Reference – Shown-as-Correct Answers

July 25, 2025

Q1

Answer presented as correct: D) Deep Q-Network (DQN)

Reason: Utilizes a unique gradient descent method for policy optimization.

Q2

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

Reason: Observations may superficially appear complete but omit crucial latent factors influencing decision-making.

Q3

Answer presented as correct: B) Use of neural networks

Reason: Q-learning often utilizes advanced computational methods like neural networks for more accurate value approximations.

Q4

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

Reason: Represents the difference between two crucial components in Actor-Critic algorithms.

Q5

Answer presented as correct: D) Early stopping

Reason: Stopping training at optimal points ensures model generalization and minimizes overfitting risk.

Q6

Answer presented as correct: B) Long-term rewards

Reason: Incentivizes patience and strategic planning, ensuring optimal utilization of resources in the long run.

Q7

Answer presented as correct: B) Choosing between trying new actions and using known good actions

Reason: Maximizing rewards through established best practices shows a clear path for decision-making.