CSPB 3202 - Truong - Artificial Intelligence

Dashboard / My courses / 2244:CSPB 3202 / 24 June - 30 June / Reading Quiz 5-MDP

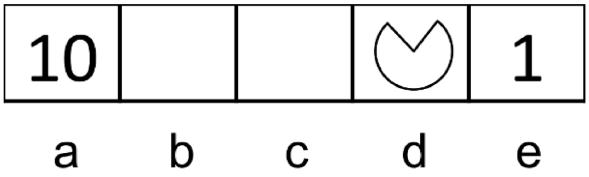
Started on	Thursday, 27 June 2024, 9:24 PM
State	Finished
Completed on	Thursday, 27 June 2024, 9:46 PM
Time taken	21 mins 26 secs

Question **1**

Correct

Marked out of 6.00

Consider the gridworld MDP for which Left and Right actions are 100% successful. Specially, 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.

Note: V_n(d) is the value for d at the n-th iteration.

$V_0(d) =$ 0 ✓

$V_1(d) =$ 0 ✓

$V_2(d) =$ 1 ✓

$V_3(d) =$ 1 ✓

$V_4(d) =$ 10 ✓

$V_5(d) =$ 10 ✓

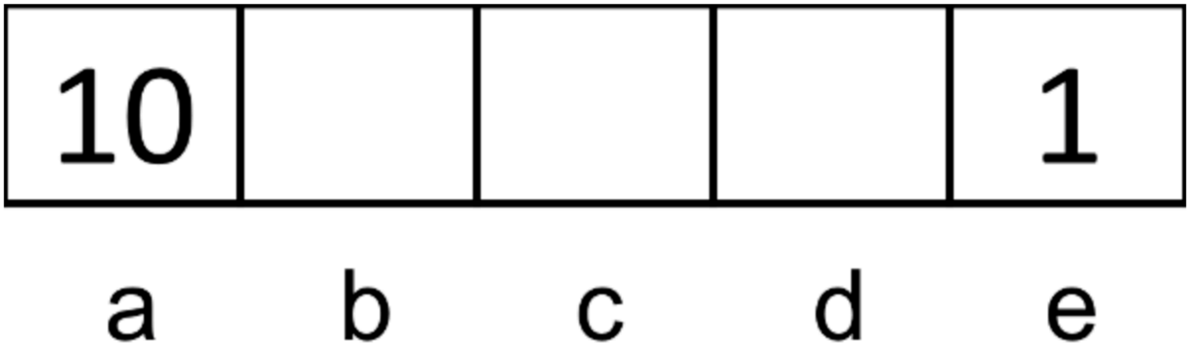
Question **2**

Correct

Marked out of 5.00

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 ✓

$V^*(b) = V_\infty(b) =$ 2 ✓

$V^*(c) = V_\infty(c) =$ 0.4 ✓

$V^*(d) = V_\infty(d) =$ 0.2 ✓

$V^*(e) = V_\infty(e)$ 1 ✓

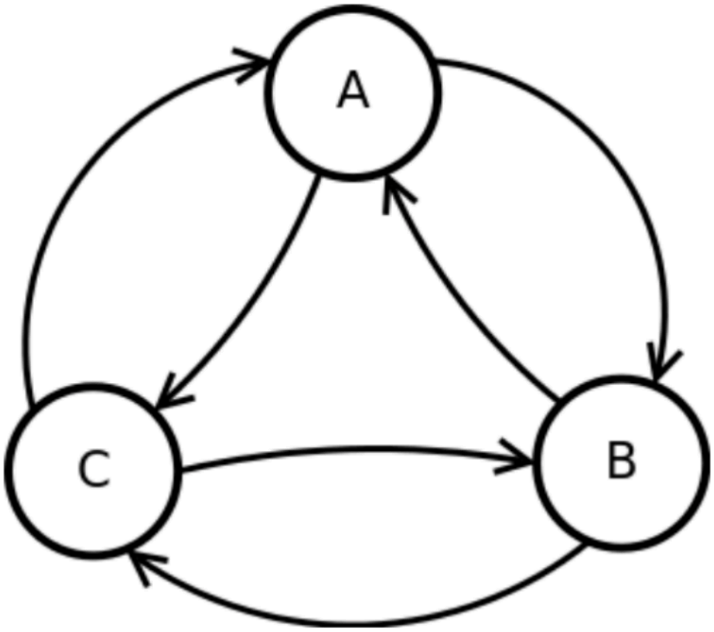
Question 3

