



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 18. Reinforcement L](#)[< Previous](#)

3. Q value iteration by sampling

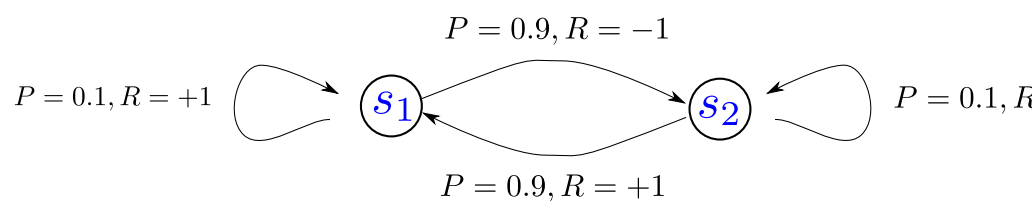
[Bookmark this page](#)

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

Q value iteration by sampling**Video**
[Download video file](#)
Transcripts
[Download SubRip \(.srt\) file](#)
[Download Text \(.txt\) file](#)

Let us consider a toy example which might not be very realistic but which nevertheless illustrates the Q-value iteration for RL using sampling approach.

For this example, assume that there are only two states, s_1 , s_2 and only one action possible in each state. Let a_{s_1} , a_{s_2} be the actions that could be taken from s_1 and s_2 respectively.



The state transition probabilities are listed below and are also shown in the figure above.

$$T(s_1, a_{s_1}, s_1) = 0.1$$

exact functions. However, for this toy example we will assume that the Q-value is directly provided with the above specified values of and has to resort to sampling function.

Let's say that the agent starts out from state and collects few samples. Each sample the following tuple which indicates that the agent received a reward when it reached state by taking action from the state .

The collected samples are described as follows in the order in which they are presented in the iteration algorithm.

Let be used to denote the sample of (). Then recall that

For all of the following problems, assume that the discount factor , and the Q-values are initialized to to start with. That is,








Numerical Example

1/1 point (graded)

Enter below the value of after the first sample is processed by the Q-value

[< Previous](#)[Next >](#)[Submit](#)

You have used 1 of 3 attempts

show all posts	
	<u>What is the means of $Q(s,a)$?</u> <u>how to compute $Q(s,a)$?</u>
	<u>This is the Q-Learning Algorithm right?</u> <u>This is the Q-Learning Algorithm right?</u>
	<u>How is alpha determined?</u> <u>Do we use cross-validation?</u>
	<u>Regarding s_2</u>
	<u>How do we know reward for each sample $R(s, a, s')$?</u> <u>From lecture, we know that RL is different from MDP since we only know or define . We don't know T and R. I</u>
	<u>hint</u> <u>Be careful which Q are used to calculate S and the remaining part...</u>
	<u>Estimation of maxQ for a given sample</u> <u>Hi all, Can somebody explain to me how to obtain the estimate of maxQ for a given sample? Kind regards, To</u>

Legal

[Terms of Service & Honor Code](#)

[Privacy Policy](#)

[Accessibility Policy](#)

[Trademark Policy](#)

[Sitemap](#)

[Cookie Policy](#)

[Do Not Sell My Personal Information](#)

Connect

[Blog](#)

[Contact Us](#)

[Help Center](#)

[Security](#)

[Media Kit](#)

