7. Value Iteration

Value Iteration

0:00 / 0:00 1.25x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

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

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

where $V_k^*(s)$ 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.

Note: The reward function is defined to be $R(s,a,s') = R(s)$, $R(s=5) = 1$ and $R(s) = 0$ otherwise. Note that we get the reward when we are leaving from the current state. When it reaches the rightmost cell, it stays for one more time step and then receives a reward of $+1$ and comes to a halt.

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

Let $V_k^*(i)$ denote the value function estimate at state i at the k^{th} step of the value iteration algorithm. Let $V_0^*(i)$ 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 $\left[V_k^*(1) \quad V_k^*(2) \quad V_k^*(3) \quad V_k^*(4) \quad V_k^*(5) \right]$.

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

$$V_0^* \quad = \quad \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then, using the value iteration update rule, we get

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

Note: Note that as soon as the agent takes the first action to reach cell 5, it stays for one more step and 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

1.0/1 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 $\begin{bmatrix} V_3^*(1) & V_3^*(2) & V_3^*(3) & V_3^*(4) & V_3^*(5) \end{bmatrix}$.

(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:

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

$$V_3^*(2) = 0 + \gamma * V_2^*(3)$$
$$V_3^*(2) = 0 + 0.5 * 0 = 0$$

$$V_3^*(3) = 0 + \gamma * V_2^*(4)$$
$$V_3^*(3) = 0 + 0.5 * 0.5 = 0.25$$

$$V_3^*(4) = 0 + \gamma * V_2^*(5)$$
$$V_3^*(4) = 0 + 0.5 * 1 = 0.5$$

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 to 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

Complexity of Value Iteration

1/1 point (graded)

Let the number of states and actions be $|S|$ and $|A|$, respectively. Choose from the following the **complexity of an iteration** of the value iteration algorithm.

- ( ) $\mathcal{O}\left(|S|^3 \cdot |A|\right)$
- ( ) $\mathcal{O}\left(|S| \cdot |A|\right)$
- (•) $\mathcal{O}\left(|S|^2 \cdot |A|\right)$ ✔

Solution:

We update the expected reward for each state in every iteration – there are $|S|$ states. For each state, we investigate a maximum of $|A|$ possible actions and for each such action there are $|S|$ possible transitions at the most. Therefore, the complexity of an iteration is $\mathcal{O}\left(|S|^2 \cdot |A|\right)$.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Another Example of Value Iteration (Software Implementation)

3 points possible (graded)

Consider the same one-dimensional grid with reward values as in the first few problems in this vertical. However, consider the following change to the transition probabilities: At any given grid location the agent can choose to either stay at the location or move to an adjacent grid location. If the agent chooses to stay at the location, such an action is successful with probability $1/2$ and

- if the agent is at the leftmost or rightmost grid location it ends up at its neighboring grid location with probability $1/2$,
- if the agent is at any of the inner grid locations it has a probability $1/4$ each of ending up at either of the neighboring locations.

If the agent chooses to move (either left or right) at any of the inner grid locations, such an action is successful with probability $1/3$ and with probability $2/3$ it fails to move, and

- if the agent chooses to move left at the leftmost grid location, then the action ends up exactly the same as choosing to stay, i.e., staying at the leftmost grid location with probability $1/2$, and ends up at its neighboring grid location with probability $1/2$,
- if the agent chooses to move right at the rightmost grid location, then the action ends up exactly the same as choosing to stay, i.e., staying at the rightmost grid location with probability $1/2$, and ends up at its neighboring grid location with probability $1/2$.

Let $\gamma = 0.5$.

Run the value iteration algorithm for 100 iterations. Use any computational software of your choice.

Enter the value of V^*_{100} as an array
$\begin{bmatrix} V^*_{100}(1) & V^*_{100}(2) & V^*_{100}(3) & V^*_{100}(4) & V^*_{100}(5) \end{bmatrix}$.

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

Answer: [0.016667,0.05,0.2,0.8,1.2]

Are the values different if we iterate 200 times? Consider only the first three decimal digits to answer this question.

- Yes
- No ✔

How about if we only performed 10 iterations? Are the values different when compared to 100 iterations? Consider only the first three decimal digits to answer this question.

- Yes ✔
- No

Solution:

After 100 iterations,

$$V_{100}^{*} = [0.016667, 0.05, 0.2, 0.8, 1.2]\,.$$

If we run the algorithm for another 100 iterations, we see that the values do not change in the first three decimal digits.

However, at only 10 iterations,

$$V_{10}^{*} = [0.015886, 0.049121, 0.199040, 0.799023, 1.199023]\,.$$

Submit You have used 0 of 4 attempts

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

Show all posts by recent activity

[Tips from Staff:] Some clarification for the reward function and transition probability. 34

Pinned Staff

Tips for Software implementation question 30

Pinned

How this class of problem extends to continuous cases (still in discrete time) ? 3
Here we have a limited set of states and actions, while in many real problems we have both c...

Community TA

[staff] 2
The page showed the software implementation problem, but when I reloaded the page beca...

[STAFF] Last question not displayed 5
I submitted by second attempt few minutes back and now I see the following message instea...

[SOLUTION] Excel script with solution to this problem 31

"When it reaches the rightmost cell, it receives a reward of +1 and comes to a halt. " 4
It is this statement that should be corrected. What is described in the lecture, and in the prob...

[SOLUTION] Python 11
Following on from riccardo_riccobello's kind posting of his Excel solution, I thought it would b...

[STAFF] Another Example of Value Iteration - provide the correct code
Hi Staff, I've solved this problem four times. One of these times, I was using [pymdptoolbox... 13

[Staff] Q. Number of Steps to Convergence 2

My understanding about Value Iteration Algorigh (based in problem "Software Implementation")
I was many time trying to understand VIA and how it works to solve the questions, but somet... 1