Neural Networks and Deep Learning with Business Applications

SP Jain School of Global Management

Assignment L8

Temporal Difference Learning in a Simple Environment

In this assignment, you will explore the concept of Temporal Difference (TD) Learning, a fundamental technique in reinforcement learning. You will implement a TD(0) learning algorithm in a defined environment and analyze its behavior over time.
Environment Description: The environment consists of four states (0, 1, 2, 3) with the following dynamics:

- Transitions:
  - From 0 to 1 or 3
  - From 1 to 0 or 2
  - From 2 to 1 or 3
  - From 3 to 0 or 2
- Rewards:
  - +1 for reaching state 3
  - -0.1 for each movement in other transitions

Tasks for Submission:

1. Implement a TD(0) learning algorithm to estimate the value of each state under a random policy.
   - Use a learning rate (α) of 0.1 and a discount factor (γ) of 0.9.
   - Run the algorithm for 1000 episodes.
2. Analyze the results after 1000 episodes.
   - What are the values of each state?
   - How do these values represent the expected returns under the given policy?

Submission Format: Your submission should include the following components:

- Source code of your TD(0) algorithm implementation.
- A report describing the algorithm and the analysis of the results.
- Any relevant visualizations or diagrams that support your analysis.

Additional Instructions:

- Ensure that your code is well-commented and follows best practices for readability.
- The report should be concise, focusing on the key aspects and findings of your work.