

## True/False and Multiple Choice Questions

**Instructions:** Select the correct option(s) for each question. Some questions may have multiple correct answers. Provide a short explanation for all.

**Question 1i)** Temporal difference method introduces variance but reduces bias.

- True) True
- False) False

**Question 1ii)** In an episodic environment, the goal of policy-based RL algorithms is to maximize the expected return of initial states.

- True) True
- False) False

**Question 2i)** Which of the follow is/are off-policy RL method(s)?

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

**Question 2ii)** What are the general principles for designing the policy in policy-based RL methods:

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

**Question 3i)** What is the “maximization bias” in Q-learning, and how do we address it?