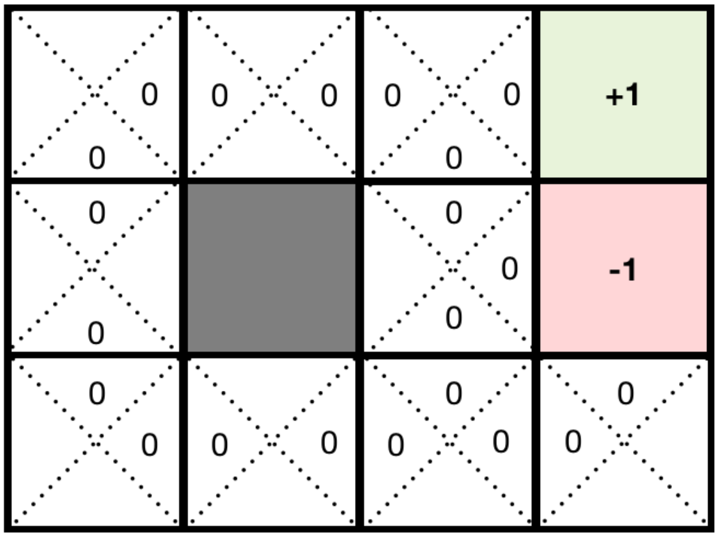
3. Q-Learning

Recall the Q-learning update rule:

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

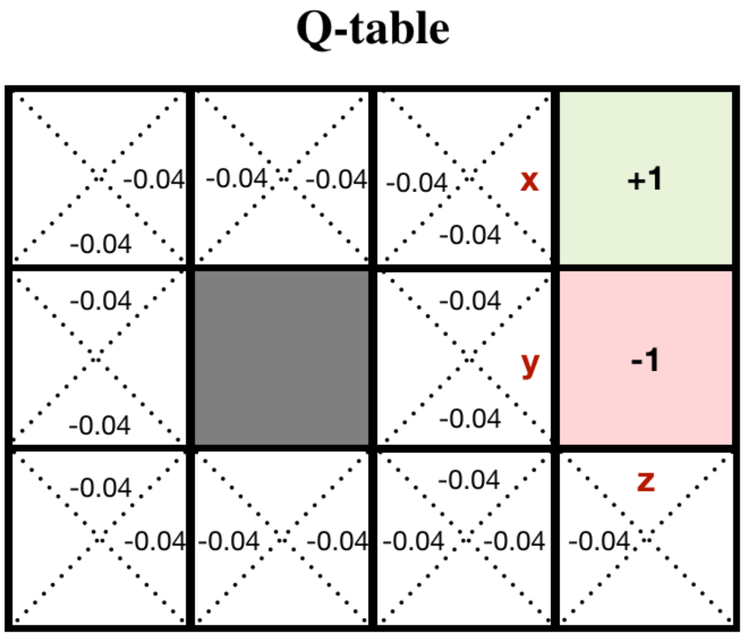
let $\alpha=1$ and $\gamma=1$ in this problem. In the figure below, at each box, we can go up, down, left and right unless the path is blocked and we initialize the Q value for all the actions in all states as 0. The Q value for the 4 directions are labeled in each box below. Moving into the upper right 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 points possible (graded)


Q-table

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

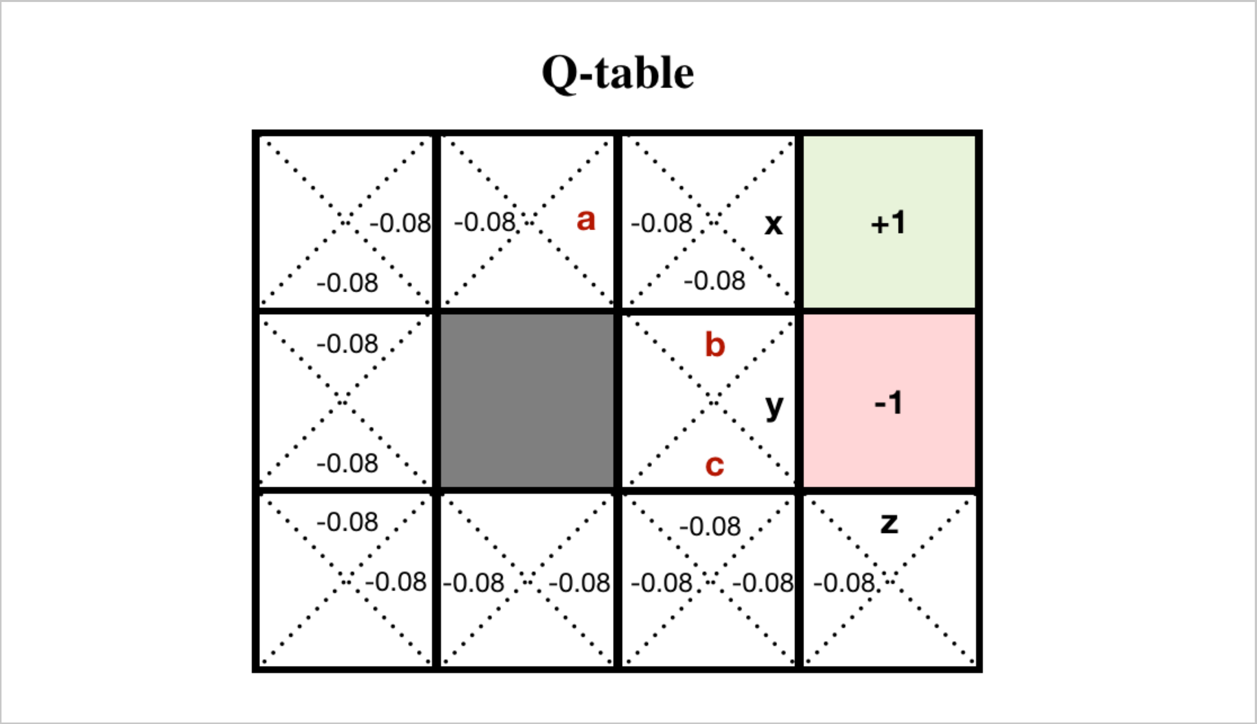
Solution:

