TP 2

Implémentation de Q-learning

Objectif

Ce TP vise d'implémenter l'algorithme d'apprentissage par renforcement Q-learning. Le problème à résoudre est la grille à 6 états du premier TP. Il vous est demandé également de créer d'un joueur IA de morpion qui apprend par renforcement.

Pour ce TP, vous avez un oracle (expert) qui vous permet de récupérer le next-state et la récompense immédiate en lien avec le problème de grille à 6 états. La fonction "main" qui permet de lancer l'algorithme de Q-learning. Vous devez compléter le code de l'algorithme.

Dans l'état, ce code est fonctionnel, mais n'apprend rien. Une action aléatoire est choisie à chaque étape sans apprentissage.

Question 1 : prendre en mail le projet et le code fourni

Découvrir l'ensemble de fichiers et le code fourni. Lancer le code. Une boucle d'affichage est lancée. Décrire ce qu'elle affiche. Quelle est sa condition d'arrêt ?

Question 2 : implementer Q-learning

Compléter la fonction playEpisde de Q-learning en respectant l'algorithme donné dans le support du cours.

Question 3 : Tester Q-learning

Tester qlearning sur le problème de Grid6. Est-ce que votre agent a appris la meilleure politique ? Expliquez.

Question 4 : Tester différents paramètres

Tester qlearning avec différentes valeurs d'epsilon. Expliquez les résultats.

Question 5 : Joueur de morpion I

Créer et implémenter morpion1.py. Votre agent Q-learning représente le joueur X. Le joueur O est un joueur aléatoire (joue une action possible aléatoire avec une distribution uniforme) et ses actions sont implémentées via la fonction next_state. Une récompense de 1 est donnée si le joueur X réalise une action finale permettant de gagner le match, sinon toutes autres actions valent 0.

Exemple : fin d'épisode avec récompense =1

X|X|X
O| |
O| |O

Question 6 bonus : Tracer et comparer l'apprentissage

Pour visualiser l'évolution de l'apprentissage, vous pouvez afficher en plot le nombre d'actions réalisées par épisode.

Il est aussi intéressant dans certains problèmes de tracer en plot la somme de récompense par épisode. Est-ce le cas pour gird6 ? Est-ce le cas pour le joueur de morpion ?

Question 7 bonus : Joueur de morpion II

Créer un agent Q-learning pour le joueur O qui apprendra face à votre joueur morpion X.