

EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the [Privacy Policy](#). ×



[Unit 5 Reinforcement Learning](#) (2

[Course](#) > [weeks](#))

> [Homework 6](#) > 3. Q-Learning

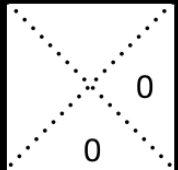
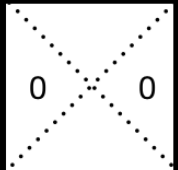
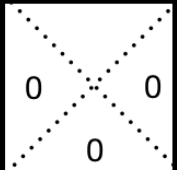
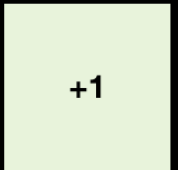
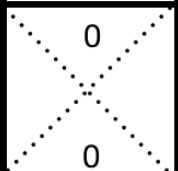
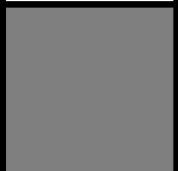
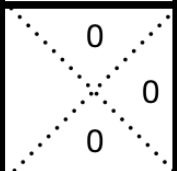
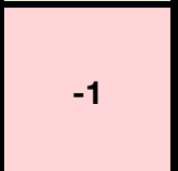
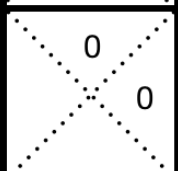
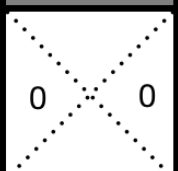
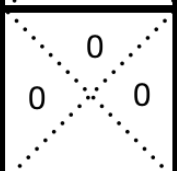
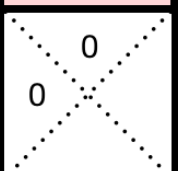
3. Q-Learning

Recall the Q-learning update rule:

$$Q_{i+1}(s, a) = Q_i(s, a) + \alpha [R(s, a, s') + \gamma \max_{a'} Q_i(s', a') - Q_i(s, a)]$$

let $\alpha = 1$ and $\gamma = 1$ in this problem. In the figure below, at each box, we can go up, down, left and right unless the path is blocked and we initialize the Q value for all the actions in all states as 0. The Q value for the 4 directions are labeled in each box below. Moving into the upper right 2 boxes will result in a reward of +1 and -1, and each move will also cost 0.04, or in another word, a reward of -0.04.

Q-table

1st Iteration

3/3 points (graded)

Q-table

-0.04 -0.04 -0.04	-0.04 -0.04 -0.04	-0.04 -0.04 -0.04	x -0.04 -0.04 +1
-0.04 -0.04 -0.04		-0.04 -0.04 -0.04	y -0.04 -0.04 -1
-0.04 -0.04 -0.04	-0.04 -0.04 -0.04	-0.04 -0.04 -0.04	z -0.04 -0.04 -0.04

After 1st iteration, enter the Q value at the position represented by x , y and z below:

$x =$

✓ Answer: 0.96

$y =$

✓ Answer: -1.04

$z =$

✓ Answer: -1.04

Solution:

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

2nd Iteration

3/3 points (graded)

Q-table

<div> <div>-0.08</div> <div>-0.08</div> </div>	<div> <div>-0.08</div> <div>a</div> </div>	<div> <div>-0.08</div> <div>x</div> </div>	+1
<div> <div>-0.08</div> <div>-0.08</div> </div>		<div> <div>b</div> <div>y</div> <div>c</div> </div>	-1
<div> <div>-0.08</div> <div>-0.08</div> </div>	<div> <div>-0.08</div> <div>-0.08</div> </div>	<div> <div>-0.08</div> <div>-0.08</div> </div>	z

After 2nd iteration, enter the Q value at the position represented by *a*, *b* and *c* below:

$a =$ ✓ Answer: 0.92

$b =$ ✓ Answer: 0.92

$c =$ ✓ Answer: -0.08

Solution:

