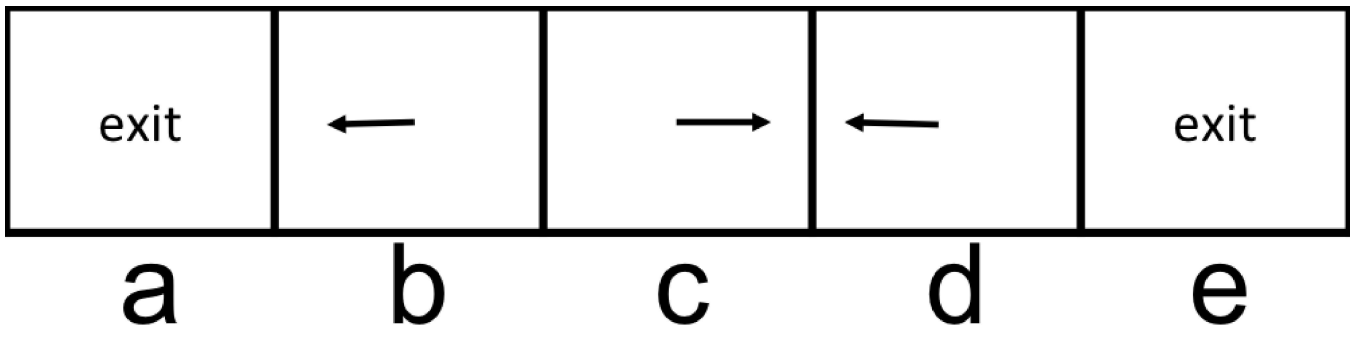
The discount factor (γ) is 0.9.



We will execute one round of policy iteration.

Part 1: Policy Evaluation

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



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

10 ✔

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

9 ✔

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

0 ✔

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

0 ✔

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

1 ✔

Submit

✔ Correct (5/5 points)

problem

5/5 points (ungraded)

Part 2: Policy Improvement

Perform a policy improvement step. The current policy's values are the ones from Part 1 (so make sure you first correctly answer Part 1 before moving on to Part 2).

$\pi_{i+1}(a) =$

- [x] Exit
- [ ] Right

$\pi_{i+1}(b) =$

- [x] Left
- [ ] Right

$\pi_{i+1}(c) =$

- [x] Left
- [ ] Right

$\pi_{i+1}(d) =$

- [ ] Left
- [x] Right

$\pi_{i+1}(e) =$

- [ ] Left
- [x] Exit

Submit

✔ Correct (5/5 points)