3 Hours **Total Marks: 80**

Note:

1. Question number one is compulsory
2. Solve any 3 out of remaining five
3. Figures to the right indicates full marks

Q1	a	What is the significance of the action space in reinforcement learning?	**05**
	b	What is the role of the learning rate in updating action-value estimates?	**05**
	c	How do rewards drive the agent's behaviour toward achieving its goals?	**05**
	d	What is the exploration-exploitation trade-off in reinforcement learning?	**05**
Q2.	a	How can you implement asynchronous updates in a reinforcement learning algorithm?	**10**
	b	You are using Q-Learning to optimize your strategy for playing a slot machine with 3 arms. How would you initialize your Q-values and update them based on the rewards received over 100 plays?	**10**
Q3	a	Imagine you're designing a simple game where a player controls a character to navigate through a maze to reach a treasure chest. The player receives a reward of +5 points upon reaching the treasure chest and -1 point for each move taken. Assume the player starts at the entrance of the maze. i. If the player reaches the treasure chest in 5 moves, what is their total reward? ii. What is the maximum possible reward the player can earn, assuming they take the fewest number of moves to reach the treasure chest? iii. If the player gets stuck in the maze and takes 15 moves to reach the treasure chest, what is their total reward?	**10**
	b	Explain the components of an MDP in detail.	**10**
Q4	a	What are the challenges of applying Q-Learning to large-scale problems, and how can they be mitigated?	**10**
	b	What are the main components of GPI? Give an example of a reinforcement learning algorithm that uses GPI.	**10**
Q5	a	What are the challenges in implementing job-shop scheduling in a real-world scenario?	**10**
	b	What are gradient bandit methods, and how do they differ from action-value-based methods?	**10**
Q6	a	How does policy evaluation differ in deterministic and stochastic environments?	**10**
	b	What is the purpose of an action-value function Q(s,a) in reinforcement learning?	**10**
