

EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the [Privacy Policy](#). ×

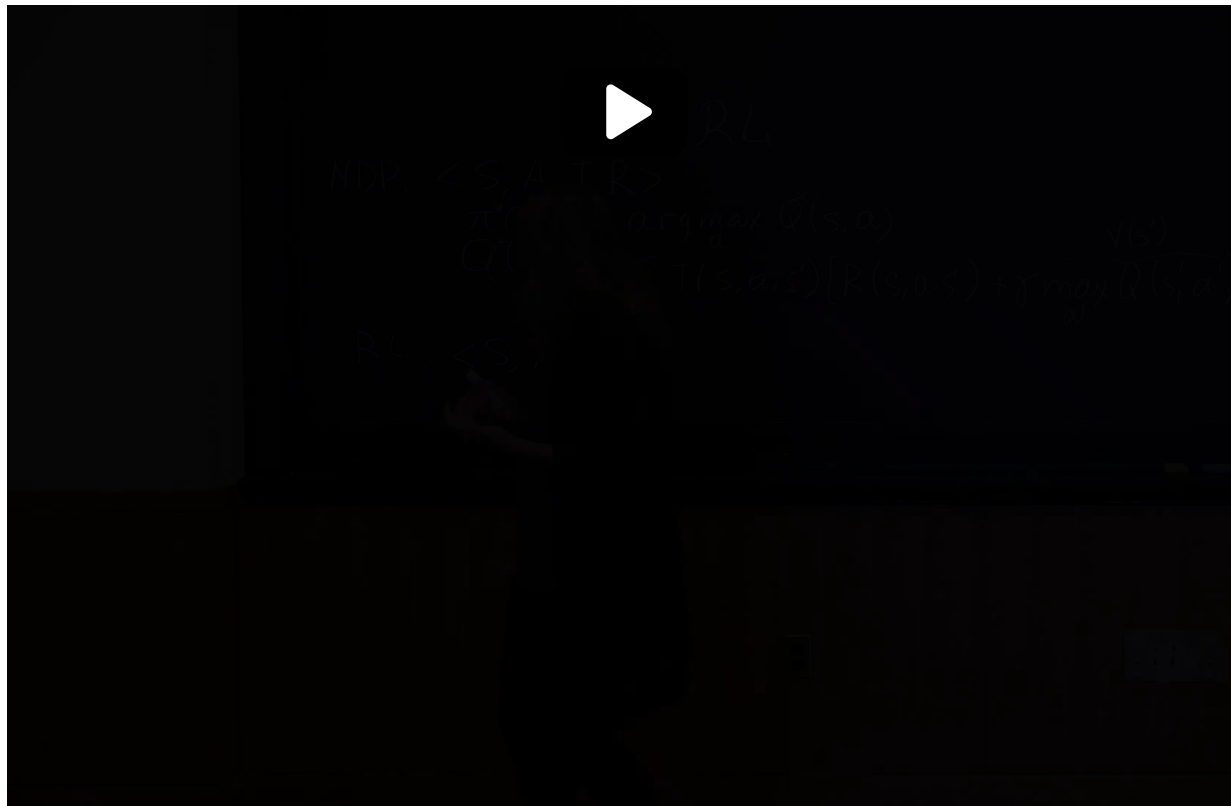


[Unit 5 Reinforcement Learning \(2 weeks\)](#) > [Lecture 18. Reinforcement Learning](#) > 1. Revisiting MDP Fundamentals

## 1. Revisiting MDP Fundamentals

### Revisiting MDP Fundamentals

eventually you  
need to provide us with the policy.  
So how are we doing these trials?  
And which and what to try is an integral  
part  
of the computation.  
So what we will do today, we will  
talk about how to do Q value iteration  
algorithm in this new setup where  
**experimenting in the world will be  
part of the computation.**



[End of transcript. Skip to the start.](#)



## Video

[Download video file](#)

## Transcripts

[Download SubRip \(.srt\) file](#)

[Download Text \(.txt\) file](#)

## Markovian Assumption

1/1 point (graded)

From the following options select one or more statement(s) which are true about markov decision processes:

- ☐ The transition probability of reaching a state  $s'$  from a given state  $s$  would depend both on  $s$  and all the states visited before  $s$
- ☒ 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  $s$
- ☒ 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, the following hold true:

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 only on  $s$  and are independent of the states that were visited before  $s$ .

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  $\pi^*$ , the optimal value function  $V^*$  and the optimal  $Q$ -function  $Q^*$

☒  $\pi^*(s)$  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 reward for the current step

☒  $V^*(s) = \max_a Q^*(s, a)$  holds for all states  $s$  ✓

☒  $V^*(s) = \max_a [\sum_{s'} T(s, a, s') (R(s, a, s') + \gamma V^*(s'))]$  must hold true for the optimal value function when  $\gamma < 1$  ✓



### 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,

$$V^*(s) = \max_a Q^*(s, a)$$

holds true for any state  $s$  and,

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

must hold true for the optimal value function when  $\gamma < 1$

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] Markov property and imprecise wording.

1

