Paper / Subject Code: 53173 / Reinforcement Learning (DLOC - V)

Time: 3 Hours **Marks: 80**

Note: 1. Question 1 is compulsory

2. Answer any three out of the remaining five questions.

3. Assume any suitable data wherever required and justify the same.

Q1 a) Describe generalized policy iteration. **[5]**

b) Suppose $\gamma = 0.9$ and the reward sequence is R1 = 2 followed by an infinite sequence of 7s. What are G1 and G0? **[5]**

c) What are the key features of reinforcement learning? **[5]**

d) Describe temporal-difference (TD) prediction with an example. **[5]**

Q2 a) Consider a k-armed bandit problem with k = 4 actions, denoted 1, 2, 3, and 4. Consider applying to this problem a bandit algorithm using ε -greedy action selection, sample-average action-value estimates, and initial estimates of Q1(a) = 0, for all a. Suppose the initial sequence of actions and rewards is A1 = 1, R1 =1, A2 = 2, R2 = 1, A3 = 2, R3 =2, A4 = 2, R4 = 2, A5 = 3, R5 = 0. On some of these time steps the ε case may have occurred, causing an action to be selected at random. On which time steps did this definitely occur? On which time steps could this possibly have occurred? **[10]**

b) What is reinforcement learning. Explain the elements of reinforcement learning. **[10]**

Q3 a) Illustrate through an example the use of Monte Carlo Methods. **[10]**

b) Describe the application of reinforcement learning to the real world problem of Job-Shop Scheduling. **[10]**

Q4 a) Describe temporal-difference (TD) control using Q-Learning. **[10]**

b) Explain how upper confidence bound (UCB) action selection generally performs better than ε-greedy action selection with a suitable example. **[10]**

Q5 a) Describe Monte Carlo Estimation of Action Values. **[10]**

b) Describe asynchronous dynamic programming with an example. **[10]**

Q6 a) What are Goals and Rewards? Explain with a suitable example. **[10]**

b) Explain Tracking a Nonstationary Problem with an example. **[10]**