← Graded Quiz

Graded Quiz • 30 min

Due Oct 9, 11:59 PM MDT

Graded Quiz

TOTAL POINTS 10

1. Which approach can find an optimal deterministic policy? (select all that apply) 1 point

   □ Exploring Starts

   □ ϵ-greedy exploration

   □ Off-policy learning with an ϵ-soft behavior policy and a deterministic target policy

2. When can Monte Carlo methods, as defined in the course, be applied? (Select all that apply) 1 point

   □ When the problem is continuing and there are sequences of states, actions, and rewards

   □ When the problem is continuing and there is a model that produces samples of the next state and reward

   □ When the problem is episodic and there are sequences of states, actions, and rewards

   □ When the problem is episodic and there is a model that produces samples of the next state and reward

3. Which of the following learning settings are examples of off-policy learning? (Select all that apply) 1 point

   □ Learning about multiple policies simultaneously while following a single behavior policy

   □ Learning the optimal policy while continuing to explore

   □ Learning from data generated by a human expert

4. If a trajectory starts at time t and ends at time T, what is its relative probability under the target policy π and the behavior policy b? 1 point

   ○ $\sum_{k=t}^{T-1} \frac{\pi(A_k \mid S_k)}{b(A_k \mid S_k)}$

   ○ $\frac{\pi(A_t \mid S_t)}{b(A_t \mid S_t)}$

   ○ $\prod_{k=t}^{T-1} \frac{\pi(A_k \mid S_k)}{b(A_k \mid S_k)}$

   ○ $\frac{\pi(A_{T-1} \mid S_{T-1})}{b(A_{T-1} \mid S_{T-1})}$

5. When is it possible to determine a policy that is greedy with respect to the value functions v_π, q_π for the policy π? (Select all that apply) 1 point

□ When state values v_π and a model are available

□ When state values v_π are available but no model is available.

□ When action values q_π and a model are available

□ When action values q_π are available but no model is available.

6. Monte Carlo methods in Reinforcement Learning work by...

1 point

○ Performing sweeps through the state set

○ Averaging sample returns

○ Averaging sample rewards

○ Planning with a model of the environment

7. Suppose the state s has been visited three times, with corresponding returns 8, 4, and 3. What is the current Monte Carlo estimate for the value of s?

1 point

○ 3

○ 15

○ 5

○ 3.5

8. When does Monte Carlo prediction perform its first update?

1 point

○ After the first time step

○ When every state is visited at least once

○ At the end of the first episode

9. In Monte Carlo prediction of state-values, **memory** requirements depend on (select all that apply)

1 point

□ The number of states

□ The number of possible actions in each state

□ The length of episodes

10. In an ϵ-greedy policy over $\mathcal{A}$ actions, what is the probability of the highest valued action if there are no other actions with the same value?

1 point

○ $1-\epsilon$

○ ϵ

○ $1-\epsilon+\frac{\epsilon}{\mathcal{A}}$

○ $\frac{\epsilon}{\mathcal{A}}$

□ I, **Dhawal Gupta**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code

Save Submit