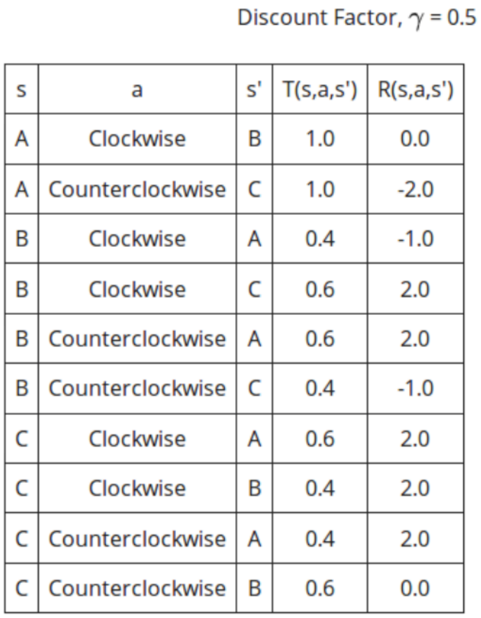
Correct

Marked out of 6.00

We recommend you work out the solutions to the following questions on a sheet of scratch paper, and then enter your results into the answer boxes.

Consider the following transition diagram, transition function and reward function for an MDP.

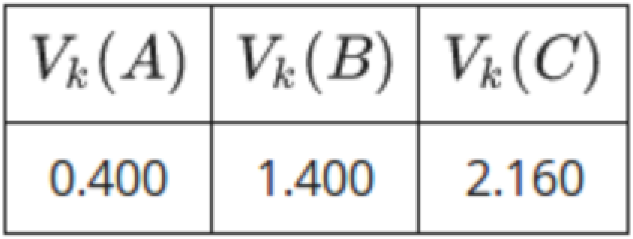




Discount Factor, γ = 0.5

s	a	s'	T(s,a,s')	R(s,a,s')
A	Clockwise	B	1.0	0.0
A	Counterclockwise	C	1.0	-2.0
B	Clockwise	A	0.4	-1.0
B	Clockwise	C	0.6	2.0
B	Counterclockwise	A	0.6	2.0
B	Counterclockwise	C	0.4	-1.0
C	Clockwise	A	0.6	2.0
C	Clockwise	B	0.4	2.0
C	Counterclockwise	A	0.4	2.0
C	Counterclockwise	B	0.6	0.0

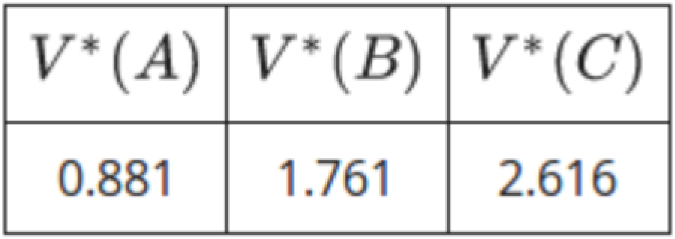
Suppose that after iteration k of value iteration we end up with the following values for Vk :



$V_k(A)$	$V_k(B)$	$V_k(C)$
0.400	1.400	2.160

What is $V_{k+1}(A)$? 0.7 ✓

Now, suppose that we ran value iteration to completion and found the following value function, V*.



