

Duration 3 Hours

Total marks 80

Note: (1) Question No. 1 is compulsory.

(2) Attempt any three questions out of the remaining five questions

(3) Figures to the right indicates full marks

- Q1**
- a Explain types of reinforcement learning. **5**
  - b What are the methods used for policy evaluation? **5**
  - c How does TD prediction differ from Monte Carlo prediction? **5**
  - d Describe the main components of an elevator dispatching system. **5**
- Q2.**
- a You are playing a slot machine with three arms. Each time you pull an arm, you either win \$1 or lose \$1 with equal probability. You decide to randomly choose an arm to pull each time. If you play the slot machine 100 times, how much money do you expect to win or lose on average? **10**
  - b Explain with an example scenario where Monte Carlo control might be applied. **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 +10 points upon reaching the treasure chest and -1 point for each move taken. Assume the player starts at the entrance of the maze. **10**
    - 1. If the player reaches the treasure chest in 15 moves, what is their total reward?
    - 2. If the player reaches the treasure chest in 20 moves, what is their total reward?
    - 3. What is the maximum possible reward the player can achieve in this game?
    - 4. What would be the reward if the player gets stuck in the maze indefinitely?
  - b Explain the advantages and disadvantages of asynchronous updates in dynamic programming. **10**
- Q4**
- a Describe the Q-learning algorithm for TD control **10**
  - b Explain the exploration-exploitation trade-off and its significance in reinforcement learning **10**
- Q5**
- a Explain the difference between first-visit and every-visit Monte Carlo policy evaluation methods. **10**
  - b Explain how scheduling algorithms optimize routes, minimize delivery times, and allocate resources effectively to meet customer demands while reducing operational costs. **10**

- Q6** a Describe the components of an MDP, including states, actions, transition probabilities, and rewards. **10**
- b What are the advantages and disadvantages of action-value methods compared to other reinforcement learning techniques? **10**
-