



Lecture 04: Temporal Difference Methods

#cs #rl

83 kez oynandı • 136 oyuncu









Herkese açık bir kahoot

Sorular (8)

1 - Quiz

MC vs TD: What is correct?









60 sn

-  MC is more sensitive to initial values than TD. 
-  TD has a lower variance than MC. 
-  MC is usually more efficient than TD. 
-  MC has a larger bias. 

2 - Quiz

Bootstrapping vs Sampling: What is correct?









60 sn

-  MC bootstraps and TD samples 
-  MC and TD do not bootstrap 
-  MC samples, but TD does not. 
-  MC does not bootstrap and TD samples. 

3 - Quiz

Temporal difference learning is biased because ...

60 sn

-  the update uses the expected return of the next state. 
-  the update uses the estimated return of the next state. 
-  it bootstraps its estimation of the immediate reward. 
-  it has a lower variance. 

4 - Doğru/Yanlış

In contrast to MC Control, SARSA can be easily used in an off-policy setting.

60 sn

-  True ✗
-  False ✓

5 - Quiz

What is the correct SARSA equation?





60 sn

-  1 ✗
-  2 ✗
-  3 ✗
-  4 ✓

6 - Quiz

Off-policy Learning...





60 sn

-  is more sample-efficient, but can be less stable. ✓
-  always needs some off-policy correction like Imp. Sam. ✗
-  can reuse experience generated from old policies. ✓
-  can only learn from older policies of the same agent. ✗

7 - Quiz

Q-Learning: What is correct?

60 sn

-  A_t is sampled accordingly to target policy. ✗
-  A_t is sampled accordingly to behavior policy. ✓
-  A' is sampled accordingly to target policy. ✓
-  A' is sampled accordingly to behavior policy. ✗

8 - Quiz

How does Double Q-learning help?

60 sn

It decorrelates the noise of Q and $\arg\max_a Q$.

It is always underestimating the true Q-value.



By the estimation of two Q-functions, the noise adds to 0.



It is always overestimating the true Q-value.

