



<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??

- Policy gradient algorithms as critics and policy iteration algorithms as actors.
- ullet Policy gradient algorithms as actors and policy iteration algorithms as critics.  $\checkmark$
- Discounted returns as actors and policy interaction algorithms as critics.

O Policy gradient algorithms as actors and expected value functions as critics.

Submit

You have used 1 of 2 attempts

## Question 3

1/1 point (graded)

Intuitively, the likelihood ratio method has which two of the following policies?

- Following the gradient decreases the likelihood of following trajectories with high variance.
- Following the gradient increases the likelihood of finding trajectories with high reward.
- Following the gradient decreases the likelihood of following trajectories with high bias.
- Following the gradient decreases the likelihood of finding trajectories with low or negative reward.



Submit

You have used 1 of 2 attempts

## Question 4

1/1 point (graded)

Which of the following are properties of the Reinforce algorithm?

- Uses a policy  $\pi(s_t)$  during an episode to collect information on states, actions and rewards.
- Ocomputes the return for each episode using the rewards collected.

 Updates the model parameters in the director of the policy gradient. All of the above. You have used 1 of 2 attempts Submit 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. Using the expected returns given the policy as a baseline discourages trajectories with return away from the baseline. You have used 1 of 2 attempts Submit Question 6 1/1 point (graded)

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

function.

The difference between the gradient of the log likelihood and the state value

• The difference between the Q-value and the gradient of the log likelihood. The difference between the Q-value and the state value function. The difference between the Q-value and the discounted return. You have used 2 of 2 attempts Submit 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 leads to solutions which maximize the advantage function. 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.

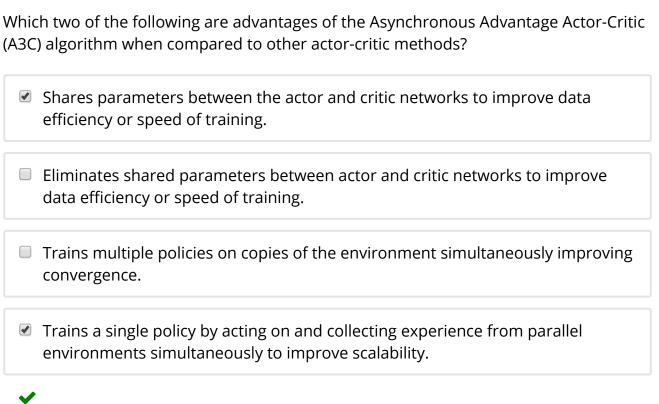


Submit

You have used 1 of 2 attempts

## Question 8

1/1 point (graded)



Submit

You have used 1 of 2 attempts

Learn About Verified Certificates

© All Rights Reserved