**Rapport de comparaison : Q-Learning et SARSA sur le jeu Taxi-v3**

**Implémentation**

**Q-Learning**

Q-Learning est un algorithme off-policy qui met à jour la fonction de valeur en utilisant l'action optimale attendue, indépendamment de l'action réellement prise. L'agent choisit ses actions selon une politique d’epsilon-greedy, qui favorise l'exploration initiale avant de se concentrer sur l'exploitation. Une version avec et sans scheduling de l’epsilon a été implémentée pour observer l'impact de la diminution progressive de l'exploration.

**SARSA**

Contrairement à Q-Learning, SARSA est un algorithme on-policy où la mise à jour de la fonction de valeur est basée sur l'action réellement prise. Il exploite également une politique epsilon-greedy, mais met à jour la valeur des états-actions en fonction de la politique suivie par l'agent.

**Paramètres**

Les deux algorithmes utilisent les mêmes paramètres de base :

* Taux d'apprentissage (alpha) : 0.5
* Facteur de discount (gamma) : 0.99
* Epsilon initial pour epsilon-greedy : 0.1 pour Q-Learning et SARSA, avec une version de Q-Learning où epsilon diminue progressivement au fil du temps.

Ces paramètres ont été trouvé de manière empirique, avoir un epsilon plus faible limite l’exploration contrôlant ainsi les résultats pour qu’il atteigne le seuil de récompenses demandé.

**Résultats**

**Performance**

Les deux versions de Q-Learning convergent plus lentement que SARSA, la version avec scheduling d’epsilon prend plus de temps pour converger mais atteint une récompense moyenne plus élevé. Le scheduling d’epsilon permet de mieux gérer la transition entre exploration et exploitation. SARSA obtient la meilleur récompense moyenne.

Performance des modèles sur leur 100 derniers entraînements :

* **Q-Learning sans scheduling** : Récompense moyenne de 3.08
* **Q-Learning avec scheduling** : Récompense moyenne de 5.49
* **SARSA** : Récompense moyenne de 7.72

Le scheduling d’epsilon dans Q-Learning améliore les performances en équilibrant progressivement exploration et exploitation, mais ne parvient toujours pas à surpasser SARSA, ce qui suggère que l'approche sur-politiques est plus efficace dans cet environnement.