



Lecture 10: Advanced Value-based Methods

70 plays · 99 players





 A public kahoot

Questions (5)

1 - Quiz

PER samples transitions with probability relative to their importance on the basis of...

60 sec

-  their Q-value. ✗
-  the entropy of the induced Boltzmann policy. ✗
-  their immediate reward. ✗
-  their TD-error. ✓

2 - True or false

In contrast to vanilla Q-learning, distributional Q-learning can model multi-modalities in value.





60 sec

-  True ✓
-  False ✗

3 - Quiz

DDPG: What is true?


60 sec

-  The actor yields the true $\arg\max_a Q$. ✗
-  DDPG updates the actor on the MSE. ✗
-  DDPG is an on-policy algorithm. ✗
-  The actor yields an approximation of $\arg\max_a Q$. ✓

4 - Quiz

What is not part of TD3?





60 sec

- | | | |
|--|---------------------------|---|
|  | Target-policy smoothing | ✗ |
|  | Clipped Double Q-learning | ✗ |
|  | Entropy regularization | ✓ |
|  | Delayed policy update | ✗ |

5 - Quiz

SAC...

60 sec

- | | | |
|--|---|---|
|  | penalizes the entropy of the policy. | ✗ |
|  | augments the reward by the entropy of the policy. | ✓ |
|  | enforces the policy to be narrow. | ✗ |
|  | limits changes in the policy between updates. | ✗ |