Submit

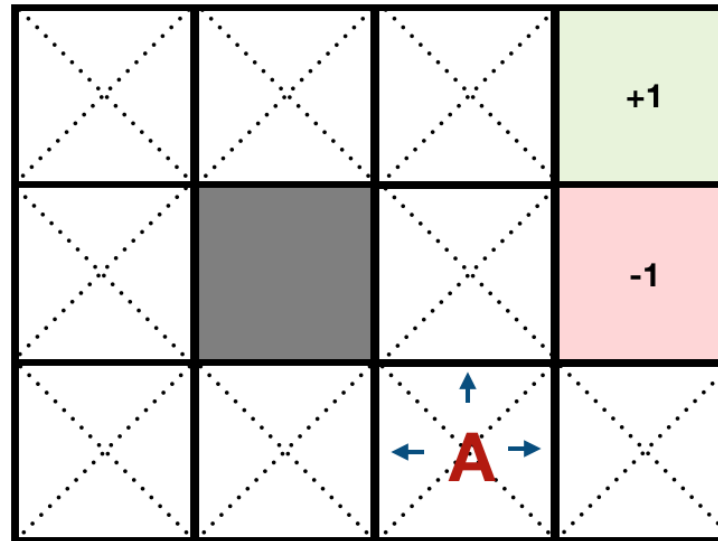
You have used 1 of 3 attempts

i Answers are displayed within the problem

2nd Iteration

1/1 point (graded)

Q-table



After convergence, at state A, which action is the optimal?

☒ UP ✓

☐ LEFT

☐ RIGHT

Solution:

Submit

You have used 1 of 1 attempt

 Answers are displayed within the problem

Epsilon-greedy method 1

1/1 point (graded)

In the ε -greedy method, a larger value of ε would generate experiences that are more consistent with the current Q-value estimates.

☐ True

☒ False 

Solution:

In the ε -greedy method, we choose a random action with probability ε and choose an action based on our current estimates with probability $1 - \varepsilon$. Therefore, it is with smaller ε that we would generate experiences which are more consistent with our current Q-value estimates.

Submit

You have used 1 of 1 attempt

 Answers are displayed within the problem

Epsilon-greedy method 2

1/1 point (graded)

In the ε -greedy method, a value of $\varepsilon = 0.999$ is likely to lead to the desired learning outcome in a highly complex environment.

☐ True

☒ False ✓

Solution:

We would pick a random action virtually every time, and in a highly complex environment, it's highly unlikely that we would properly explore the parts of the space that have high rewards.

Submit

You have used 1 of 1 attempt

 Answers are displayed within the problem

Discussion

[Hide Discussion](#)

Topic: Unit 5 Reinforcement Learning (2 weeks) :Homework 6 / 3. Q-Learning

[Add a Post](#)

Show all posts ▼	by recent activity ▼
<p>🔍 [Staff] 2nd Iteration Q Table</p> <p>If ethically permitted, can anyone explain to me why the max Q' for c isn't -.08 or 0? I tried both and was graded incorrect. As far as I can see, oth...</p>	8
<p>💬 @Staff omit Epsilon-greedy method 2 please</p> <p>Answer is depended on assumptions that are not touched in the question. Please omit this question. In general, this homework(and this module)...</p>	10
<p>💬 [staff] Iteration (2) If 'b' and 'c' are at the same state, why are not their values the same?</p>	4
<p>💬 2nd Iteration</p> <p>1. Is this the equation being used for Q2. $Q_2 = Q_1 + \alpha * (R + \gamma * \max_{a'}(Q_1(s', a')) - Q_1)$ 2. For second iteration, do we look at the state of each...</p>	5
<p>🔍 1st iteration - add attempt</p> <p>Hi, is there anyway I could have one more attempt? I didn't mean to leave negatives in my answer. Thanks!</p>	2
<p>✅ [Staff] About Epsilon-greedy method 2</p> <p>I think in a HIGHLY COMPLEX environment it is better to consider how to easily jump out from a dead-cycle(in other words: local optimal) which c...</p>	9
<p>💬 Cool Tool ;)</p> <p>https://www.mladdict.com/q-learning-simulator play around and learn how Q-learning works</p>	3
<p>🔍 [Staff] Epsilon-greedy method 2</p> <p>Is question referring to epsilon at the beginning of the learning or end of the learning or is it a fixed value?</p>	3 new_ 11
<p>🔍 Did anyone figure out 1st iteration? Is x moving or staying put? If so, where is it going?</p>	7

? <u>1st iteration: where was x, y, z before, and where is it going?</u>	3
☑ <u>[Staff] 1st Iteration - Need help</u>	7 new_ 17
? <u>Acting optimally from cells '-1' and '+1'</u>	4
💬 <u>Q-Learning: Squares marked +1 and -1 are terminal?</u> <u>No transition possible out of these two states on any of the actions? Once we get in there we stay there?</u>	3
💬 <u>The route to convergence</u>	3

© All Rights Reserved