

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2024

B.Tech Semester VIII (CSE/IT)

COURSE CODE (CREDITS): 19B1WCI837 (3)

MAX. MARKS: 35

COURSE NAME: REINFORCEMENT LEARNING

COURSE INSTRUCTORS:DHA

MAX. TIME: 2 Hours

---

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

*(b) Marks are indicated against each question in square brackets.*

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

---

- Q1. a) Define temporal difference (TD) learning and its role in reinforcement learning.  
b) Describe the TD(0) algorithm, including its key components and how it updates the value function.  
c) Discuss the advantages and limitations of TD learning compared to other reinforcement learning methods such as Monte Carlo methods. **[CO-3, Marks:3+3+3]**

- Q2. a) Explain how TD backup updates value estimates incrementally at each time step based on the difference between the estimated value of the current state and the estimated value of the next state.  
b) What is the primary objective of TD backup? How exhaustive the backup is? **[CO-3, Marks: 3+3+3]**  
c) Explain On Policy and Off Policy Learning Algorithms.

- Q3. a) Provide a real-world application where Q learning can be successfully applied? Justify.  
b) Explain how Model Free Control is different from Model Free Prediction.  
c) How does TD learning differ from SARSA algorithm?

**[CO- 3, Marks: 3+3+3]**

- Q4. Explain in detail the Sarsa( $\lambda$ ) windy world example with making all the necessary assumption.

**[CO- 4, Marks: 8]**