R22 Regulation *Subject code: 4E8GA*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

(Autonomous, Accredited by NAAC with 'A+' Grade)

B.Tech VIII Semester Regular Examinations, May 2026

REINFORCEMENT LEARNING

(CSE (AI&ML))

Maximum Marks: 60 Date:04.05.2026 Duration: 3 hours

Note: **1.This question paper contains two parts A and B.**
2. Part A is compulsory which carries 10 marks. Answer all questions in Part A.
3. Part B consists of 5 Units. Answer any one full question from each unit.
4. Each question carries 10 marks and may have a, b, c, d as sub questions.

Part-A

All the following questions carry equal marks (10X1M=10 Marks)		Marks	CO	BTL
1.a)	Define probability of an event.	1	1	1
b)	What is regret in bandit problems?	1	1	1
c)	What is a policy in reinforcement learning?	1	2	2
d)	What is meant by a value function?	1	2	1
e)	What is the Reinforcement Learning problem?	1	3	1
f)	What is the key idea behind Monte Carlo methods?	1	3	2
g)	Define the TD(0) algorithm.	1	4	1
h)	What is meant by Expected SARSA?	1	4	2
i)	Define the TD(λ) algorithm.	1	5	1
j)	What is the purpose of experience replay?	1	5	2

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	BTL
2	A. Describe key concepts of linear algebra used in machine learning. B. Explain the concept of regret and distinguish between cumulative regret and instantaneous regret.	5 5	1	1 4
	OR			
3	A. Describe the KL-UCB algorithm and how it improves over the standard UCB approach. B. Explain Thompson Sampling and how it balances exploration and exploitation.	5 5	1	1 2
4	A. Given a simple MDP with states and rewards, compute the value function for a given policy using the Bellman equation. B. Explain the concept of episodic and continuing tasks with suitable real-world examples.	5 5	2	3 2
	OR			
5	A. Differentiate between value iteration and policy iteration algorithms in terms of steps, convergence, and efficiency.	5 5	2	4 5

	B. Discuss how the choice of discount factor (γ) affects the behavior of an agent in infinite horizon problems.			
6	A. Explain the Reinforcement Learning framework, including agent, environment, states, actions, and rewards. B. Illustrate the difference between prediction and control problems with an example.	5 5	3	2 3
	OR			
7	A. Given a sequence of episodes, compute the state value estimates using Monte Carlo prediction. B. Evaluate the effectiveness of Monte Carlo methods in environments with delayed rewards.	5 5	3	3 5
8	A. Explain the concept of bootstrapping and how it is used in the TD(0) algorithm. B. Compare Monte Carlo methods and TD(0) in terms of convergence, bias, and variance.	5 5	4	2 4
	OR			
9	A. Illustrate how Q-learning updates action-value functions using a sample transition. B. Analyze the differences between Q-learning, SARSA, and Expected SARSA in terms of policy behavior and stability.	5 5	4	3 4
10	A. Demonstrate how the TD(λ) algorithm updates value estimates using eligibility traces. B. Explain linear function approximation and describe its geometric interpretation.	5 5	5	3 4
	OR			
11	A. Discuss how control with function approximation differs from tabular methods. B. Develop a reinforcement learning solution using experience replay or fitted Q-iteration for a real-world case study.	5 5	5	4 6