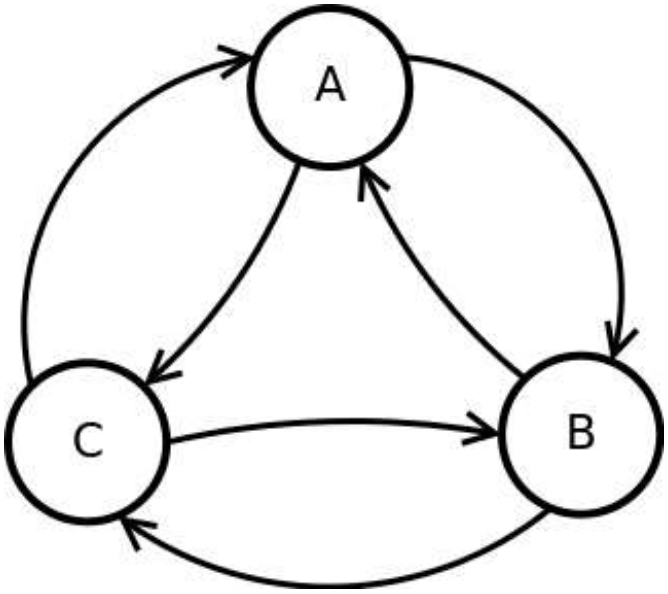
hw4_mdps_q3_value_iteration_cycle

Question 3: Value Iteration: Cycle

0.0/16.0 points (graded)

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, $\gamma = 0.5$

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

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

$V_k\,(A)$	$V_k\,(B)$	$V_k\,(C)$
2.000	2.000	0.700

Part 1: What is $V_{k+1}\,(B)$?

2.35

Answer: 2.35

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

$V^*(A)$	$V^*(B)$	$V^*(C)$
2.424	2.606	1.212

Part 2: What is Q^*(B, clockwise)?

2.606

Answer: 2.60606060606

Part 3: What is Q^*(B, counterclockwise)?

2.2908

Answer: 2.29090909091

Part 4: What is the optimal action from state B? Enter clockwise or counterclockwise.

Clockwise

Answer: Clockwise

This question is randomized. Here are the general formulas to solve these problems.

Part 1: Here is the formula to calculate $V_{k+1}(A)$.

$$V_{k+1}(A) = max\left(Q_{k+1}(A, clockwise), Q_{k+1}(A, counterclockwise)\right)$$
$$Q_{k+1}(A, clockwise) = T(A, clockwise, B)\left[R(A, clockwise, B) + \gamma V_k(B)\right] + T(A, clockwise, C)\left[R(A, clockwise, C) + \gamma V_k(C)\right]$$
$$Q_{k+1}(A, counterclockwise) = T(A, counterclockwise, B)\left[R(A, counterclockwise, B) + \gamma V_k(B)\right] + T(A, counterclockwise, C)\left[R(A, counterclockwise, C) + \gamma V_k(C)\right]$$

Part 2: Here is the formula to calculate $Q^*(A, clockwise)$.

$$Q^*(A, clockwise) = T(A, clockwise, B)\left[R(A, clockwise, B) + \gamma V^*(B)\right] + T(A, clockwise, C)\left[R(A, clockwise, C) + \gamma V^*(C)\right]$$

Part 3: Here is the formula to calculate $Q^*(A, counterclockwise)$.

$$Q^*(A, counterclockwise) = T(A, counterclockwise, B)\left[R(A, counterclockwise, B) + \gamma V^*(B)\right] + T(A, counterclockwise, C)\left[R(A, counterclockwise, C) + \gamma V^*(C)\right]$$

Part 4: The optimal action from A is the action that gives us the highest Q^* value.

Submit