|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |  |

 ****

**Presidency University**

**Bengaluru**

 **School Of Computer Science and Engineering & Information Science**

**SUMMER TERM End-Term Examinations, Aug 2024**

**Date**: 05.08.2024

**Time**: 9.30 am to 12.30 pm

**Max Marks**: 100

**Weightage**: 50%

**Odd Semester**: 2023 - 24

**Course Code**: CSE 3011

**Course Name**: Reinforcement Learning

**Department: SOCSE**

 **Instructions:**

1. *Read the all questions carefully and answer accordingly.*
2. *Do not write any matter on the question paper other than roll number.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Q.No** | **Questions** | **Marks** | **CO** | **RBT** |
| 1 | 1. Define the key elements of RL.
 | 4 | CO1 | L1 |
| 1. Explain the typical steps involved in a reinforcement learning algorithm, such as initializing parameters, interacting with the environment, updating values, and policy improvement. How do these steps contribute to the learning process and the development of an effective policy?
 | 6 | CO1 | L2 |
| 1. Apply the concepts of supervised, unsupervised, and reinforcement learning to a real-world scenario, such as predicting customer churn, clustering customer behavior, and optimizing a recommendation system. For each paradigm, describe how you would set up the problem, choose appropriate algorithms, and evaluate the results.
 | 10 | CO1 | L3 |
| OR |
| 2 | 1. What are the different types of reinforcement learning (RL) environments commonly used in RL research and applications?
 | 4 | CO1 | L1 |
| 1. Describe the mathematical essentials of reinforcement learning, including concepts such as reward functions, value functions, policies, and the Bellman equation. How do these mathematical components contribute to modeling and solving RL problems?
 | 6 | CO1 | L2 |
| 1. Implement a grid world environment as a Markov Decision Process (MDP) using a Python library of your choice. Define the states, actions, rewards, and transition probabilities. Then, use a basic RL algorithm, such as value iteration or policy iteration, to find an optimal policy for this environment. How would you validate and interpret the results of your implementation?
 | 10 | CO1 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 3 | 1. What does 'MC' stand for in the context of reinforcement learning methods, and what is the primary purpose of using MC methods?
 | 4 | CO2 | L1 |
| 1. Describe how the Monte Carlo (MC) prediction algorithm estimates the value function of a given policy. What are the key steps involved in this process, and how does the concept of 'returns' contribute to estimating the value function?
 | 6 | CO2 | L2 |
| 1. Apply both the First-Visit Monte Carlo and Every-Visit Monte Carlo methods to estimate the value function for a simple game environment with 10 states. Implement a Python script to generate episodes, calculate returns, and update the value function estimates for each method. Compare the results obtained from both methods and discuss any differences observed in the estimates.
 | 10 | CO2 | L3 |

 OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 4 | 1. What is the purpose of Monte Carlo exploring starts in reinforcement learning, and how does it help in ensuring adequate exploration of the state-action space?
 | 4 | CO2 | L1 |
| 1. Explain the concept of incremental mean updates and how they are used to estimate the average value of a variable over time. How do incremental mean updates differ from computing the mean directly, and what are the advantages of using this approach in reinforcement learning
 | 6 | CO2 | L2 |
| 1. Consider a game environment with various scenarios, such as winning, losing, and drawing. Implement a simulation to test how different strategies perform in each case. Use this simulation to apply different reinforcement learning algorithms and evaluate their effectiveness across these scenarios. How would you modify your strategy based on the outcomes observed in each case?
 | 10 | CO2 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5 | 1. Match the difference between SARSA and Q-Learning.
 | 4 | CO3 | L1 |
| 1. Explain the difference between prediction and control tasks in reinforcement learning. How do the objectives of these tasks differ, and what are some common approaches used to address each type of task?
 | 6 | CO3 | L2 |
| 1. Using Monte Carlo (MC) prediction, write a Python script to estimate the value function of a policy for a blackjack game. Assume that the agent follows a specific policy, such as always hitting until the sum of the cards is 17 or more. How would you implement the sampling of episodes, calculation of returns, and updating of the value estimates? Additionally, explain how you would evaluate the performance of the trained agent
 | 10 | CO3 | L3 |

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 6 | 1. What are some advantages of Monte Carlo (MC) methods over Dynamic Programming (DP) methods in reinforcement learning?
 | 4 | CO3 | L1 |
| 1. Explain the different types of Monte Carlo (MC) prediction methods used in reinforcement learning, such as first-visit MC and every-visit MC. How do these methods differ in terms of how they calculate value estimates from sampled episodes?
 | 6 | CO3 | L2 |
| 1. Implement the SARSA (State-Action-Reward-State-Action) algorithm for a simple grid world environment where an agent must learn to navigate to a goal. Define the environment, initialize the Q-table, and update the Q-values using the SARSA update rule. How would you ensure the agent explores sufficiently and avoids getting stuck in local optima? Also, describe how you would visualize and evaluate the learned policy.
 | 10 | CO3 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 7 | 1. List out the four steps of algorithms in the upper confidence bound.
 | 4 | CO4 | L1 |
| 1. Explain how Thompson Sampling balances exploration and exploitation in a multi-armed bandit problem. Why is this balance important, and how does the algorithm use probability distributions to achieve it?
 | 6 | CO4 | L2 |
| 1. Given a multi-armed bandit problem with 5 arms, design a simple simulation using the epsilon-greedy method. Set the exploration rate (epsilon) to 0.1 and outline how you would implement the algorithm to select actions and update the value estimates based on rewards. How would you evaluate the performance of your strategy?
 | 10 | CO4 | L3 |

 OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 8 | 1. Choose any four applications for MAB?
 | 4 | CO4 | L1 |
| 1. Describe how Softmax exploration is used in reinforcement learning to handle the exploration-exploitation trade-off. How does the temperature parameter influence the action selection process, and why is this aspect important in learning optimal policies?
 | 6 | CO4 | L2 |
| 1. Install the gym-bandits package in your Python environment. Then, use it to create an environment with a multi-armed bandit setup. Write a Python script to interact with this environment using a simple agent that follows a random action selection strategy. How would you run the script and interpret the results to assess the agent's performance?
 | 10 | CO4 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 9 | 1. State the Bellman equation of the value function and Q function.
 | 4 | CO1 | L1 |
| 1. What are the key components of the Bellman equation used in a Markov Decision Process (MDP) for value function estimation?
 | 6 | CO1 | L2 |
| 1. Given a deterministic grid world environment with defined states, actions, and rewards, use the Bellman equation to calculate the value function of a specific state. Implement the calculations step-by-step, assuming a given policy. How would you verify that your computed value function is accurate for the given environment?
 | 10 | CO1 | L3 |

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 10 | 1. What are two common limitations of Monte Carlo methods in reinforcement learning?
 | 4 | CO2 | L1 |
| 1. Explain the concept of off-policy Monte Carlo control in reinforcement learning. How does it differ from on-policy methods, and what are the advantages of using off-policy approaches in learning algorithms?
 | 6 | CO2 | L2 |
| 1. Imagine you are designing a new recommendation algorithm for an online streaming service. How would you apply the principles of the Exploration-Exploitation dilemma to balance between recommending popular content (exploitation) and introducing users to new, less-known content (exploration)? Provide a detailed strategy for how you would approach this problem.
 | 10 | CO2 | L3 |