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

# PRESIDENCY UNIVERSITY

## BENGALURU

**End - Term Examinations – December 2025**

**Date:** 10- 12- 2025 **Time:** 1.00pm to 04.00pm

| | | |
|---|---|---|
| **School:** SOIS | **Program:** BCA-AIML | |
| **Course Code:** CSA3074 | **Course Name:** REINFORCEMENT LEARNING | |
| **Semester**: V | **Max Marks**: 100 | **Weightage**: 50% |

| CO - Levels | CO1 | CO2 | CO3 | CO4 | CO5 |
|---|---|---|---|---|---|
| Marks | 26 | 24 | 26 | 24 | NA |

**Instructions:**

*(i) Read all questions carefully and answer accordingly.*

*(ii) Do not write anything on the question paper other than roll number.*

## Part A

**Answer ALL the Questions. Each question carries 2marks.** **10Q x 2M=20M**

| | | | | |
|---|---|---|---|---|
| **1.** | What are the key elements of a reinforcement learning system? | **2 Marks** | **L1** | **CO1** |
| **2.** | Differentiate between deterministic and stochastic policies. | **2 Marks** | **L3** | **CO1** |
| **3.** | What is an incremental mean update in Monte Carlo prediction? | **2 Marks** | **L1** | **CO1** |
| **4.** | Explain Monte Carlo control. | **2 Marks** | **L2** | **CO2** |
| **5.** | Define Temporal Difference (TD) learning. | **2 Marks** | **L1** | **CO3** |
| **6.** | Write the update equation for the Q-learning algorithm. | **2 Marks** | **L3** | **CO3** |
| **7.** | Why is the Multi Armed Bandit problem important in reinforcement learning? | **2 Marks** | **L1** | **CO3** |
| **8.** | Describe the function of the temperature ($\tau$) in softmax exploration. | **2 Marks** | **L2** | **CO2** |
| **9.** | Write the UCB action selection formula. | **2 Marks** | **L3** | **CO4** |
| **10.** | Discuss exploration–exploitation trade-off in reinforcement learning. | **2 Marks** | **L3** | **CO4** |

## Part B

**Answer the Questions.** **Total Marks 80M**

| | | | | | |
|---|---|---|---|---|---|
| **11.** | **a.** | Describe the Frozen Lake environment and explain how it illustrates a Markov Decision Process (MDP) in reinforcement learning. | **20 Marks** | **L2** | **CO1** |
| | | **Or** | | | |
| **12.** | **a.** | Discuss the importance of On-Policy and Off-Policy Monte Carlo Control in Reinforcement Learning. | **20 Marks** | **L3** | **CO1** |
| **13.** | **a.** | Explain the types of Monte Carlo prediction techniques and analyze the differences between first-visit and every-visit approaches. | **20 Marks** | **L2** | **CO2** |
| | | **Or** | | | |
| **14.** | **a.** | Briefly explain the Monte Carlo control framework. With the help of an algorithm, explain how on-policy Monte Carlo control achieves optimal policy improvement using the ε-greedy method. | **20 Marks** | **L2** | **CO2** |
| **15.** | **a.** | Elucidate the SARSA algorithm with its update rule and working procedure. How is the optimal policy derived from SARSA in reinforcement learning? | **20 Marks** | **L3** | **CO3** |
| | | **Or** | | | |
| **16.** | **a.** | Compare Dynamic Programming (DP), Monte Carlo (MC), and Temporal Difference (TD) methods with respect to their principles, assumptions, and performance in reinforcement learning tasks. | **20 Marks** | **L3** | **CO3** |
| **17.** | **a.** | Summarize the theory and operation of the Upper Confidence Bound (UCB) algorithm. How does it achieve an effective trade-off between exploration and exploitation in the Multi-Armed Bandit problem? | **20 Marks** | **L2** | **CO4** |
| | | **Or** | | | |
| **18.** | **a.** | Explain the detailed working of the Thompson Sampling algorithm. Using a suitable example, describe how it selects actions and updates beliefs in the Multi-Armed Bandit problem. Write about its advantages and disadvantages. | **20 Marks** | **L2** | **CO4** |