



MITx 6.86x

Machine Learning with Python-From Linear Models to Deep Learning[Course](#)[Progress](#)[Dates](#)[Discussion](#)[Resources](#)[🏠 Course](#) / [Unit 5. Reinforcement Learning \(2 we...](#) / [Lecture 17. Reinforcement L](#)[◀ Previous](#)

4. Utility Function

[🔖 Bookmark this page](#)

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

Video note: In the video below at 1:25, Prof Barzilay miswrote "final horizon" on the board. It should be meant **finite horizon**.


Utility Function




Video

 [Download video file](#)

Transcripts

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

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

Finite Horizon vs Discounted Reward

1/1 point (graded)

The main problem for MDPs is to optimize the agent's behavior. To do so, we first need to define what we are trying to maximize in terms of accumulated rewards. We will define a **utility function** as the expectation of the sum of discounted rewards.

We consider two different types of utility functions:

1. **Finite horizon based utility** : The utility function is the sum of rewards after acting for a fixed number of steps. For example, in the case when the rewards depend only on the states, the utility function is the sum of rewards over the horizon.

How do these two types of utility function depend on the time steps?
(Choose all that apply.)



The action at state that maximizes a finite horizon based utility can depend on the actions that have been taken.



The action at state that maximizes a finite horizon based utility does **not** depend on the actions that have been taken.



The action at state that maximizes a discount reward based utility does **not** depend on the actions that have been taken.



The action at state that maximizes a discount reward based utility can depend on the actions that have been taken.



Submit

You have used 1 of 3 attempts

Discounted Utility

[< Previous](#)

[Next >](#)

1/1 point (graded)

Recall that the discounted reward in the case when

is given by:



edX

[About](#)

[Affiliates](#)

[edX for Business](#)

[Open edX](#)

[Careers](#)

[News](#)

Legal

[Terms of Service & Honor Code](#)

[Privacy Policy](#)

💬 Utility Function

💬 R(s0) expression

What does Gamma represent in this expression? Why is the reward bounded at $R_{\max} / (1-\gamma)$?



© 2023 edX LLC. All rights reserved.

深圳市恒宇博科技有限公司 [粤ICP备17044299号-2](#)