$V^*(A)$	$V^*(B)$	$V^*(C)$
0.881	1.761	2.616

What is $Q^*(A, clockwise)$? 0.8805 ✓

What is the optimal action from state A? Clockwise ✓

Question **4**

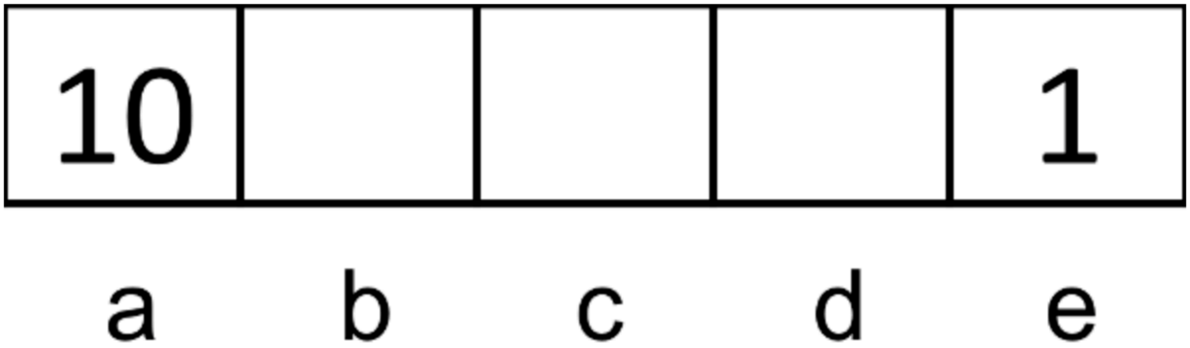
Correct

Marked out of 10.00

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

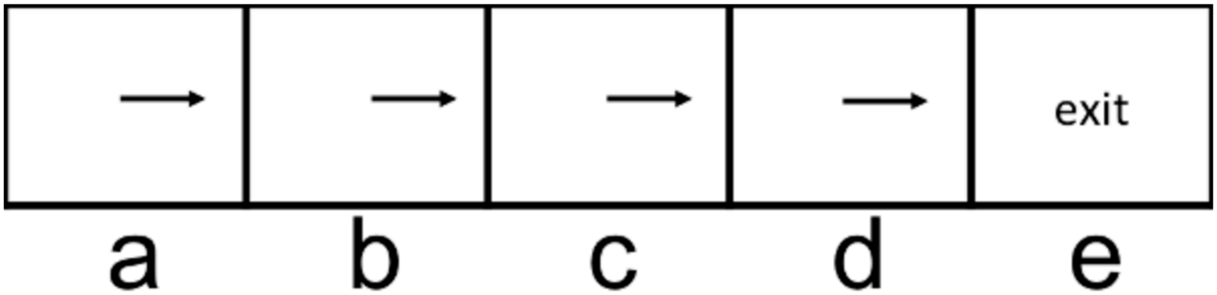
Specifically, the available actions in each state are to move to the neighboring grid squares. From state , there is also an exit action available, which results in going to the terminal state and collecting a reward of 10. Similarly, in state , the reward for the exit action is 1. Exit actions are successful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

