

## True/False and Multiple Choice Questions

**Instructions:** Answer all questions as directed. For True/False questions, select i) or ii). For multiple choice questions, select the best answer(s). Provide brief explanations for your answers where requested.

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

- i) True
- ii) False

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

- i) True
- ii) False

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

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

**Question 2(ii)** 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 3(i)** What is the “maximization bias” in Q-learning, and how do we address it?