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 <a href="Privacy Policy">Privacy Policy</a>.





<u>Unit 5 Reinforcement Learning (2</u>

Lecture 17. Reinforcement Learning

Course > weeks)

> <u>1</u>

> 8. Q-value Iteration

## 8. Q-value Iteration Q-value Iteration

and do exactly the same computation and directly get Qs.

And that's exactly what you will do in your exercise.

You will reformulate this algorithm and get a new algorithm which is called Q value iterations.

But the idea-- exactly the same.

And we will use this algorithm in our discussion

about reinforcement learning next time.

Thank you.





Video <u>Download video file</u> Transcripts

<u>Download SubRip (.srt) file</u>

<u>Download Text (.txt) file</u>

In this problem, we compute the optimal policy using Q-value iteration.

First, let's recall some of the bellman equations:

$$Q^{st}\left(s,a
ight) \;\;=\;\; \sum_{s'} T\left(s,a,s'
ight) \left(R\left(s,a,s'
ight) + \gamma V^{st}\left(s'
ight)
ight)$$

We also know that,  $V^{st}\left(s
ight)=\max_{a}Q^{st}\left(s,a
ight)$ . Therefore,

$$Q^{st}\left(s,a
ight) \;\;=\;\; \sum_{s^{\prime}}T\left(s,a,s^{\prime}
ight)\left(R\left(s,a,s^{\prime}
ight)+\gamma\mathrm{max}_{a^{\prime}}Q^{st}\left(s^{\prime},a^{\prime}
ight)
ight)$$

## Q value iteration update rule

1/1 point (graded)

Q-value iteration update rule from the above equation could be derived similar to the value iteration update rule.

Pick the most suitable Q-value iteration update rule from the options below:

$$egin{aligned} & igotimes Q_{k+1}^*\left(s,a
ight) = \sum_{s'} T\left(s,a,s'
ight) \left(R\left(s,a,s'
ight) + \gamma \mathrm{max}_{s'} Q_k^*\left(s',a
ight)
ight) \end{aligned}$$

$$igcup Q_{k+1}^{st}\left( s,a
ight) =\sum_{s^{\prime}}T\left( s,a,s^{\prime}
ight) \left( R\left( s,a,s^{\prime}
ight) +\gamma V^{st}\left( s^{\prime}
ight) 
ight)$$

$$ullet \ Q_{k+1}^*\left(s,a
ight) = \sum_{s'} T\left(s,a,s'
ight) \left(R\left(s,a,s'
ight) + \gamma \mathrm{max}_{a'} Q_k^*\left(s',a'
ight)
ight)$$

$$egin{aligned} & igotimes_{k+1}^*\left(s,a
ight) = \sum_{s'} T\left(s,a,s'
ight) \left(R\left(s,a,s'
ight) + \gamma Q_k^*\left(s',a
ight)
ight) \end{aligned}$$

## **Solution:**

First let's recall the following bellman equations:

$$Q^{st}\left(s,a
ight) \ = \ \sum_{s'} T\left(s,a,s'
ight) \left(R\left(s,a,s'
ight) + \gamma V^{st}\left(s'
ight)
ight)$$

We also know that,  $V^{st}\left(s
ight)=\max_{a}Q^{st}\left(s,a
ight)$ . Therefore,

$$Q^{st}\left(s,a
ight) \ = \ \sum_{s'} T\left(s,a,s'
ight) \left(R\left(s,a,s'
ight) + \gamma \mathrm{max}_{a'} Q^{st}\left(s',a'
ight)
ight)$$

Q value iteration would use previous iteration of the Q-value on the right hand side of the above equation to update the Q value estimate of the current step. Hence, the Q value update for  $k^{th}$  step would look like:

$$Q_{k+1}^{st}\left(s,a
ight) = \sum_{s^{\prime}} T\left(s,a,s^{\prime}
ight) \left(R\left(s,a,s^{\prime}
ight) + \gamma \mathrm{max}_{a^{\prime}} Q_{k}^{st}\left(s^{\prime},a^{\prime}
ight)
ight)$$

Submit

You have used 1 of 2 attempts

**1** Answers are displayed within the problem

<b>Topic:</b> Unit 5 Reinforcement Learning (2 weeks) :Lecture 17. Reinforcement Learning 1 / 8. Q-value Iteration	Hide Discussion
	Add a Post
Show all posts ▼	by recent activity ▼
There are no posts in this topic yet.	
×	

© All Rights Reserved