$V^{\pi_1}(b) =$ 1 ✓

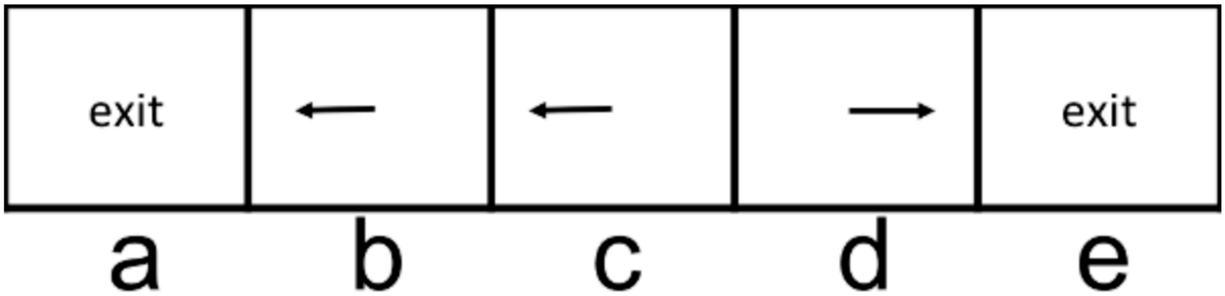
$V^{\pi_1}(c) =$ 1 ✓

$V^{\pi_1}(d) =$ 1 ✓

$V^{\pi_1}(e) =$ 1 ✓

Part 2

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



$V^{\pi_2}(a) =$ 10 ✓

$V^{\pi_2}(b) =$ 10 ✓

$V^{\pi_2}(c) =$ 10 ✓

$V^{\pi_2}(d) =$ 1 ✓

$V^{\pi_2}(e) =$ 1 ✓

Question **5**

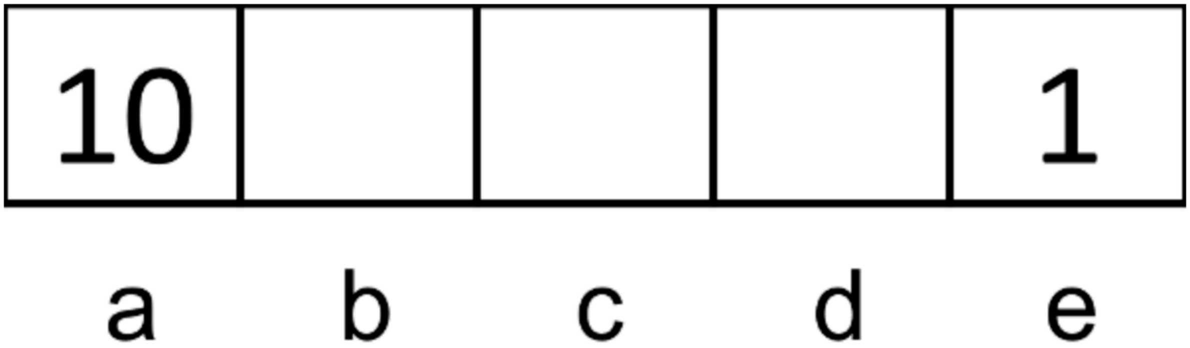
Correct

Marked out of 5.00

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

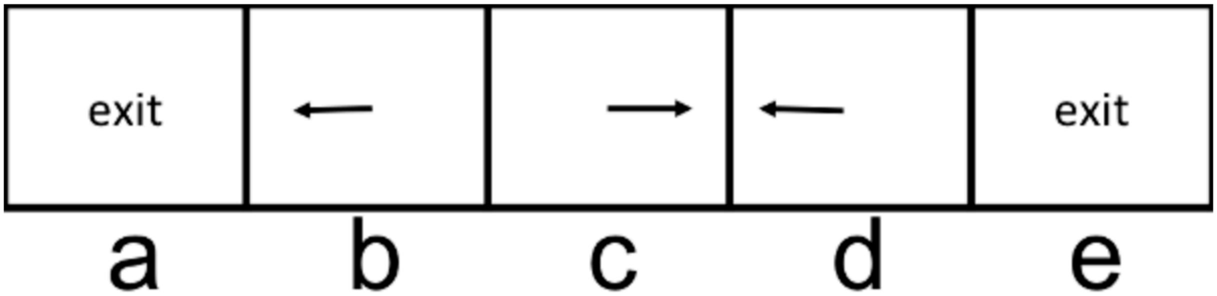
Specifically, the available actions in each state are to move to the neighboring grid squares. From state , there is also an exit action available, which results in going to the terminal state and collecting a reward of 10. Similarly, in state e, the reward for the exit action is 1. Exit actions are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

The discount factor γ is 0.9.



We will execute one round of policy iteration.

