JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2025

B.Tech-VIII Semester (CSE/IT)

COURSE CODE (CREDITS): 19B1WCI837(3) MAX. MARKS: 15

COURSE NAME: Reinforcement Learning

COURSE INSTRUCTORS: Kuntal Sarkar MAX. TIME: 1 Hour

Note: *(a) All questions are compulsory.*

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

Q.No	Question	CO	Marks
Q1	How do Temporal Difference (TD) methods like SARSA differ from Monte Carlo methods?	CO-2	2
Q2	What is the exploration vs. exploitation trade-off in reinforcement learning?	CO-1	2
Q3	(a) Explain the working principles of the Actor-Critic algorithm in reinforcement learning. (b) Derive the mathematical formulation of the standard policy gradient method.	CO-3	2+2
Q4	(a) What is the difference between a value-based and a policy-based reinforcement learning approach? (b) Explain the concept of a discount factor in reinforcement learning.	CO-1	1.5+1.5
Q5	Given A={a1,a2,a3},Greedy action: a2, ϵ=0.2, Q(s,a)=2, Q(s′,a′)=4, R=3, α=0.1, γ=0.9 Compute Q(s,a) using SARSA and action selection probabilities.	CO-2	2
Q6	Explain the working principles of the Double Q Learning algorithm in reinforcement learning.	CO-2	2