hw4_mdps_q3_value_iteration_cycle

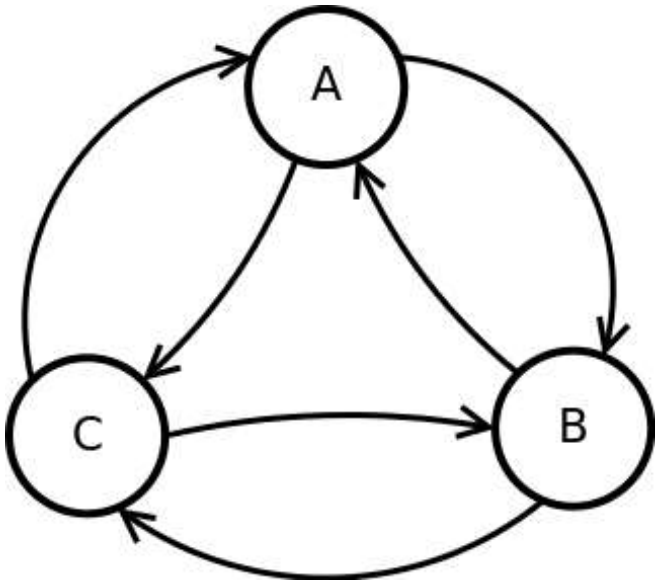
Question 3: Value Iteration: Cycle

16/16 points (ungraded)

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	0.8	-2.0
A	Clockwise	C	0.2	1.0
A	Counterclockwise	B	0.2	-1.0
A	Counterclockwise	C	0.8	0.0
B	Clockwise	C	1.0	1.0
B	Counterclockwise	A	0.6	-2.0
B	Counterclockwise	C	0.4	-2.0
C	Clockwise	A	0.6	-1.0
C	Clockwise	B	0.4	1.0
C	Counterclockwise	B	1.0	0.0

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

$V_k\,(A)$	$V_k\,(B)$	$V_k\,(C)$
-0.100	1.000	0.500

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

0.5 ✔

Correct: Your answer evaluated to 0.500, which is close enough to the correct answer, 0.500.

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

$V^*(A)$	$V^*(B)$	$V^*(C)$
0.200	1.333	0.667

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

0.1268 ✔

Correct: Your answer evaluated to 0.127, which is close enough to the correct answer, 0.127.

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

0.667 ✔

Correct: Your answer evaluated to 0.667, which is close enough to the correct answer, 0.667.

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

counterclockwise ✔

Submit

✔ Correct (16/16 points)