JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATIONS- 2026

B.Tech-V1/VIII Semester (CSE/IT/BT/ECE/CE)

COURSE CODE (CREDITS): ~~25B1WCI640~~/19B1WCI837 MAX MARKS: 35

COURSE NAME: Reinforcement Learning

COURSE INSTRUCTOR: Saurav MAX. TIME: 2 Hours

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

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

(c) Use of calculator is allowed

Q.No	Question	CO	Marks
Q1	Answer all. 1. Define episode in reinforcement learning. 2. State the Bellman equation. 3. What is the discount factor in reinforcement learning? 4. What is a Q-value in Q-learning? 5. What is the difference between on-policy and off-policy learning?	CO1, CO3, CO4	5
Q2	Compare SARSA and Q-Learning based on: 1. policy type, 2. update rule, 3. exploration behavior, 4. convergence characteristics.	CO4	6
Q3	(i) The joint PMF is given by: $P(X=x,Y=y) = (x+y)/12$ for $x=1,2$ and $y=1,2$ Find:- a)Marginal PMF of X b) E[X|Y=1] (ii) A fair die is rolled until all six faces have appeared at least once. Let X be the total number of rolls required. Find E[X].	CO2	3+3
Q4	(i) Prove that the Bellman Optimality Operator is a contraction mapping. (ii) Prove that the Bellman Optimality Equation has a unique fixed point corresponding to the optimal value function V^*.	CO4	3+4

Q5	A Markov Decision Process has two states s1,s2 and one action per state. Transition and reward details: • From s1: move to s2, reward = 4 • From s2: remain in s2, reward = 2 Discount factor γ=0.5 and initial values $V_0(s1)=V_0(s2)=0$. Perform two iterations of Value Iteration.	CO3	5
Q6	Design a Reinforcement Learning agent for autonomous traffic signal control in a smart city. Clearly define the state space, action space, reward function, and learning objective. Discuss the challenges involved in real-world deployment.	CO3	6