



Lecture 09: Policy Gradient Methods

78 plays · 108 players









 A public kahoot

Questions (6)

1 - Quiz

In Policy Gradient Methods, we...

60 sec

-  always parameterize value-function and policy. 
-  implicitly learn the policy by a value-function. 
-  explicitly parameterize the policy. 
-  never use a value-function. 

2 - True or false

We can only get policy gradients for continuous policies.

60 sec

-  True 
-  False 

3 - Quiz

We can employ a baseline...









60 sec

-  to reduce variance. 
-  to reduce bias. 

4 - Quiz

What is true about baselines?







60 sec

-  Baselines change the expectation of the policy gradient. 
-  Baselines must not depend on actions. 
-  The baseline can be a constant scalar. 
-  Only the true value function can act as a baseline. 

5 - Quiz

Why do we introduce the surrogate in PPO?








60 sec

-  We can't sample states from a policy we haven't applied. 
-  To make more cautious updates. 
-  π_{old} yields better trajectories than π . 

6 - Quiz

Why do we constrain the surrogate (KL-divergence, clipping)?

60 sec

-  The optimization becomes infeasible otherwise. 
-  To make more cautious updates. 
-  The surrogate is only a local approximation of ξ . 
-  The surrogate is an overestimation of ξ . 