Reinforcement Learning TP3

Authors
Eithan Nakache - Camil Ziane
Promotion 2025

EPITA
École pour l'Informatique et les Techniques Avancées

1 Q-Learning / Epsilon Scheduling Agent

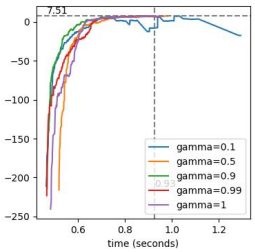

Figure 1: Gamma

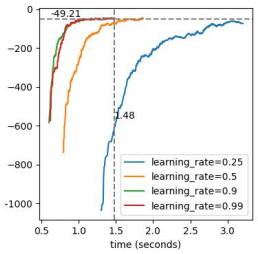

Figure 2: Learning Rate

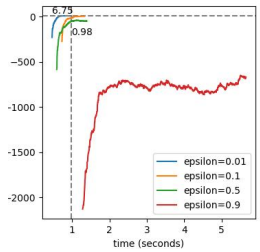

Figure 3: Epsilon

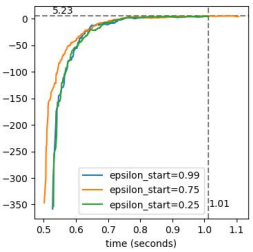

Figure 4: Start Epsilon

Figure 5: Comparison of model rewards across different hyperparameter values over time.

The final parameters were: **Gamma = 0.9**, **Learning Rate = 0.99**, **Epsilon = 0.01**, and **Start Epsilon = 0.25**.

2 SARSA Agent

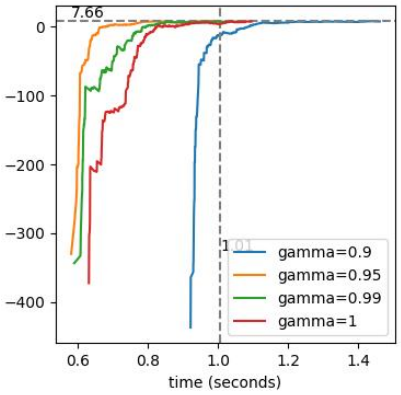

Figure 6: Gamma

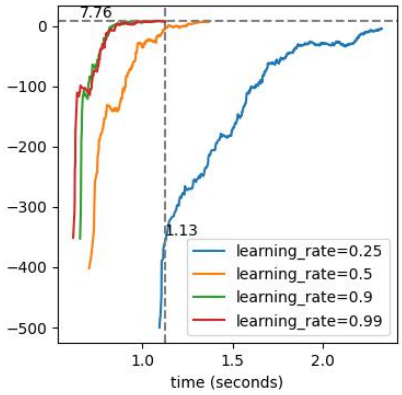

Figure 7: Learning Rate

Figure 8: Comparison of different hyperparameter values for SARSA.

The best SARSA agent was found with **Gamma = 0.9** and **Learning Rate = 0.99**.

3 Comments and Comparison

The best hyperparameters for each agent were obtained iteratively, adjusting one hyperparameter at a time while keeping the previously optimized values fixed. We observed that for the taxi problem, a model with less randomness is favored, and gamma does not play a major role.

Agent	Best Reward	Best Time (seconds)
Q-Learning Agent	**7.44**	1.47
Q-Learning Agent (Epsilon Scheduling)	4.11	**1.0126**
SARSA Agent	7.40	1.0961

Table 1: Comparison of best rewards and times for different agents. (With optimized hyperparameters)