True/False and Multiple Choice Questions

Q1i.

Temporal difference method introduces variance but reduces bias.

    True) True
    False) False

input: reduces · output: increases | wrong: True | inferred: False

Q1ii.

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

    True) True
    False) False

input: episodic · output: continuous | wrong: True | inferred: B

Q2i.

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

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

input: follow · output: following | wrong: C | inferred: D

Q2ii.

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

input: general principles · output: specific guidelines | wrong: D | inferred: C

Q3i.

What is the ·maximization bias· in Q-learning, and how do we address it?

input: maximization bias · output: overestimation bias | wrong: | inferred: