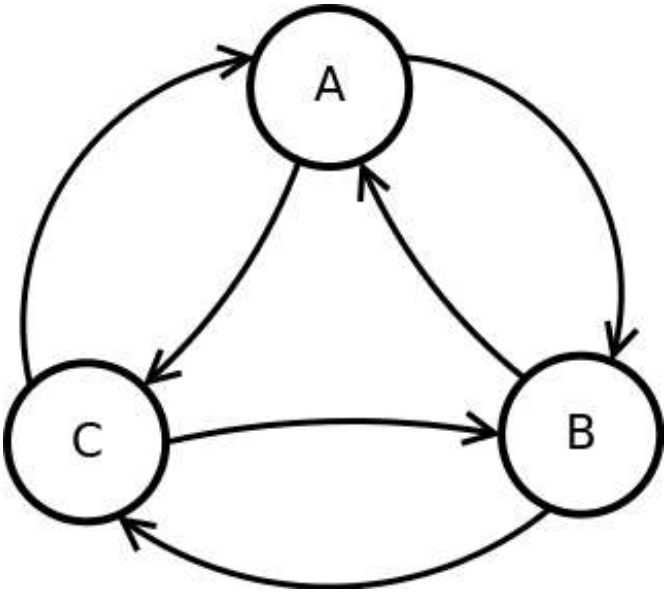
hw4_mdps_q8_policy_iteration_cycle

Question 8: Policy 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	0.8	0.0
A	Clockwise	C	0.2	2.0
A	Counterclockwise	B	0.2	-1.0
A	Counterclockwise	C	0.8	1.0
B	Clockwise	A	0.2	-1.0
B	Clockwise	C	0.8	0.0
B	Counterclockwise	A	0.8	-2.0
B	Counterclockwise	C	0.2	-1.0
C	Clockwise	A	0.6	2.0
C	Clockwise	B	0.4	-1.0
C	Counterclockwise	B	1.0	0.0

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
Clockwise	Clockwise	Counterclockwise

$V_k^\pi(A)$	$V_k^\pi(B)$	$V_k^\pi(C)$
0.320	-0.160	-0.100

Part 1: What is $V_{k+1}^\pi(A)$?

0.326

Answer: $\mathbf{0.326}$

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

$V_\infty^\pi(A)$	$V_\infty^\pi(B)$	$V_\infty^\pi(C)$
0.305	-0.212	-0.106

Now let's execute policy improvement.
Part 2: What is Q_∞^π(A, clockwise)?

0.3046

Answer: 0.304635761589

Part 3: What is Q_∞^π(A, counterclockwise)?

0.5364

Answer: 0.53642384106

Part 4: What is the updated action for state A? Enter clockwise or counterclockwise.

Counterclockwise

Answer: Counterclockwise

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

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

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

We only take into account the clockwise action from A, because that is the action according to our policy.

Part 2: Here is the formula to calculate $Q_\infty^\pi(A, clockwise)$.

$$Q_\infty^\pi(A, clockwise) = T(A, clockwise, B)\left[R(A, clockwise, B) + \gamma V_\infty^\pi(B)\right] + T(A, clockwise, C)\left[R(A, clockwise, C) + \gamma V_\infty^\pi(C)\right]$$

Part 3: Here is the formula to calculate $Q_\infty^\pi(A, counterclockwise)$.

$$Q_\infty^\pi(A, counterclockwise) = T(A, counterclockwise, B)\left[R(A, counterclockwise, B) + \gamma V_\infty^\pi(B)\right] + T(A, counterclockwise, C)\left[R(A, counterclockwise, C) + \gamma V_\infty^\pi(C)\right]$$

Part 4: The updated action for state A will be the action that results in the higher Q_∞^π.

Submit