Course › Unit 5 Reinforcement Learning (2 weeks) › Lecture 17. Reinforcement Learning 1 › 7. Value Iteration

7. Value Iteration

Value Iteration

marks

a Q star S A. What it says for a given state

is go select, check all the Q star S a's and select

the action which maximizes it.

So what eventually I go to after this algorithm?

After this algorithm, when it converged,

I computed the Q values, and then I computed the policy.

And now I know how to act in my MDP.

So that's what we have done here.

End of transcript. Skip to the start.

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Recall the value iteration update rule from the lecture:

$$V_{k+1}^{*}\left(s\right)=\max_{a}\left[\sum_{s'}T\left(s,a,s'\right)\left(R\left(s,a,s'\right)+\gamma V_{k}^{*}\left(s'\right)\right)\right]$$

Consider the following example discussed in the lecture:

Agent's starting state				$+1$

An agent is trying to navigate a one-dimensional grid consisting of 5 cells. At each step, the agent has only one action to choose from which will make it move to its adjoining cell on the right until it reaches the rightmost cell in which case it receives a reward of $+1$ and the agent comes to a halt.

Let V be an array that specifies the value function for the 5 states that the agent can be in. Let $V\left(i\right)$ denote the value function of the state i where the agent is in the i^{th} cell from left.
Let V_{k}^{*} denote the value function estimate at the k^{th} step of the value iteration algorithm. Let V_{0}^{*} denote the initialization of this estimate and assume that the discount factor is $\gamma=0.5$.

Recall that, from the lecture

$$\begin{array}{lcl} V_0^* & = & [0 \quad 0 \quad 0 \quad 0 \quad 0] \\ V_1^* & = & [0 \quad 0 \quad 0 \quad 0 \quad 1] \\ V_2^* & = & [0 \quad 0 \quad 0 \quad 0.5 \quad 1] \end{array}$$

Assume that all elements in V are initialized to 0.

Value Function Update

1/1 point (graded)

Run the 3^{rd} iteration of the value iteration algorithm to get V_3^* and answer the following questions:

Enter the value of V_3^* as an array $[s_0, s_1, \ldots]$

[0, 0, 0.25, 0.5, 1]

✓ **Answer:** [0, 0, 0.25, 0.5, 1]

Solution:

Note that a non-zero reward is obtained only in state s_4 when transitioning to s_5.

The 3^{rd} step of the value iteration could be worked out as follows:

$$\begin{aligned} V_3^*(s_1) &= 0 + \gamma * V_2^*(s_2) \\ V_3^*(s_1) &= 0 + 0.5 * 0 = 0 \end{aligned}$$

$$\begin{aligned} V_3^*(s_3) &= 0 + \gamma * V_2^*(s_4) \\ V_3^*(s_3) &= 0 + 0.5 * 0.5 = 0.25 \end{aligned}$$

$$\begin{aligned} V_3^*(s_4) &= 0 + \gamma * V_2^*(s_5) \\ V_3^*(s_4) &= 0 + 0.5 * 1 = 0.5 \end{aligned}$$

The same computation for the rest of the states.

Submit You have used 1 of 3 attempts

ℹ Answers are displayed within the problem

Number of steps till convergence

1/1 point (graded)

Enter below the number of steps it takes starting from V_0^* for the value function updates to converge to the optimal value function V^*:

5 ✓ **Answer:** 5

Solution:

Note that after the 5^{th} step, the reward from the rightmost cell in the grid gets propagated to the leftmost state after which the value function estimate V_k^* stops updating. Hence, for this example it takes 5 steps for the value function estimate to converge to the optimal value function.

Submit You have used 1 of 2 attempts

ℹ Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 5 Reinforcement Learning (2 weeks) :Lecture 17. Reinforcement Learning 1 / 7. Value Iteration

Add a Post

< All Posts

[Staff] Number of steps till convergence

question posted about 4 hours ago by **disguiser**

Should we count V_0 as a step? Should we count V_{k+1} step that has the same value as V_k?

This post is visible to everyone.

romfirst

about an hour ago

The number of steps is the smallest k such as $V_k^* = V^*$.

Add a comment

Showing all responses

Preview

Submit