Quiz - 2

1. **What are the parameters to a state-value function?**

   The state-value function, denoted as V(s), takes a state (s) as input and provides the expected cumulative future reward from that state.

2. **What are the parameters to an action-value function?**

   The action-value function, denoted as Q(s, a), takes a state (s) and an action (a) as input and provides the expected cumulative future reward from taking action (a) in state (s).

3. **What is meant by a policy (π)? Is it deterministic or stochastic?**

   A policy is a strategy or a mapping from states to actions. It defines the agent's behavior in an environment. A policy can be deterministic, meaning it prescribes a specific action for each state, or stochastic, meaning it provides a probability distribution over actions for each state.

4. **Under an optimal policy π^*, it is not necessary that every state has the highest state-value compared to any other policy. (TRUE/FALSE)**

   False. Under an optimal policy π^*, every state must have the highest state-value compared to any other policy. This is a defining characteristic of optimality.`

5. **You are trying to train an agent that has 1000 chances to play a game. Player/Agent tries to throw a ball standing at center and gets different points when the ball reaches different regions as follows:**

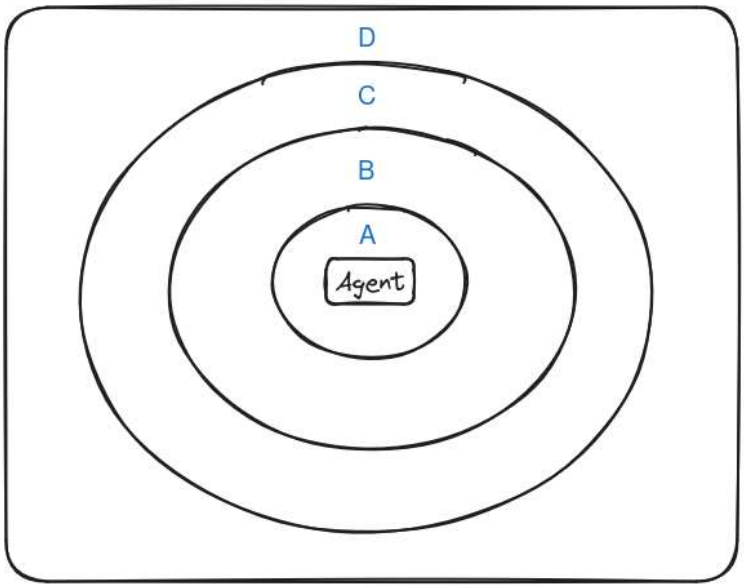
   A:10

   B:20

   C: -30

   D: 40

There is always a 90% chance that the ball reaches the region that was aimed for and a 10% chance that it reaches an adjacent next region(either above or below).



- **How would you choose the rewards to formulate this as an RL problem? Select the reward you would give the agent for throwing the ball to each of these regions such that it learns to get maximum total points by playing the game a 1000 times.**
- **Which region would be the best to aim for each time?**

  Note: This is a stateless RL problem as there are no states involved here. Therefore no state-transition probabilities.

  Assign rewards based on the points obtained in each region. For example:

  Region A: 10 points

  Region B: 20 points

  Region C: -30 points

  Region D: 40 points

  These rewards should reflect the desirability of each outcome.

  Best region to aim for: The best region to aim for would be the one with the highest expected value. In this case, it would be Region D with 40 points.

6. **Can there be more than one optimal policy for a given value function?**

   Yes, there can be more than one optimal policy for a given value function. Different policies may lead to the same expected cumulative future reward, making them equally optimal.