CSCN8020 – Assignment 2: Q-Learning Report

Github link: **https://github.com/theRedeemer997/CSCN-8020-Assignment-2.git**

Name: Manu Mathew

Course: CSCN8020 – Reinforcement Learning

Environment: Taxi-v3 (500 discrete states, 6 actions)

1. Introduction

This report presents the implementation and evaluation of a Q-Learning agent in the Taxi-v3 environment. The objective is to train an agent to efficiently pick up and drop off passengers by learning optimal routes using reinforcement learning principles. The Q-Learning algorithm updates its value estimates based on observed rewards, following the Bellman equation.

2. Experimental Setup

• Environment: Taxi-v3 (500 states, 6 actions)
• Reward structure: +20 for successful drop-off, -10 for illegal actions, -1 per step
• Baseline parameters:
  - Learning Rate $\alpha = 0.1$
  - Exploration Factor $\varepsilon = 0.1$
  - Discount Factor $\gamma = 0.9$
• Deliberate parameter variations:
  - Learning Rate $\alpha \in [0.01, 0.001, 0.2]$
  - Exploration Factor $\varepsilon \in [0.2, 0.3]$
• Training episodes: 3000 per configuration

3. Results and Metrics

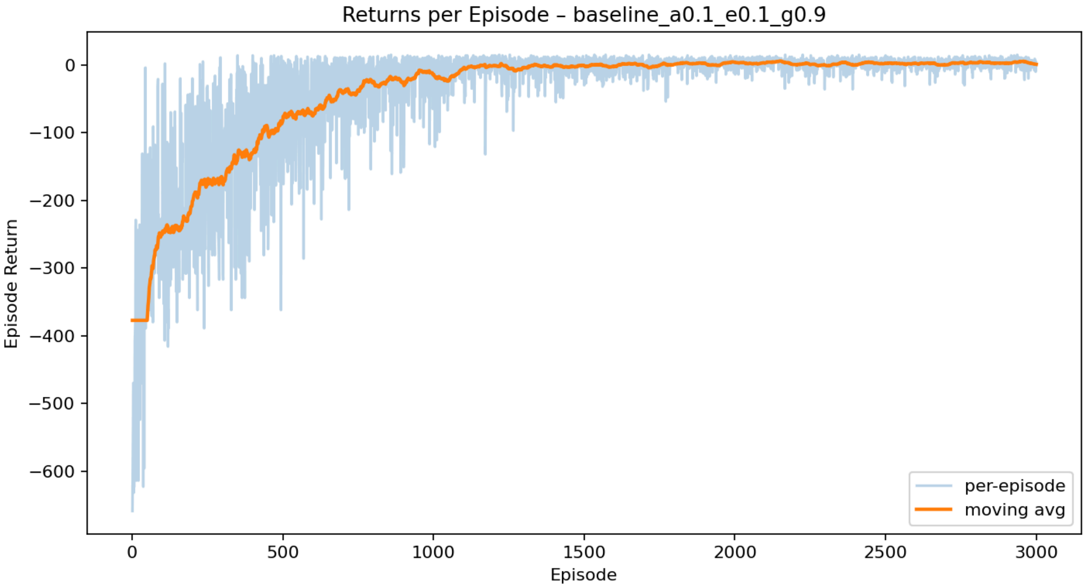
The following metrics were recorded for each training configuration:

Configuration	α	ε	γ	Avg Return	Avg Steps	Total Steps
Baseline	0.1	0.1	0.9	-36.97	40.42	121,251
α=0.01	0.01	0.1	0.9	-203.55	153.50	460,510
α=0.001	0.001	0.1	0.9	-263.15	186.66	559,982
α=0.2	0.2	0.1	0.9	-20.22	28.88	86,650
ε=0.2	0.1	0.2	0.9	-51.76	43.66	130,972
ε=0.3	0.1	0.3	0.9	-69.32	46.78	140,356
Best	0.2	0.1	0.9	-20.22	28.88	86,650

