1. Revisiting MDP Fundamentals

Revisiting MDP Fundamentals

0:00 / 0:00 1.25x

Video

Download video file

Transcripts

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

Review: Markovian Assumption

1/1 point (graded)

Which of the following are true about Markov decision processes? (Choose all that apply.)

- [ ] The transition probability of reaching a state s' from a given state s would depend both on s and all the states visited before s
- [x] The transition probability of reaching a state s' from a given state s would only depend on s and is independent of the states visited before state s
- [ ] The rewards received starting from state s would depend both on s and all the states visited before s
- [x] The rewards received starting from state s would depend only on s and are independent of the states that were visited before s.

Solution:

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^*

- [x] $\pi^*(s)$ records the action that would lead to the best expected utility starting from the state s
- [ ] $\pi^*(s)$ records the action that would necessarily lead to the best immediate reward for the current step
- [x] $V^*(s) = \max_a Q^*(s,a)$ holds for all states s
- [x] $V^*(s) = max_a [\sum_{s'} T(s,a,s')(R(s,a,s') + \gamma V^*(s'))]$ 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

$$\begin{aligned} V^*(s) &= \max_a Q^*(s,a) \\ &= \max_a \left[\sum_{s'} T(s,a,s')(R(s,a,s') + \gamma V^*(s'))\right] \qquad \text{where } 0 \leq \gamma < 1. \end{aligned}$$

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

Show all posts by recent activity

[STAFF] Option d is wrong for policy and value function 7

Grader problem
Marked the correct answer in my first attempt and was graded wrong. 3