

> 2

Unit 5 Reinforcement Learning (2

Lecture 18. Reinforcement Learning

<u>Course</u> > <u>weeks</u>)

1. Revisiting MDP Fundamentals

Audit Access Expires May 11, 2020

You lose all access to this course, including your progress, on May 11, 2020.

1. Revisiting MDP Fundamentals Revisiting MDP Fundamentals



Video

Download video file

Transcripts

<u>Download SubRip (.srt) file</u> <u>Download Text (.txt) file</u>

Review: Markovian Assumption

1/1 point (graded)

Vhich of the following are true about Markov	v decision processes? (Choose all that apply.)
The transition probability of reaching a on $oldsymbol{s}$ and all the states visited before $oldsymbol{s}$	state s^\prime from a given state s would depend both
The transition probability of reaching a on s and is independent of the states vi	state s^\prime from a given state s would only depend isited before state s
The rewards received starting from stat visited before $oldsymbol{s}$	se s would depend both on s and all the states
The rewards received starting from stat independent of the states that were visit	

Solution:

1

Recall from the previous lecture that under Markovian assumptions, both the transition probability of reaching a state s' from a given state s, and the rewards received starting from state s, would only depend on s and is independent of the states visited before state s. This assumption allows us to specify the transition probabilities and rewards by $T\left(s,a,s'\right)$ and $R\left(s,a,s'\right)$.

Submit

You have used 1 of 2 attempts

• Answers are displayed within the problem

Policy Function and Value Function

1/1 point (graded)

From the following options select one or more statement(s) which are true about the optimal policy function π^* , the optimal value function V^* and the optimal Q-function Q^*

- $ightharpoons \pi^st\left(s
 ight)$ records the action that would lead to the best expected utility starting from the state s
- $\pi^{st}\left(s
 ight)$ records the action that would necessarily lead to the best immediate reward for the current step
- $abla V^{st}\left(s
 ight) = \max_{a} Q^{st}\left(s,a
 ight)$ holds for all states s
- $rac{1}{N}V^{st}\left(s
 ight) =max_{a}\left[\sum_{s}T\left(s,a,s^{\prime}
 ight) \left(R\left(s,a,s^{\prime}
 ight) +\gamma V^{st}\left(s^{\prime}
 ight)
 ight)
 ight]$ must hold true for the optimal value function when $0 < \gamma < 1$



Solution:

The goal of the optimal policy function is to maximize the expected discounted reward, even if this means taking actions that would lead to lower immediate next-step rewards from few states.

Recall that from the previous lecture that for all s, the (optimal) value function is

$$egin{aligned} V^{*}\left(s
ight) &=& \max_{a}Q^{*}\left(s,a
ight) \ &=& \max_{a}\left[\sum_{s^{\prime}}T\left(s,a,s^{\prime}
ight)\left(R\left(s,a,s^{\prime}
ight)+\gamma V^{*}\left(s^{\prime}
ight)
ight)
ight] \qquad ext{where }0\leq\gamma<1. \end{aligned}$$

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

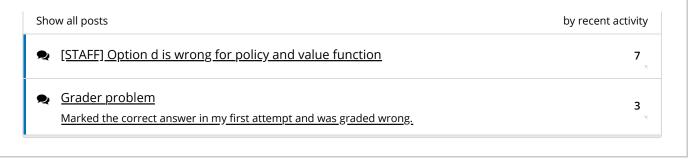
Hide Discussion

Topic: Unit 5 Reinforcement Learning (2 weeks): Lecture 18. Reinforcement Learning 2 / 1. Revisiting MDP Fundamentals

Add a Post

1. Revisiting MDP Fundamentals | Lecture 18. Re...

https://courses.edx.org/courses/course-v1:MITx+...



© All Rights Reserved

4 of 4 2020-05-09, 9:49 a.m.