

True/False and Multiple Choice Questions

Q1i.

Temporal difference method introduces variance but reduces bias.

- True) True
- False) False

Q1ii.

In an episodic environment, the goal of

policy-based RL algorithms is to maximize the exp

- True) True
- False) False

Q2i.

Which of the following is/are off-policy RL method(s)?

- A) Sarsa
- B) REINFORCE
- C) Q-learning
- D) Actor-Critics

Q2ii.

What are the general principles for designing the policy in policy-based

RL method

- A) differentiable
- B) non-deterministic
- C) easy to sample from
- D) additive

Q3i.

What is the “ maximization bias”

in Q-learning, and how do we address