Question: MDP Properties

Which of the following statements are true for an MDP?

- [ ] If the only difference between two MDPs is the value of the discount factor then they must have the same optimal policy.

- [x] For an infinite horizon MDP with a finite number of states and actions and with a discount factor γ that satisfies $0 < \gamma < 1$, value iteration is guaranteed to converge.

0: everything converges to 0; 1: infinite actions

- [ ] When running value iteration, if the policy (the greedy policy with respect to the values) has converged, the values must have converged as well.

- [ ] None of the above

Question: Value Iteration Properties

Which of the following are true about value iteration? We assume the MDP has a finite number of actions and states, and that the discount factor satisfies $0 < \gamma < 1$.

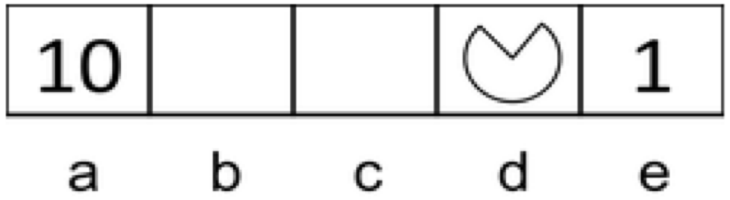
- [x] Value iteration is guaranteed to converge.

- [x] Value iteration will converge to the same vector of values (V^*) no matter what values we use to initialize V.

- [ ] None of the above

Question: Solving MDPs

Consider the gridworld MDP for which **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.

$V_0(d)=$ 0

$V_1(d)=$ 0

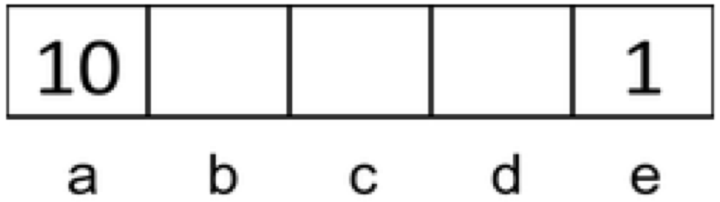
$V_2(d)=$ 1

$V_3(d)=$ 1

$V_4(d)=$ 10

Question: Value Iteration

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)= 10

V*(b)= 2

V*(c)= 0.4

V*(d)= 0.2

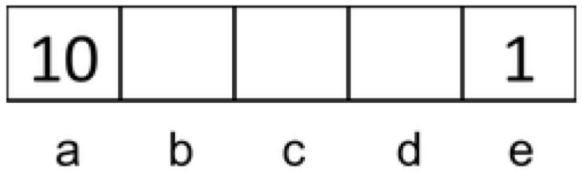
V*(e)= 1

Question: Policy Evaluation

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

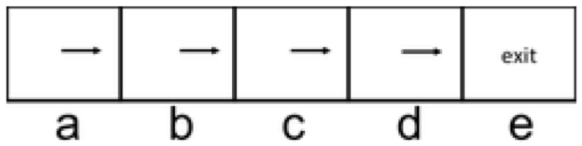
Specifically, the available actions in each state are to move to the neighboring grid squares. From state a, there is also an exit action available, which results in going to the terminal state and collecting a reward of 10. Similarly, in state e, the reward for the exit action is 1. Exit actions are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

V(b) = 1

V(c) = 1

V(d) = 1

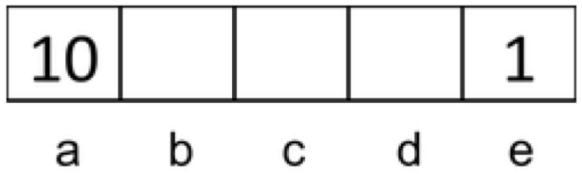
V(e) = 1

Question: Policy Evaluation

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

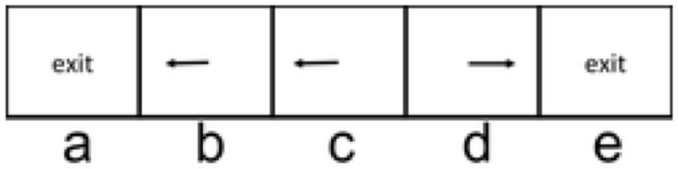
Specifically, the available actions in each state are to move to the neighboring grid squares. From state a, there is also an exit action available, which results in going to the terminal state and collecting a reward of 10. Similarly, in state e, the reward for the exit action is 1. Exit actions are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

The discount factor (γ) is 1.



Part 2

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



$V^{\pi_2}(a) =$ 10

V(b) = 10

V(c) = 10

V(d) = 1

V(e) = 1