7. Value Iteration

Exercises due May 3, 2023 08:59 -03 Past due

Value Iteration

0:00 / 0:00 1.0x

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				$+1$

Then, using the value iteration update rule, we get

Note: Note that as soon as the agent takes the first action to reach cell , it stays for o
and does not take any more action, so we set for all .

Value Function Update

1.0/1 point (graded)
Run the iteration of the value iteration algorithm to get and answer the followin

Enter the value of as an array .

(For example, type `[0,2,0,3,4]` for the array .)

[0, 0, 0.25, 0.5, 1]

Submit You have used 3 of 3 attempts

Number of Steps to Convergence

1/1 point (graded)
Enter below the number of steps it takes starting from for the value function updat
optimal value function :

5

Submit You have used 2 of 2 attempts

Complexity of Value Iteration

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 problem
However, consider the following change to the transition probabilities: At any given gric
choose to either stay at the location or move to an adjacent grid location. If the agent c
location, such an action is successful with probability and

- if the agent is at the leftmost or rightmost grid location it ends up at its neighboring probability ,
- if the agent is at any of the inner grid locations it has a probability each of endi neighboring locations.

If the agent chooses to move (either left or right) at any of the inner grid locations, suc with probability and with probability it fails to move, and

- if the agent chooses to move left at the leftmost grid location, then the action ends choosing to stay, i.e., staying at the leftmost grid location with probability , and neighboring grid location with probability ,
- if the agent chooses to move right at the rightmost grid location, then the action en as choosing to stay, i.e., staying at the rightmost grid location with probability , neighboring grid location with probability .

Note in this setting, we assume that the game does not halt after reaching the rightmo

Let .

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

Enter the value of as an array

(For example, type `[0,2,0,3,4]` for the array . Type at least 4 decima

Are the values different if we iterate 200 times? Consider only the first three decimal di question.

- Yes
- No

Discussion

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

< Previous

Next >

Show all posts