Code No.: AD851PE	**R20**	**H.T.No.**			**8**	**R**						

CMR ENGINEERING COLLEGE: : HYDERABAD
UGC AUTONOMOUS
IV–B.TECH–II–Semester End Examinations (Advanced Supply) – June- 2025
REINFORCEMENT LEARNING
(AI&DS)

[Time: 3 Hours] **[Max. Marks: 70]**

Note: This question paper contains two parts A and B.
Part A is compulsory which carries 20 marks. Answer all questions in Part A.
Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

PART-A (20 Marks)

1. a) What are the KL- UCB algorithms? [2M]
b) What is a probability distribution? Name two discrete and two continuous distributions. [2M]
c) What are the key components of an MDP? [2M]
d) Which is more efficient in practice: value iteration or policy iteration? Why? [2M]
e) What are the limitations of model-based RL? [2M]
f) What is the key assumption behind Monte Carlo methods? [2M]
g) Why is Q-learning called an off-policy algorithm? [2M]
h) Why is Expected SARSA considered more stable than standard SARSA? [2M]
i) What is the main limitation of tile coding? [2M]
j) What happens when λ=1 in TD(λ)? [2M]

PART-B (50 Marks)

2. Given A Multi armed bandit scenario with five arms and their respective reward distributions, apply the Upper Confidence Bound Algorithm to select the best arm to select the best arm for maximizing cumulative records. [10M]

OR

3. What is Thompson Sampling? How does Thompson Sampling differ from UCB-based methods? [10M]

4. Describe the agent-environment interface in the context of Finite Markov Decision Processes. [10M]

OR

5. Why Bellman optimality important for agent to learn and improve their decision making abilities. [10M]

6. Explain the difference between first-visit and every-visit Monto Carlo policy evaluation methods. [10M]

OR

7. What are the advantages of Monte Carlo prediction methods? How does Monte Carlo control differ from Monte Carlo prediction? [10M]

8. Describe the Q-Learning algorithm for TD control. [10M]

OR

9. How does Expected SARSA differ from standard SARSA? How does Expected SARSA compute the expected value update? What is the advantage of Expected SARSA over regular SARSA? [10M]

10. Write the true online TD(λ) algorithm. [10M]

OR

11. Illustrate the concept of n step returns in reinforcement learning. [10M]

****