EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the <u>Privacy Policy</u>.

×



<u>Course</u> > <u>Policy</u>... > <u>Knowle</u>... > Knowle...

# **Knowledge Checks**

# Question 1

1/1 point (graded)

Which two of the following are advantages of policy gradient methods over valuefunction based methods??

- Policy gradient methods are scalable to problems with high dimensions or continuous state spaces.
- ✓ Policy gradient methods can learn stochastic policies.
- ☐ Policy gradient methods converge to the global optimum policy.
- ☐ Policy gradient methods are more sample efficient.



Submit

You have used 1 of 2 attempts

## Question 2

1/1 point (graded)

Which reinforcement learning methods does Actor-Critic algorithms combine??

O Policy gradient algorithms as critics and policy iteration algorithms as actors.

eturns as actors and policy interaction algorithms as critics.
nt algorithms as actors and expected value functions as critics.
have used 1 of 2 attempts
lihood ratio method has which two of the following policies?
e gradient decreases the likelihood of following trajectories with
e gradient increases the likelihood of finding trajectories with high
e gradient decreases the likelihood of following trajectories with
e gradient decreases the likelihood of finding trajectories with low eward.
have used 1 of 2 attempts

1/1 point (graded)

Which of the following are properties of the Reinforce algorithm?

Ouses a policy $oldsymbol{\pi}\left(s_{t} ight)$ during an episode to collect information on states, actions and rewards.
Computes the return for each episode using the rewards collected.
O Updates the model parameters in the director of the policy gradient.
○ All of the above. ✔
Submit You have used 1 of 2 attempts
Question 5  1/1 point (graded)  Which two of the following are methods to reduce the variance of the REINFORCE algorithm?
Use the minimum variance policy gradient to minimize variance of the return.
☑ Discount returns to encourage trajectories with good actions and discourage trajectories with bad actions.
✓ Using the discounted expected returns given the policy as a baseline discourages trajectories with return below the baseline.
<ul> <li>Using the expected returns given the policy as a baseline discourages trajectories with return away from the baseline.</li> </ul>
Submit You have used 1 of 2 attempts

#### Question 6

1/1 point (graded)

Which of the following is a correct definition of the advantage function?

<ul> <li>The difference between the gradient of the log likelihood and the state value function.</li> </ul>
The difference between the Q-value and the gradient of the log likelihood.

The difference between the Q-value and the state value function.
--

_			_					
	Tha	difforance	hotwoon	tha O	Value	and tha	discounted	roturn
	1116	unierence	nermeen	uie Q	-vaiue	and the	uiscounteu	return.

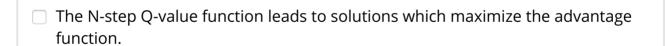
Submit

You have used 1 of 2 attempts

## Question 7

1/1 point (graded)

Which two of the following are the following are advantages of using an N-step Q-value function in an actor-critic algorithm?



- ✓ The N-step Q-value function bootstraps and does not need to sample to the end of an episode to compute an estimate of Q.
- ✓ The N-step Q-value function trades off bias for lower variance.
- ☐ The N-step Q-value function trades off variance for lower bias.



Submit

You have used 1 of 2 attempts

# Question 8

1/1 point (graded)

Which two of the following are advantages of the Asynchronous Advantage Actor-Critic (A3C) algorithm when compared to other actor-critic methods?

✓	Shares parameters between the actor and critic networks to improve data
	efficiency or speed of training.

Eliminates shared parameters between actor and critic networks to improve	/e
data efficiency or speed of training.	

Trains multiple policies on copies of the environment simultaneously improving
convergence.

Trains a single policy by acting on and collecting experience from parallel environments simultaneously to improve scalability.



Submit

You have used 1 of 2 attempts

© All Rights Reserved