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 ✓

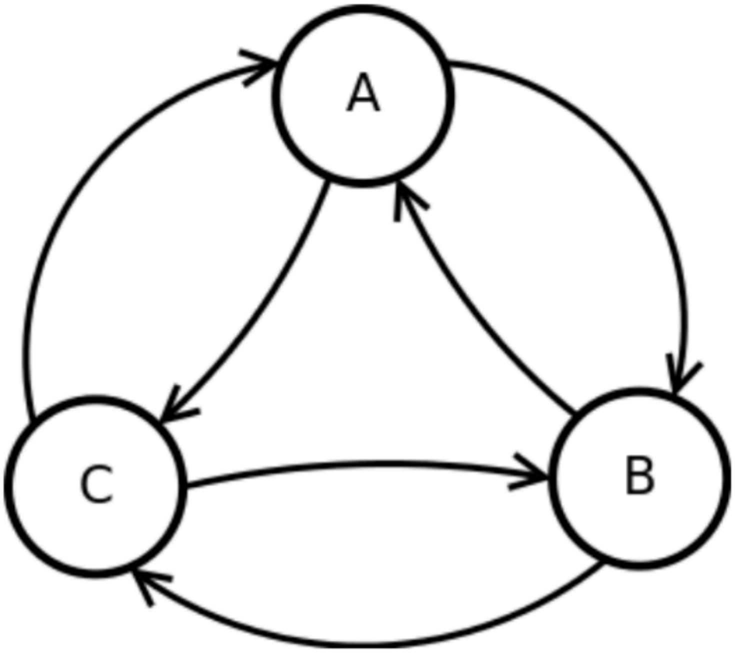
$V^{\pi_i}(e) =$ 1 ✓

Question **6**

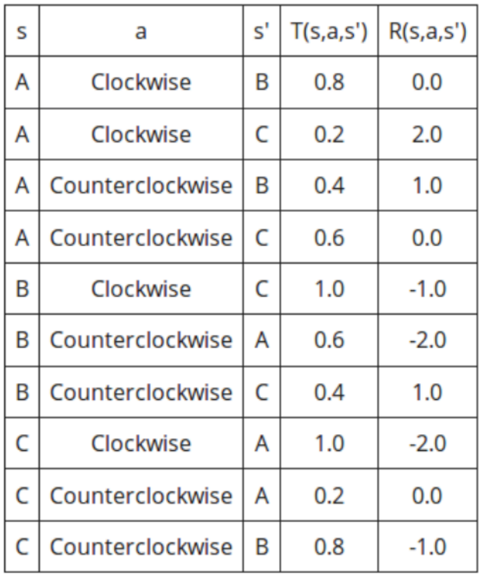
Correct

Marked out of 8.00

Consider the following transition diagram, transition function and reward function for an MDP.

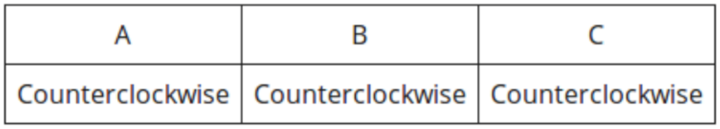


Discount Factor, γ = 0.5

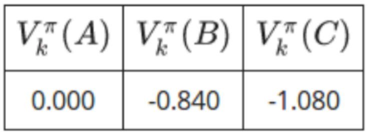


s	a	s'	T(s,a,s')	R(s,a,s')
A	Clockwise	B	0.8	0.0
A	Clockwise	C	0.2	2.0
A	Counterclockwise	B	0.4	1.0
A	Counterclockwise	C	0.6	0.0
B	Clockwise	C	1.0	-1.0
B	Counterclockwise	A	0.6	-2.0
B	Counterclockwise	C	0.4	1.0
C	Clockwise	A	1.0	-2.0
C	Counterclockwise	A	0.2	0.0
C	Counterclockwise	B	0.8	-1.0

Suppose we are doing policy evaluation, by following the policy given by the left-hand side table below. Our current estimates (at the end of some iteration of policy evaluation) of the value of states when following the current policy is given in the right-hand side table.



