Review Test Submission: Quiz 12-14

User	BRIAN THOMAS LOUGHRAN
Course	EN.605.649.82.FA20 Introduction to Machine Learning
Test	Quiz 12-14
Started	12/9/20 7:42 PM
Submitted	12/9/20 7:58 PM
Due Date	12/12/20 11:59 PM
Status	Completed
Attempt Score	Grade not available.
Time Elapsed	15 minutes out of 30 minutes
Instructions	Ten multiple choice or true/false questions will be presented on material from Module 13 and 14 in the course. Please complete the quiz in the time allotted. To best evaluate your understanding, you should try to complete the quiz without using notes or online resources; although, using such resources is permitted if necessary. To encourage this, only 30 minutes will be allotted to complete the quiz. You will have two attempts.
Results Displayed	Submitted Answers, Incorrectly Answered Questions

Question 1

0 out of 10 points

✗ What is meant by "temporal difference error?"

Selected Answer: D. It is the mean squared error when learning a value function.

Question 2

10 out of 10 points

✓ When an agent is learning using reinforcement learning, it generally needs to balance exploration and exploitation. What is the class of problems that captures this balancing act called?

Selected Answer: C. Bandit problems.

Question 3

0 out of 10 points

✗ Solving Markov Decision Processes uses dynamic programming as its principal optimization strategy. Which of the following is a key characteristic of an MDP that make dynamic programming a good choice?

Selected Answer: B. Nondeterministic transitions

Question 4

0 out of 10 points

✗ How might reinforcement learning be posed as a supervised learning problem?

Selected Answer: B. It can’t. It is fundamentally different.

Question 5

10 out of 10 points

✓ What is the main way Q-learning and SARSA differ?

Selected Answer: E. Q-learning is off-policy and SARSA is on-policy.

Question 6

10 out of 10 points

✓ What is the most important condition for proving that Q-learning and SARSA will converge to the optimal policy?

Selected Answer: B. Every state-action pair is visited and updated infinitely often.

Question 7

10 out of 10 points

✓ What do eligibility traces do?

Selected Answer: A.
They provide a mechanism for updating entire sequences of states and actions on each visit to a new state. The extent to which the updates of the parts of these sequences occur is based on how recently these parts were updated previously.

Question 8

0 out of 10 points

✗ What is the purpose of applying a discount factor in the Bellman equation of a Markov Decision Process?

Selected Answer: B. It refines the magnitude of the updates for various states.

Question 9

10 out of 10 points

✓ Value iteration uses a threshold on the Bellman error magnitude to determine when to terminate, but policy iteration does not. Why is policy iteration able to ignore the Bellman error magnitude in its termination decision?

Selected Answer: D.
Policy iteration terminates when the policy stops changing. Since the policy is based on the current value function and a new value function is computed based on an updated policy, once the policy stops changing, so does the value function.

Question 10

10 out of 10 points

What does it mean for a learning algorithm to be off-policy?

Selected Answer: E. When generating sequences for learning, the update rule sometimes use a choice other than what the current policy returns.

Wednesday, December 9, 2020 7:58:43 PM EST

← **OK**