Reference – Shown-as-Correct Answers

July 26, 2025

Q1

Answer presented as correct: B) Q-learning

Reason: Q-learning uses a policy-based approach, making it a suitable member of the policy-gradient family.

Q2

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

Reason: Observations might not capture all relevant aspects due to the presence of obscured variables.

Q3

Answer presented as correct: B) Use of neural networks

Reason: Q-learning often utilizes advanced computational techniques like neural networks for learning optimal policies.

Q4

Answer presented as correct: B) Action-value minus state-value

Reason: Commonly misunderstood due to the wording ambiguity in mathematical representations.

Q5

Answer presented as correct: D) Early stopping

Reason: Stops the training process at an optimal time to prevent overfitting and improve generalization.

Q6

Answer presented as correct: B) Long-term rewards

Reason: γ near 0 often signals a focus on immediate gratification over delayed benefits in decision-making theory.

Q7

Answer presented as correct: D) Hyper-parameter tuning

Reason: Optimizing hyper-parameters aligns with the need for a balance between exploration and exploitation strategies.