JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2025

B.Tech-VIII Semester (CSE/IT)

COURSE CODE (CREDITS): 19B1WCI837(3) MAX. MARKS: 35

COURSE NAME: Reinforcement Learning

COURSE INSTRUCTORS: Kuntal Sarkar 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

Q.No	Question	CO	Marks
Q1	(a) Explain how Monte Carlo Tree Search works.	CO-8	2
	(b) Implement a basic version of MCTS to solve a simple turn –based game like Tic-Tac-Toe. How does the algorithm balance exploration and exploitation?		3
	(c) What is the RL^2 algorithm in the context of meta reinforcement learning? Explain how it differs from traditional reinforcement learning?		2
	(d) How would you design a neural architecture to implement RL^2 and in what types of tasks does it outperform standard RL algorithms?		3
Q2	(a) Describe Feudal RL architecture and it differs from standard flat RL.	CO-7	2
	(b) What are the key advantages and challenges of using Feudal RL in complex environments?		2
	(c) How does the A* algorithm work? What if the search space in A* algorithm is not a grid and is a graph?		2
Q3	(a) How does Dueling DQN improve upon the standard DQN?	CO-6	2
	(b) Explain how the value and advantage streams		2

	are combined. (c) In what scenarios does Dueling DQN provide a significant benefit over standard DQN?		2
Q4	(a) Describe a typical pipeline for Inverse RL and give a practical example where IRL is preferable over RL.	CO-9	2
	(b) Describe all the methods of IRL (write pseudo code).		4
Q5	(a) How does IRL compare to GAIL in terms of performance and practicality?	CO-9	2
	(b) Suppose MAML shows slow convergence or poor adaptation after inner loop updates. What are 3 possible reasons ?		2
Q6	You are performing MCTS in a simple turn-based game. At a particular node, you have the following statistics for 3 child actions: **Action** **Visits N_i** **Total Reward W_i** A1 10 7.0 A2 20 14.0 A3 5 5.0 The total number of visits to the parent node is N=35. Where: • W_i: Total reward of child i • N_i: Number of times child i has been visited • N: Total number of times the parent has been visited • c= $\sqrt{2}$ (exploration constant) Compute the UCB value for each action and select the best action to explore next.	CO-8	3