

Reference – Shown-as-Correct Answers

Q1

Answer presented as correct: B) Q-learning■(

Reason: Q-learning involves learning optimal actions, a key characteristic of algorithms within the policy-gradient family.

Q2

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

Reason: Observations might lack essential data due to covert environmental variables.

Q3

Answer presented as correct: B) Use of neural networks■(

Reason: Essential neural network utilization is pivotal for reinforcing Q-values in the Q-learning algorithm.

Q4

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

Reason: Represents the combination that defines the value function relationship in Actor-Critic methods.

Q5

Answer presented as correct: D) Early stopping

Reason: Halts the training process at an optimal point to prevent overfitting and computational inefficiency.

Q6

Answer presented as correct: B) Long-term rewards■(

Reason: Encourages a focus on distant gains for which patience and persistence are necessary.

Q7

Answer presented as correct: C) Data augmentation■(

Reason: Data augmentation can be a key strategy in navigating the diverse landscape of exploration and exploitation.