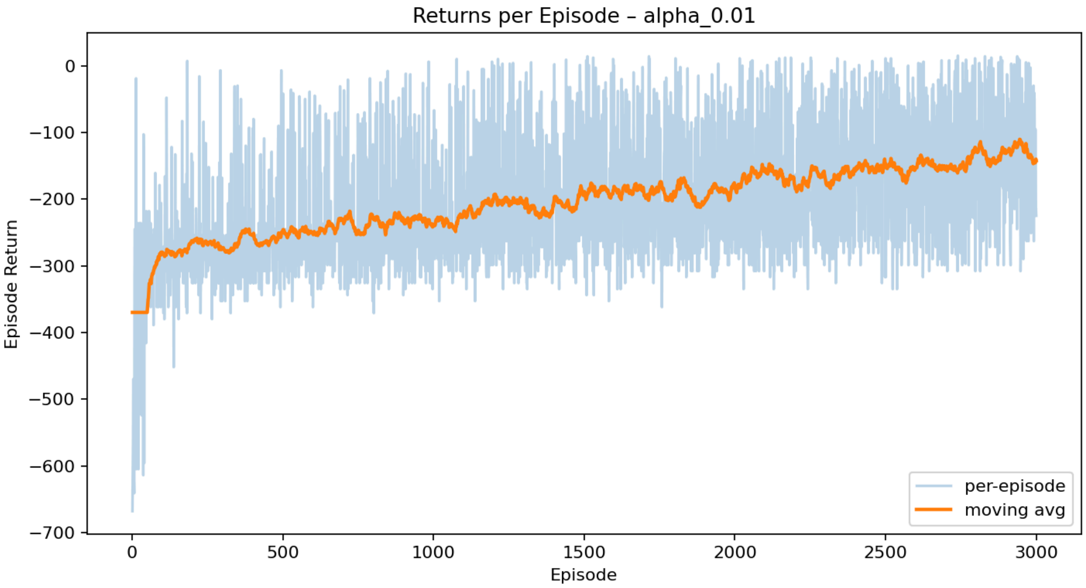
The table indicates that the highest performance was achieved with α=0.2, ε=0.1, γ=0.9, showing faster convergence, fewer steps, and higher average return compared to the baseline.

4. Learning Curves (Returns per Episode)

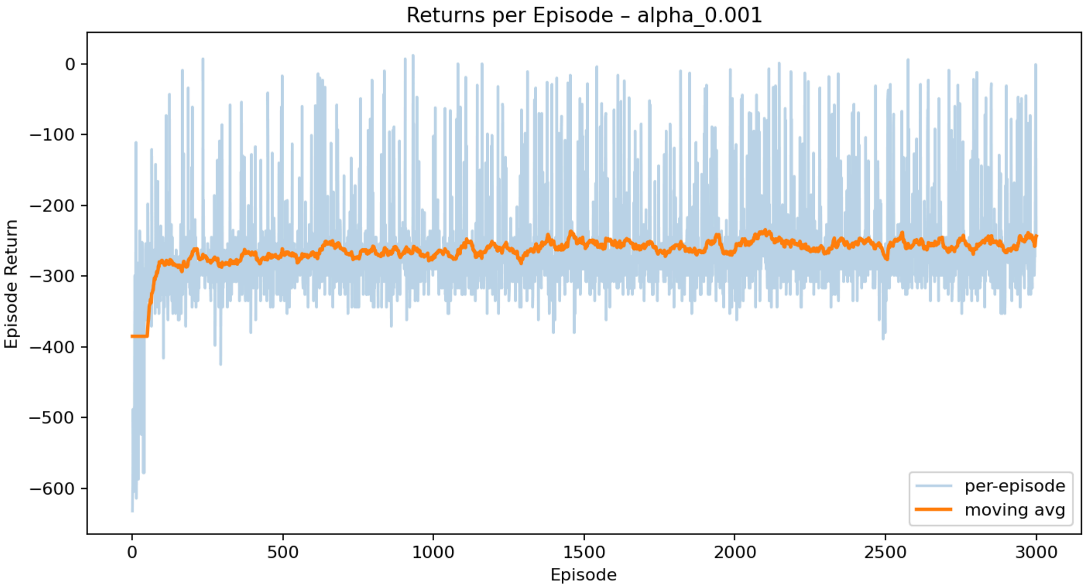
- Baseline (α=0.1, ε=0.1, γ=0.9)



