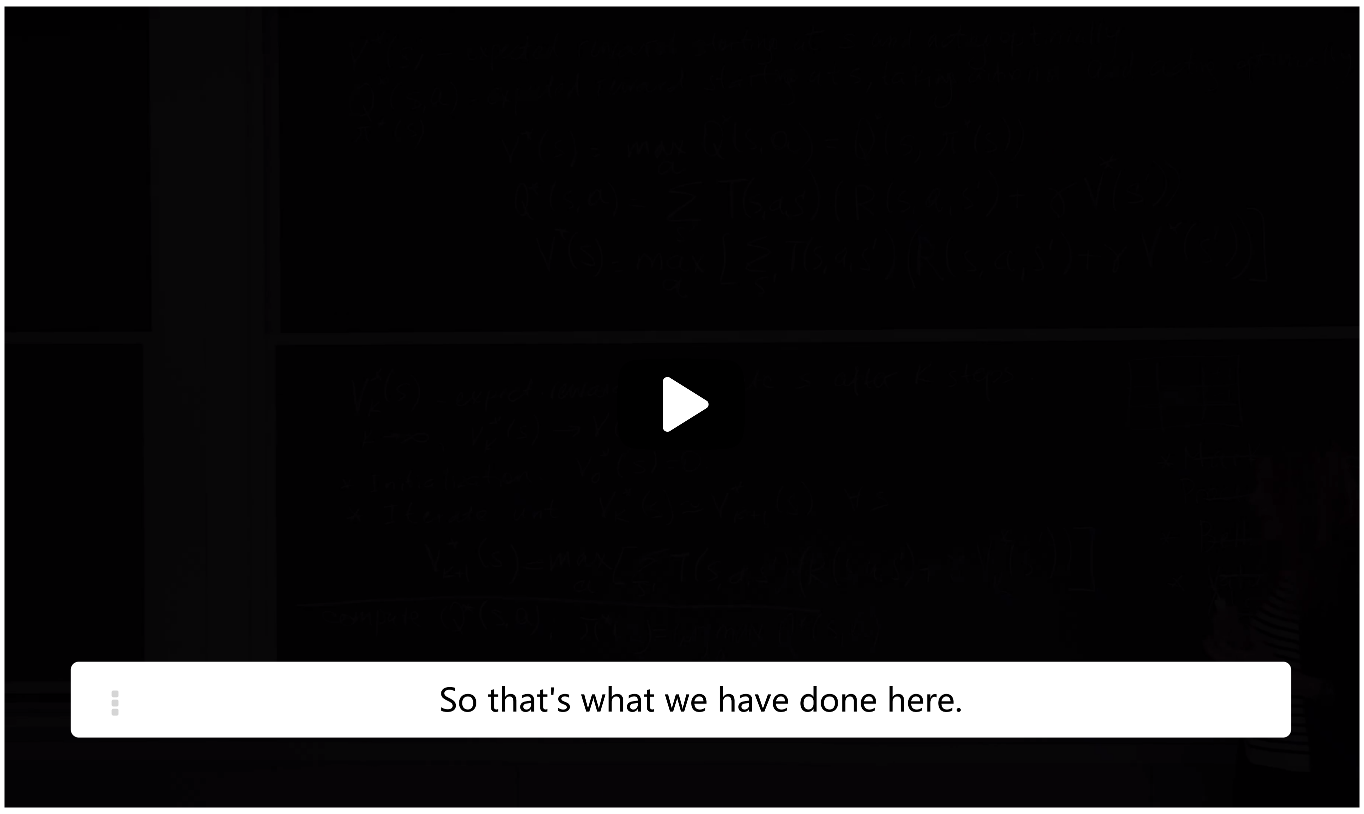
7. Value Iteration

Value Iteration



17:59 / 17:59 1.0x

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

xuetangX.com 学堂在线

Recall from lecture the **value iteration update rule** :

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

where $V_k^*\left(s\right)$ is the expected reward from state s after acting optimally for k steps.

Recall the 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, i.e. it moves to the cell on the immediate right. When it reaches the rightmost cell, it receives a reward of $+1$ and comes to a halt.

Let $V^*\left(i\right)$ denote the value function of state i, the i^{th} cell starting from left.

Let $V_k^*\left(i\right)$ denote the value function estimate at state i at the k^{th} step of the value iteration algorithm. Let $V_0^*\left(i\right)$ denote the initialization of this estimate.

Use the discount factor $\gamma = 0.5$.

We will write the functions V_k^* as arrays below, i.e. as $\begin{bmatrix} V_k^*(1) & V_k^*(2) & V_k^*(3) & V_k^*(4) & V_k^*(5) \end{bmatrix}$.

Initialize by setting $V_0^*(i) = 0$ for all i:

$$V_0^* = [0 \quad 0 \quad 0 \quad 0 \quad 0].$$

Then, using the value iteration update rule, we get

$$\begin{aligned} V_1^* &= [0 \quad 0 \quad 0 \quad 0 \quad 1], \\ V_2^* &= [0 \quad 0 \quad 0 \quad 0.5 \quad 1] \end{aligned}$$

Note (Aug 22): Note that as soon as the agent takes the first action to reach cell 5, it halts and does not take any more action, so we set $V_{k+1}^*(5) = V_k^*(5)$ for all $k \geq 1$.

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 $\begin{bmatrix} V_3^*(0) & V_3^*(1) & V_3^*(2) & V_3^*(3) & V_3^*(4) \end{bmatrix}$.

(For example, type `[0,2,0,3,4]` for the array $\begin{bmatrix} 0 & 2 & 0 & 3 & 4 \end{bmatrix}$.)

[0,0,1/4,1/2,1] ✔ **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:

$$\begin{aligned} V_3^*(1) &= 0 + \gamma * V_2^*(2) \\ V_3^*(1) &= 0 + 0.5 * 0 = 0 \end{aligned}$$

$$\begin{aligned} V_3^*(2) &= 0 + \gamma * V_2^*(3) \\ V_3^*(2) &= 0 + 0.5 * 0 = 0 \end{aligned}$$

$$\begin{aligned} V_3^*(3) &= 0 + \gamma * V_2^*(4) \\ V_3^*(3) &= 0 + 0.5 * 0.5 = 0.25 \end{aligned}$$

$$\begin{aligned} V_3^*(4) &= 0 + \gamma * V_2^*(5) \\ V_3^*(4) &= 0 + 0.5 * 1 = 0.5 \end{aligned}$$

and $V_3^*(5) = V_2^*(5) = 1$.

The same computation for the rest of the states.

Submit

You have used 1 of 3 attempts

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

Answers are displayed within the problem

Discussion

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

Show Discussion