8. Q-value Iteration
Q-value Iteration

0:00 / 0:00 1.25x

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

The derivation of the Q-value iteration update rule from the equation above is similar to the derivation of the value iteration update rule.

First, recall the Bellman equations:

$$\begin{aligned} V^*(s) &= \max_a Q^*(s,a) \\ Q^*(s,a) &= \sum_{s'} T(s,a,s') \left(R(s,a,s') + \gamma V^*(s')\right). \end{aligned}$$

Plugging first equation into the second, we get:

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

Now, let $Q_k^*(s,a)$ be the expected rewards from state s followed by action a, and then acting optimally for k steps afterwards. (Hence, $V_k^*(s) = \max_a Q_k^*(s,a)$.)

Q-Value Iteration Update Rule

1/1 point (graded)

Referring to the equations above, what should the Q-value iteration update rule be?

- $Q_{k+1}^*(s,a) = \sum_{s'} T(s,a,s') \left(R(s,a,s') + \gamma \max_{s'} Q_k^*(s',a)\right)$
- $Q_{k+1}^*(s,a) = \sum_{s'} T(s,a,s') \left(R(s,a,s') + \gamma V^*(s')\right)$
- **(selected)** $Q_{k+1}^*(s,a) = \sum_{s'} T(s,a,s') \left(R(s,a,s') + \gamma \max_{a'} Q_k^*(s',a')\right)$
- $Q_{k+1}^*(s,a) = \sum_{s'} T(s,a,s') \left(R(s,a,s') + \gamma Q_k^*(s',a)\right)$

✔

Solution:

Q-value iteration would use the previous iteration of the Q-value on the right hand side of the equation

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

to update the Q value estimate of the current step. Hence, the Q value update for k^{th} step would look like:

$$Q^*_{k+1}(s,a) = \sum_{s'} T(s,a,s') \left(R(s,a,s') + \gamma \max_{a'} Q^*_k(s',a') \right).$$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 5 Reinforcement Learning (2 weeks) :Lecture 17. Reinforcement Learning 1 / 8. Q-value Iteration

Add a Post

Show all posts by recent activity

- why not? — 1
  Why not switch between calculating Q from V, then V from Q, until convergence? We've done ...
- Why do we sum up by s'? — 1
  So the transition is not deterministic? That is not like we have probabilities of taking action "u...
- What is a'? — 2
  What's the difference between a and a'?