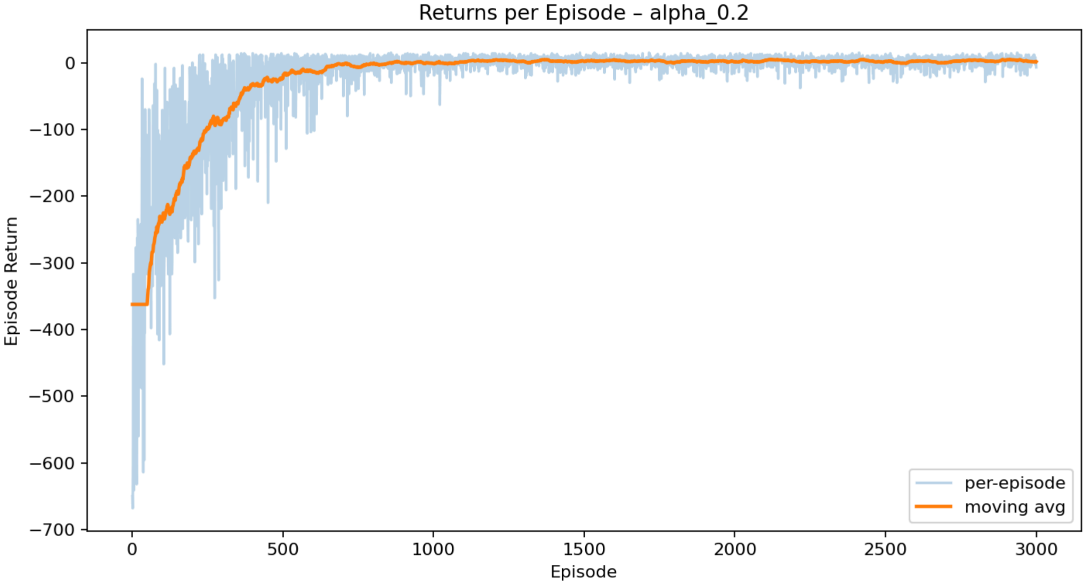
- Learning Rate α=0.01



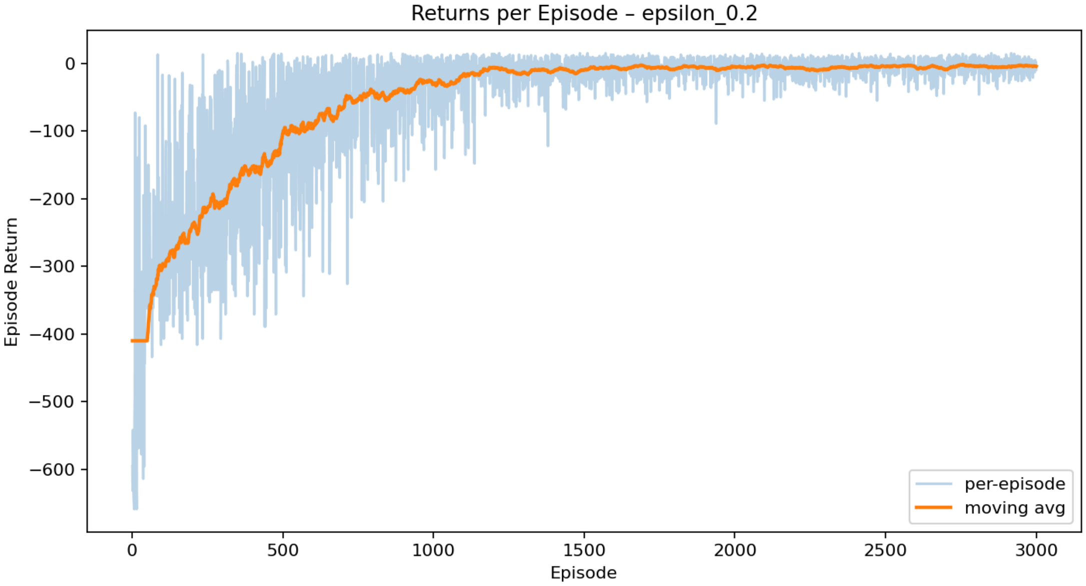
- Learning Rate α=0.001



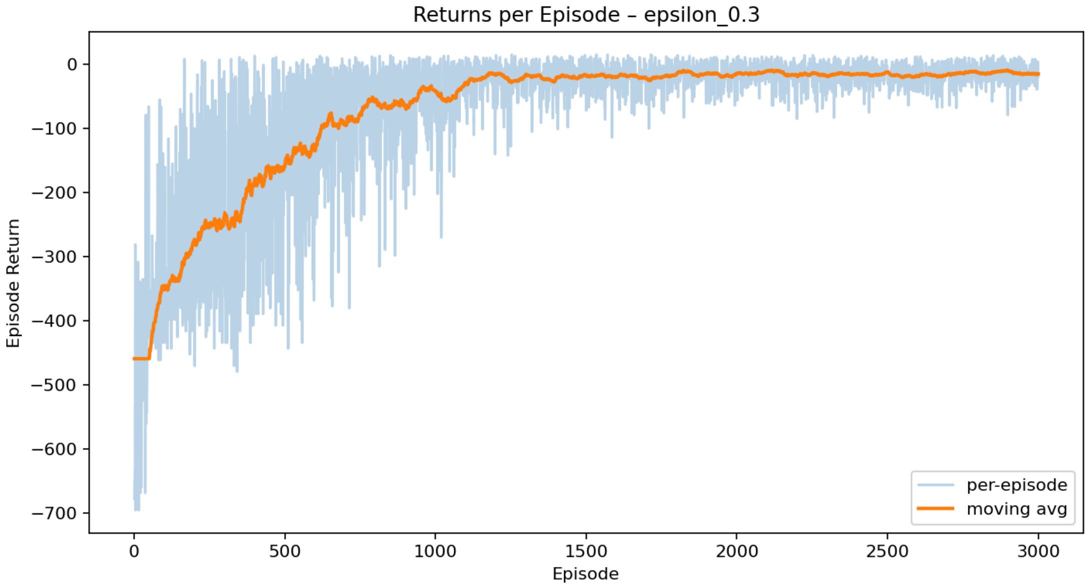
- Learning Rate α=0.2



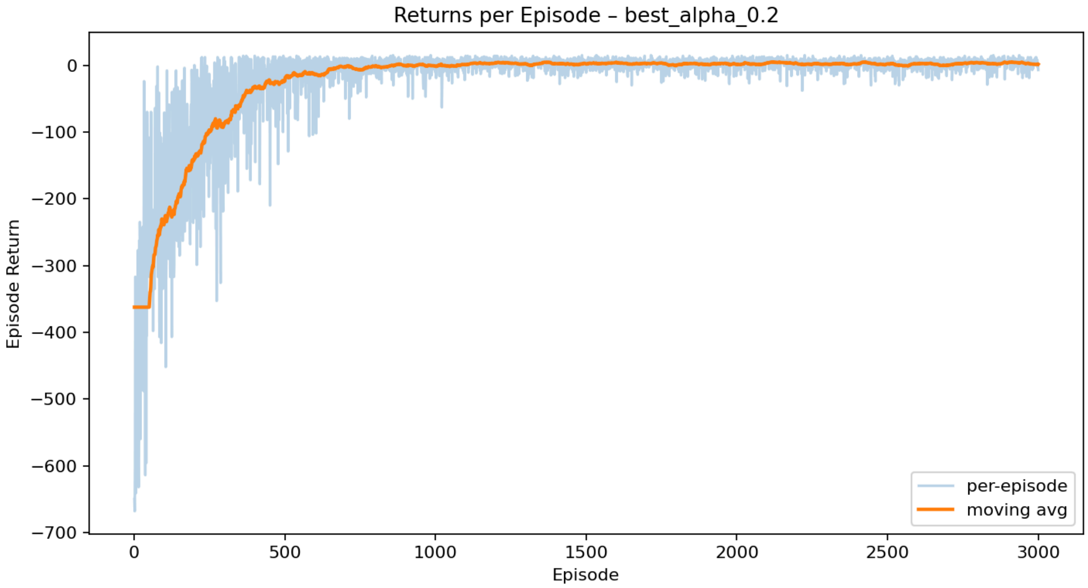
- Exploration Factor ε=0.2



- Exploration Factor ε=0.3



- Best Configuration (α=0.2, ε=0.1, γ=0.9)



5. Observations and Discussion

• When α was decreased to 0.01 and 0.001, the agent learned very slowly. Smaller learning rates resulted in minimal Q-value updates, producing poor returns and longer trajectories.
• Increasing α to 0.2 accelerated learning, enabling faster convergence and the best performance.
• Increasing ε to 0.2 and 0.3 increased exploration, which helped early in training but caused more random actions and slightly reduced average returns.

- The baseline (α=0.1, ε=0.1, γ=0.9) provided stable performance but slower convergence.
- The best balance between exploration and exploitation was achieved with α=0.2, ε=0.1, γ=0.9, offering the highest average return and lowest average steps.

6. Conclusion

Through systematic experimentation, the optimal Q-learning hyperparameter combination was found to be α=0.2, ε=0.1, and γ=0.9. This configuration achieved the highest average return (-20.22) and the fewest steps (28.88) among all tested setups. The results confirm that a higher learning rate improves convergence speed, while maintaining moderate exploration ensures stability. Overall, Q-learning effectively enabled the Taxi agent to learn optimal pick-up and drop-off behavior through repeated interaction with the environment.