



MITx 6.86x

**Machine Learning with Python-From Linear Models to Deep Learning**[Course](#)[Progress](#)[Dates](#)[Discussion](#)[Resources](#)[🏠 Course](#) / [Unit 5. Reinforcement Learning \(2 weeks\)](#) / [Homework 5](#)[< Previous](#)

### 3. Q-value iteration

[🔖 Bookmark this page](#)

Homework due May 3, 2023 08:59 -03 Completed

Recall the Q-value iteration update rule:

$$Q^*(s, a) = \sum_{s'} T(s, a, s') \left( R(s, a, s') + \gamma \max_{a'} Q^*(s', a') \right).$$

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

**Q-table**

0 0	0 0	0 0	+1
0 0		0 0	-1
0 0	0 0	0 0	0 0

## 1st Iteration

3/3 points (graded)

**Q-table**

-0.04 -0.04	-0.04 -0.04	-0.04 -0.04	+1
-0.04 -0.04		-0.04 -0.04	-1
-0.04 -0.04	-0.04 -0.04	-0.04 -0.04	-0.04 -0.04

## 2nd Iteration

3/3 points (graded)

**Q-table**

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

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



Submit

You have used 1 of 3 attempts

## After Convergence

1/1 point (graded)

**Q-table**

<div></div> <div></div> <div></div> <div></div>	<div></div> <div></div> <div></div> <div></div>	<div></div> <div></div> <div></div> <div></div>	<div>+1</div>
---	---	---	---------------

☐ RIGHT

You have used 1 of 1 attempt

## Epsilon-greedy method 1

1/1 point (graded)

In the  $\epsilon$ -greedy method, a larger value of  $\epsilon$  would generate experiences that are more  $\epsilon$ -greedy than current Q-value estimates.

☐ True☒ False

You have used 1 of 1 attempt

## Epsilon-greedy method 2

1/1 point (graded)

In the  $\epsilon$ -greedy method, a value of  $\epsilon = 0$  is likely to lead to the desired learning outcome.

[< Previous](#)[Next >](#)☐ True**edX**[About](#)[Affiliates](#)[edX for Business](#)[Open edX](#)

I wouldn't mind, though, if it was a bit more challenging (it feels like it was one of the easiest 100% I got in the

💬 2nd Iteration of Q-values

Is b the Q-value if you are in position y and going to the state directly above and c the Q-value if you are in p

💬 epsilon-greedy question

I feel like this question is missing a lot of context. Does it mean epsilon = .999 permanently and is never adju

💬 Hint on  $\max_a(Q_k(s,a))$

Just in case anyone has been struggling for a while (like me) to wonder why the Q values computed are not t

## Connect

[Blog](#)

[Contact Us](#)

[Help Center](#)

[Security](#)

[Media Kit](#)



© 2023 edX LLC. All rights reserved.

深圳市恒宇博科技有限公司 [粤ICP备17044299号-2](#)