Analysis of Geopolitical factors over Global Supply Chain using RL Assignment 1

Adarsh Prajapati
Winter in Data Science

December 15, 2025

1 Q Learning

Q Learning is a fundamental off-policy model-free learning algorithm which I hope you might have studied by now. If you are still confused about the terms I am using here, it's best to look them up.

You are tasked with implementing tabular Q-learning to solve the 8×8 FrozenLake problem. You can use the OpenAI gymnasium Q Learning environment.

Your goal is not merely to make the agent reach the goal occasionally, but to understand why it succeeds, when it fails, and how hyperparameters affect the result.

Look at the base code on the github of WiDS 25.

Experimental Analysis:

Conduct the following experiments and report results quantitatively.

1. **Exploration Strategy:**
   - Compare fixed ϵ with decaying ϵ_t
   - Plot success rate vs. training episodes
2. **Learning Rate Sensitivity:**
   - Evaluate performance for at least three values of α
   - Discuss stability and convergence speed
3. **Discount Factor Sensitivity:**
   - Compare policies learned using $\gamma \in \{0.90, 0.95, 0.99\}$
   - Analyze differences in risk-sensitive behavior near holes

After training:

1. Extract the greedy policy $\pi(s) = \arg\max_a Q(s, a)$
2. Visualize the policy on the 6×6 grid
3. Estimate the empirical success probability over 1,000 evaluation episodes

2 On-Policy Learning with SARSA

In this part, you will study the on-policy reinforcement learning algorithm SARSA and contrast it with Q-learning, building on your previous implementation for the FrozenLake (6×6) environment.

1. Implement the SARSA algorithm using the same environment, state representation, and reward structure as in the Q-learning task. Use an ε-greedy policy for action selection.

2. Ensure that the learning rate, discount factor, and exploration schedule are identical to those used in the Q-learning implementation, unless explicitly stated otherwise.

3. Train both agents (Q-learning and SARSA) for the same number of episodes and record:
   - Episode-wise cumulative reward
   - Success rate (reaching the goal)
   - Final learned policy

4. Compare the learned policies of SARSA and Q-learning. In particular, analyze how each algorithm behaves near holes and risky transitions in the FrozenLake environment.

5. Provide a theoretical explanation for any observed differences in performance or policy structure by addressing the following:
   - The role of on-policy versus off-policy updates
   - How exploration is incorporated into the value updates
   - The implications of stochastic transitions on learned behavior