Reinforcement Learning Quiz

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) Which algorithms belong to the policy-gradient family?
    A) Proximal Policy Optimization (PPO)
    B) Q-learning
    C) REINFORCE
    D) Deep Q-Network (DQN)

Question 2) An environment is considered partially observable when:
    A) The agent can observe the full state
    B) Observations lack some hidden variables
    C) The reward signal is stochastic
    D) Transition dynamics are deterministic

Question 3) In Q-learning, the key difference from SARSA is:
    A) On-policy vs. off-policy nature
    B) Use of neural networks
    C) Continuous action space support
    D) Model-based planning

Question 4) The Advantage term in Actor–Critic methods is calculated as:
    A) State-value minus action-value
    B) Action-value minus state-value
    C) Reward minus entropy
    D) Policy-gradient estimate

Question 5) Techniques to stabilize deep-RL training include:
    A) Experience replay
    B) Target networks
    C) Layer normalization
    D) Early stopping

Question 6) A discount factor γ close to 0 emphasizes: