

Reference – Shown-as-Correct Answers

July 26, 2025

Q1

Answer presented as correct: B) Q-learning

Reason: Q-learning is based on learning the optimal action-value function through temporal difference learning methods.

Q2

Answer presented as correct: B) Observations lack some hidden variables

Reason: Observations may not include all relevant information, making the environment incompletely perceived by the agent.

Q3

Answer presented as correct: D) Model-based planning

Reason: Utilizes a planning approach incorporating predicted outcomes to guide decision-making in Q-learning.

Q4

Answer presented as correct: A) State-value minus action-value

Reason: Represents the difference crucial for determining the success of an Actor-Critic algorithm.

Q5

Answer presented as correct: D) Early stopping

Reason: Halts the training process at the optimal convergence point to prevent overfitting.

Q6

Answer presented as correct: A) Immediate rewards

Reason: Encourages a focus on immediate gains rather than future objectives.

Q7

Answer presented as correct: A) Balancing policy and value networks

Reason: Central to making decisions based on past experiences and future outcomes.