Submit You have used 0 of 3 attempts

Answers are displayed within the problem

2nd Iteration

3 points possible (graded)



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

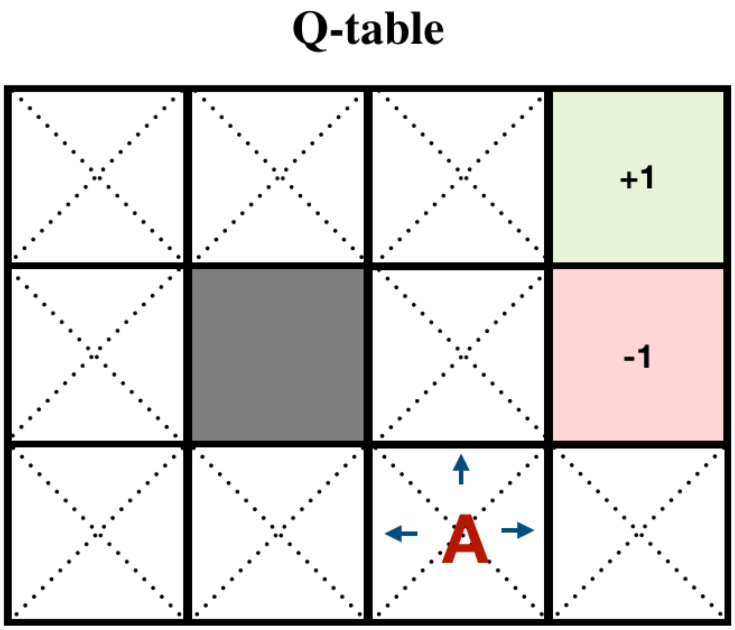
Solution:

Submit You have used 0 of 3 attempts

Answers are displayed within the problem

2nd Iteration

1 point possible (graded)



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

○ UP ✓

○ LEFT

○ RIGHT

Solution:

Submit

You have used 0 of 1 attempt

Answers are displayed within the problem

Epsilon-greedy method 1

1 point possible (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 - ε. 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 0 of 1 attempt

Answers are displayed within the problem

Epsilon-greedy method 2

1 point possible (graded)

In the ε-greedy method, a value of $\varepsilon = 0.999$ is likely to lead to the desired learning outcome (better utility) 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 0 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

- ? Does the formula given at the begining of the excersise applies here? Maybe I am wrong, but as this formula is the given for "Q value iteration by sampling" (Lectur... 4
- ☑ What should be the value for Max(Q(s',a')) when it's at the edge state? After reading the answers, I'm still confused as I could not get the results as what's supposed... 2
- 💬 Question about reward value initialization for blocked directions? How should we initialize the reward values for those blocked directions? I found when I initial... 3
- 💬 2nd Iteration Q-values Any hint how to solve 2nd iteration Q value at the position a, b, and c? Also, why reward is no... 4
- ☑ [Staff] Epsilon-greedy method 2 It appears that assumptions needed to solve this problem are "implicit". Could staff make the... 6
- ? I'm completely lost- some hint to start ? see up 7
- ? Epsilon-greedy method 1 Apparently this question is NOT related to the previous question. I thought current Q-value e... 1
- ☑ [STAFF] Progress bar for this homework. 3

? Epsilon-Greedy method 2 clarifications

Since epsilon-greedy method starts by first exploring the environment with probability epsilo... 9

💬 [Staff] Score not updated

I got 34/36 for this homework, but according to the progress chart I only got 32/36. 2

☑ [Staff] 1st Iteration doubts...

I've managed to get the Q value for position x correct, but not sure where I'm going wrong fo... 4

? 1st iteration

Please increase attempts by 1 or 2. I lost all of the attempts but couldn't get the answer 2

💬 [staff] Is iteration exercise 2 built on results from iteration 1 exercise?

Is iteration 2 exercise built on the result from iteration 1 exercise, or is iteration 2 exercise a s... 3