Week 4 Quiz

Q1.

[True or False] Learning in Reinforcement Learning occurs by interacting with an environment and learning from mistakes and successes.

A. True

B. False

Q2.

[True or False] Discounting helps reinforcement learning algorithms to converge more quickly.

A. True

B. False

Q3.

[True or False] You are using reinforcement learning for a self-driving car. The position of the car would be its _________.

A. state

B. action

C. reward

D. transition

Q4.

In a Markov Decision Process, what does the term 'policy' refer to?

A. A) A sequence of actions that the agent will take.

B. B) The mathematical function used to calculate the reward for each state.

C. C) The strategy that the agent employs to determine its action in each state.

D. D) The probability distribution over the set of all possible states.

Q5.

What does the value iteration algorithm in Markov Decision Processes iteratively compute?

A. A) The average reward for each action in all states.

B. B) The probabilities of transitioning from one state to another.

C. C) The optimal values of states until they converge.

D. D) The cumulative reward for an entire episode.

Q6.

When can the iteration process be stopped during value iteration?

A. A) When the values of all states change insignificantly between two consecutive iterations, remaining below a small threshold.

B. B) After a fixed number of iterations, regardless of any changes in state values.

C. C) As soon as the agent receives a negative reward.

D. D) Once the agent finds the shortest path to the goal state.

Q7.

You are controlling a lunar lander (a spacecraft designed to land on the Moon's surface). While landing, the spacecraft can assume continuous positions. The goal is to land successfully between geopoints P1 and P2. Which of the following is a good choice for a reward function:

A. A) A fixed positive reward for every second the lander is airborne, regardless of its position relative to points P1 and P2.

B. B) A high positive reward when the lander successfully lands between points P1 and P2, penalties for landing outside this zone, and a very high penalty for crashing.

C. C) A positive reward proportional to the distance from the Moon's surface, encouraging the lander to stay as high as possible.

D. D) A constant reward for using minimal fuel, regardless of the landing position.

Q8.

What does a policy represent in policy-based reinforcement learning algorithms?

A. A) A mapping from states to probabilities of selecting each possible action.

B. B) A list of rewards that an agent expects to receive in the future.

C. C) The estimated time it will take for an agent to reach the goal state.

D. D) The agent's memory of the states that it has visited in the past.

Q9.

How does policy iteration differ from value iteration?

A. A) Policy iteration involves both policy evaluation and policy improvement steps, while value iteration solely computes a value function.

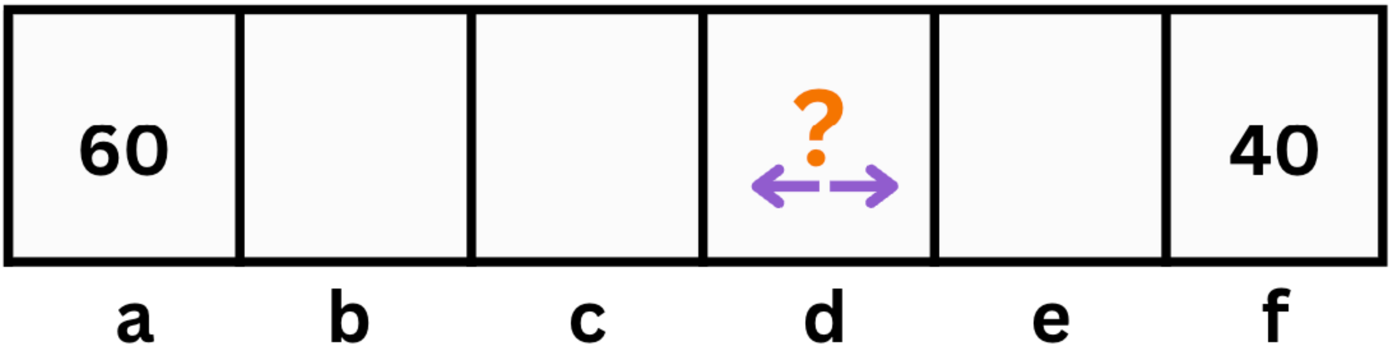
B. B) Policy iteration is a model-free approach, whereas value iteration is a model-based approach.

C. C) Policy iteration does not require a discount factor, but value iteration does.

D. D) Value iteration involves updating policies, while policy iteration updates values based on a fixed policy.

Q10.

Given the states below and a reward function R(s) = -1 for states b, c, d, and e, what is the optimal policy for state 'c' if the discount factor is 0.5?



A. Move left

B. Move right