A	B	C
Counterclockwise	Counterclockwise	Counterclockwise



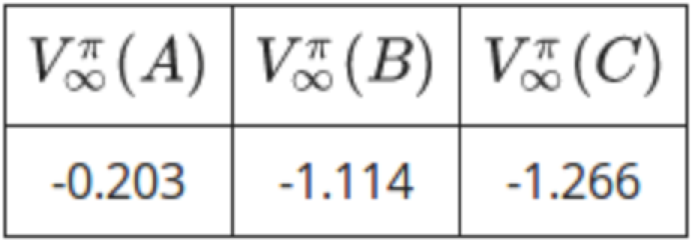
$V_k^\pi(A)$	$V_k^\pi(B)$	$V_k^\pi(C)$
0.000	-0.840	-1.080

We recommend you work out the solutions to the following questions on a sheet of scratch paper, and then enter your results into the answer boxes.

Part 1

What is $V_{k+1}^\pi(A)$? -0.092 ✓

Suppose that policy evaluation converges to the following value function, V_∞^π..



$V_\infty^\pi(A)$	$V_\infty^\pi(B)$	$V_\infty^\pi(C)$
-0.203	-1.114	-1.266

Now let's execute policy improvement.

Part 2

What is $Q^{\pi}_{\infty}(A, clockwise)$? -0.1722 ✓

Part 3

What is $Q^{\pi}_{\infty}(A, counterclockwise)$? -0.2026 ✓

Part 4

What is the updated action for state A? Clockwise ✓

Question **7**

Correct

Marked out of 2.00

Which of the following statements are true for an MDP?

Select one:

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

● For an infinite horizon MDP with a finite number of states and actions and with a discount factor γ that satisfies 0 < γ <1, value iteration is guaranteed to converge. ✓

○ 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 **8**

Correct

Marked out of 2.00

Which of the following statements are true for an MDP?

Select one or more:

☑ If one is using value iteration and the values have converged, the policy must have converged as well. ✓

□ Expectimax will generally run in the same amount of time as value iteration on a given MDP.

☑ For an infinite horizon MDP with a finite number of states and actions and with a discount factor γ that satisfies $0 < \gamma < 1$, policy iteration is guaranteed to converge. ✓

□ None of the above

Question **9**

Correct

Marked out of 3.00

John, James, Alvin and Michael all get to act in an MDP $(S, A, T, \gamma, R, s_0)$.

John runs value iteration until he finds V^*which satisfies $\forall s \in S V^*(s) = \max_{a \in A} \sum_{s'} T(s, a, s')(R(s, a, s') + \gamma V^*(s'))$ and acts according to $\pi_{\text{John}} = \arg\max_{a \in A} \sum_{s'} T(s, a, s')(R(s, a, s') + \gamma V^*(s'))$

James acts according to an arbitrary policy π_{James}.

Alvin takes James's policy $\pi_{\text{James}}$and runs one round of policy iteration to find his policy π_{Alvin}.

Michael takes John's policy and runs one round of policy iteration to find his policy π_{Michael}.

Note: One round of policy iteration = performing policy evaluation followed by performing policy improvement.

Select the answer that are guaranteed to be true:

Select one:

○ It is guaranteed that $\forall s \in S : V^{\pi_{\text{James}}}(s) \geq V^{\pi_{\text{Alvin}}}(s)$

● It is guaranteed that $\forall s \in S : V^{\pi_{\text{Michael}}}(s) \geq V^{\pi_{\text{Alvin}}}(s)$ ✓

○ It is guaranteed that $\forall s \in S : V^{\pi_{\text{Micheal}}}(s) \geq V^{\pi_{\text{John}}}(s)$

○ It is guaranteed that $\forall s \in S : V^{\pi_{\text{James}}}(s) \geq V^{\pi_{\text{John}}}(s)$

○ None of the above.

Question **10**

Correct

Marked out of 3.00

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 <γ< 1.

Select one or more:

☑ Value iteration is guaranteed to converge. ✓

☑ Value iteration will converge to the same vector of values (V^*)no matter what values we use to initialize V. ✓

□ None of the above.