Exercises due May 4, 2021 19:59 EDT

Value Iteration

17:59 / 17:59 1.50x

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

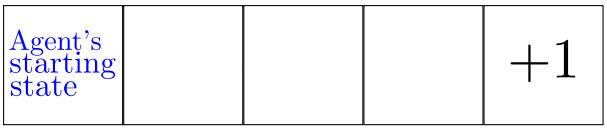
Download SubRip (.srt) file
Download Text (.txt) file

Recall 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.



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:

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

Then, using the value iteration update rule, we get

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

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

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 point possible (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 $\begin{bmatrix} 0 & 2 & 0 & 3 & 4 \end{bmatrix}$.)

Submit You have used 0 of 3 attempts

Number of Steps to Convergence

1 point possible (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^*:

Submit You have used 0 of 2 attempts

Complexity of Value Iteration

1 point possible (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.

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

Submit You have used 0 of 2 attempts

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 $\left[V_{100}^*(1) \quad V_{100}^*(2) \quad V_{100}^*(3) \quad V_{100}^*(4) \quad V_{100}^*(5) \right]$.

(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.)

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

Submit You have used 0 of 4 attempts

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

Why does my code give the wrong answer?	I have used the code below but it gave me the wrong answer. And I cannot figure out why. Can ...	6
Are rewards earned when transitioning back to the same state?	For the last question on this page, are rewards R(s) for a state earned even when transitioning b...	4
Notation question		1
about the transition probabilities		7
[TO STAFF]: is the solution for "Software Implementation" correct?	I think reward should be R(s') not R(s). Which of the following equations is correct? 1. Q(s,a)=∑T(...	5
Drawing helps to understand		1
Another Example (Value Iteration)	Dear Staff team, Kindly note that I have implemented the algorithm of value iteration, and I actu...	5
How is this a deterministic model?	Take actions R → R → S → L → S. You can tell me the probability that the agent is now in _state ...	3
Complete Algorithm	Am I correct in that the full algorithm after initialization is below. In other words, we need to itera...	2
[Staff] [URGENT]: Technical issue--Another example of value iteration question disappeared		3

? Are the transition probabilities fully specified? 9
I don't see where the transitions probabilities are specified if in the leftmost position and the ch...

? the prob of left most move to right? 3
what's the successful probability the agent at left most cell choose to move right? 1 or 1/2?

☑ Number of Steps to Convergence 5
The number of steps does it include the zeroth step where we initialize everything to zero? I thin...

Another Example of Value Iteration, V*(s=5) larger than 1?