Course › Policy ... › Knowle... › Knowle...

Knowledge Checks

Question 1

1/1 point (graded)

Which two of the following are advantages of policy gradient methods over value-function 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.

● Policy gradient algorithms as actors and policy iteration algorithms as critics. ✓

○ Discounted returns as actors and policy interaction algorithms as critics.

○ 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 $\pi\left(s_t\right)$ during an episode to collect information on states, actions and rewards.

○ Computes the return for each episode using the rewards collected.

○ 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.

□ Using the expected returns given the policy as a baseline discourages trajectories with return away from the baseline.

✓

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?

○ The difference between the gradient of the log likelihood and the state value function.

○ 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.

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 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)

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 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