GUJARAT TECHNOLOGICAL UNIVERSITY
BE- SEMESTER–VII (NEW) EXAMINATION – WINTER 2024

Subject Code:3174208 **Date:04-12-2024**

Subject Name: Reinforcement Learning

Time:10:30 AM TO 01:00 PM **Total Marks:70**

Instructions:
1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**
4. **Simple and non-programmable scientific calculators are allowed.**

			MARKS
Q.1	**(a)**	Summarize brief history of reinforcement learning.	**03**
	(b)	Illustrate the general idea of reinforcement learning using Tic-Tac-Toe example.	**04**
	(c)	Describe the Elements of Reinforcement Learning in detail.	**07**
Q.2	**(a)**	Define and formulate probability density function (PDF)	**03**
	(b)	Explain concepts of random variables in probability with suitable example.	**04**
	(c)	Discuss the concept of joint, and conditional probability with their equations.	**07**
		OR	
	(c)	Define Probability Mass Function (PMF)? Explain with suitable example.	**07**
Q.3	**(a)**	Illustrate Markov Reward Process (MRP) with suitable example.	**03**
	(b)	What are the roles of Optimal Value Functions in Markov decision process (MDP)?	**04**
	(c)	Formulate Policy Evaluation to compute the state-value function in dynamic programming.	**07**
		OR	
Q.3	**(a)**	What is The Markov Property? List the criteria to identify the Markov Property.	**03**
	(b)	Solve the Bellman equation for $v*$ for the simple Gridworld problem.	**04**
	(c)	Explain Policy Iteration in detail.	**07**
Q.4	**(a)**	Give overview of Monte Carlo (MC) methods for model free reinforcement learning.	**03**
	(b)	Illustrate with example, On policy and off policy learning.	**04**
	(c)	Write an algorithm for the first-visit MC method for estimating v_π.	**07**
		OR	
Q.4	**(a)**	Explain how to apply Importance Sampling in off-policy technique.	**03**
	(b)	Take down value iteration algorithm.	**04**
	(c)	Write an algorithm for the Every-visit MC method for estimating v_π.	**07**
Q.5	**(a)**	Define Temporal-Difference Learning (TD). Give the overview of Overview TD (0).	**03**
	(b)	What are the advantages of TD Prediction Methods?	**04**
	(c)	Explain Q-Learning for an off-policy TD control algorithm.	**07**
		OR	
Q.5	**(a)**	What it means by the term "Eligibility Traces"? What are two ways to view eligibility traces?	**03**
	(b)	Differentiate State-action-reward-state-action (SARSA) and Q-learning	**04**
	(c)	Describe N-step TD prediction technique in detail.	**07**

****