Assignment 2

CSCN8020 – Reinforcement Learning Programming

Q-Learning Variability Analysis

Name: Preetpal Singh

Student ID: 8804336

1. Introduction

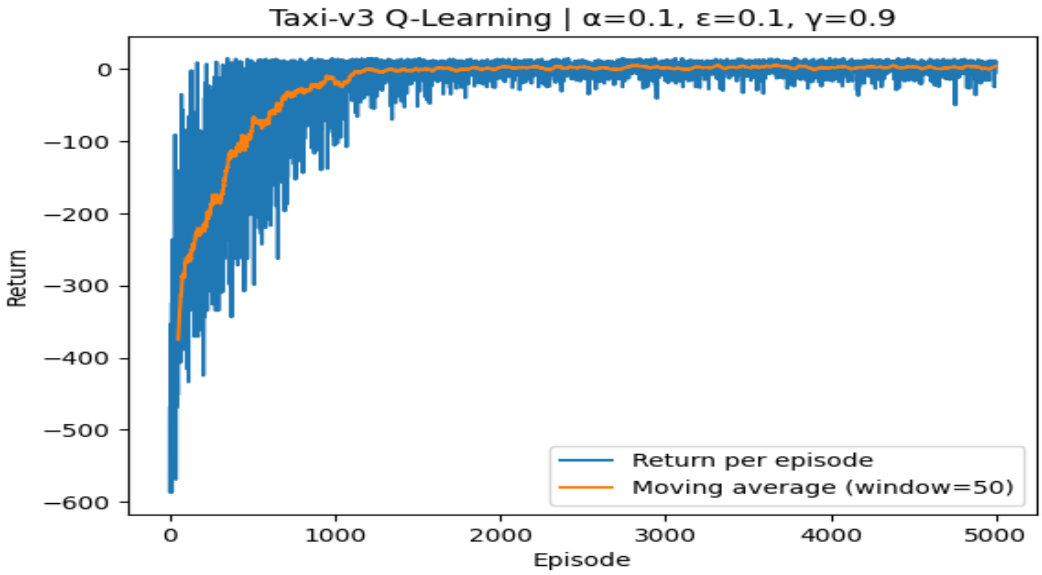
This report analyzes the variability in Q-Learning performance under different hyperparameter configurations, focusing on how the learning rate (α) and exploration factor (ε) influence convergence, average returns, and stability. Experiments were conducted using a discrete environment over 5,000 episodes with a fixed discount factor $\gamma = 0.9$.

Parameter	Description	Values Tested
Episodes	Number of training episodes	5000
α (Learning Rate)	Controls Q-value update speed	0.001, 0.01, 0.1, 0.2
ε (Exploration Rate)	Probability of random action	0.1, 0.2, 0.3
γ (Discount Factor)	Weight of future rewards	0.9 (fixed)

3. Metrics and Performance Observations

3.1 Base Run – α=0.1, ε=0.1

Mean Return: 1.98
Mean Steps: 15.06
Training Time: 17.96 seconds
This configuration demonstrated fast, stable convergence with near-optimal returns.

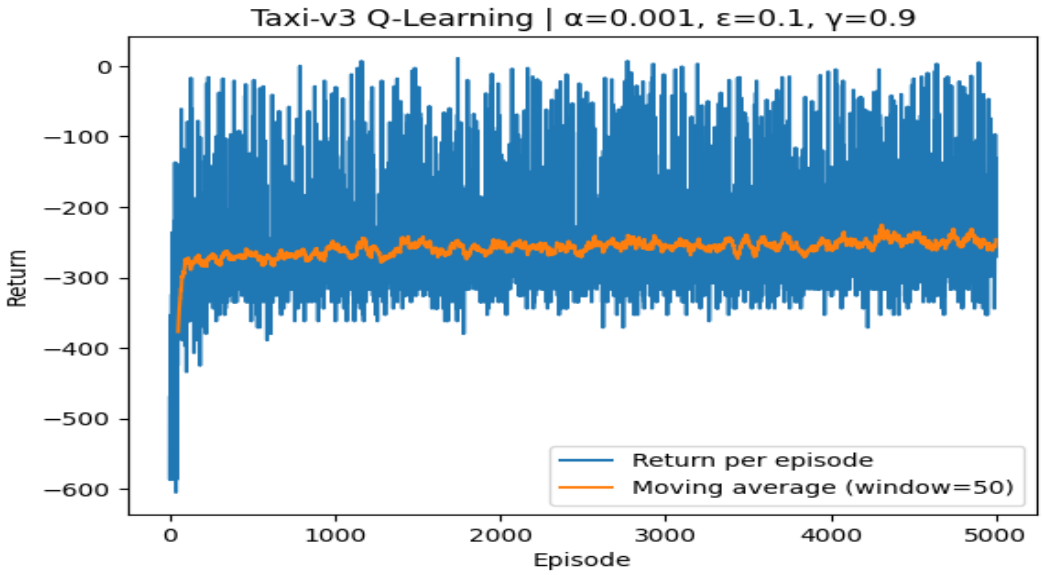


3.2 Learning Rate Sweep

α	Mean Return	Mean Steps	Train Time (s)	Observation
0.001	-253.76	185.84	124.05	Failed to converge – extremely slow learning
0.01	-72.22	70.81	71.30	Partial improvement, unstable
0.1	1.98	15.06	17.96	Optimal performance
0.2	1.97	14.98	5.38	Slightly faster, stable

3.3 Exploration Rate Sweep

ε	Mean Return	Mean Steps	Observation
0.1	1.98	15.06	Fast convergence, less exploration
0.2	-4.77	17.31	Moderate exploration, slightly worse returns
0.3	-14.91	19.98	High exploration, noisy policy



4. Discussion

A very small learning rate (α=0.001) drastically slows convergence and leads to negative returns. Moderate α values (0.1–0.2) achieve the best performance. High exploration (ε>0.3) introduces instability, while low exploration improves speed but risks local optima. The ideal trade-off is α=0.1 and ε=0.1.

5. Best Configuration

Parameter	Optimal Value	Result
α	0.1	Stable convergence, mean return 1.98
ε	0.1	Efficient exploration-exploitation balance
γ	0.9	Consistent reward discounting
Episodes	5000	Adequate for convergence

6. Conclusion

The results confirm that Q-Learning achieves optimal and stable convergence at α=0.1 and ε=0.1. Very low learning rates or excessive exploration lead to poor performance and high variability. Q-Learning demonstrates low variance, high stability, and is effective for real-time RL tasks.

7. References

1. Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press.
2. Q-Learning Assignment Output Files (2025).
3. OpenAI (2025). LLM-Aided Reinforcement Learning Report Generation.