EXERCICE 4

Mise à jour d'une Q-table avec Q-Learning

rhoumahaythem

Avril 2025

Rappel des équations fondamentales du Q-learning

1. Valeur cible de l'action

$$Q_{\text{cible}}(S, a) = R(s') + \gamma \max_{a'} Q(s', a')$$

2. Erreur de Temporal Difference (TD Error)

$$TD = Q_{\text{cible}}(S, a) - Q(S, a)$$

3. Mise à jour de la Q-valeur

$$Q(S, a) \leftarrow Q(S, a) + \alpha \cdot TD$$

—

Partie 4-1 : Mise à jour de Q(S1, droite)

Q-table avant mise à jour :

Q	gauche	droite	haut	bas
S1	-0.5	1.0	2.1	1.3
S2	0.5	0.75	-0.5	1.5
S3	-1.2	1.2	0.7	1.7

Données : État actuel : S_1 ; Action : droite ; État suivant : S_2 Récompense : -1 ; Q(S1, droite) initial : 1.0 ; $\alpha = 0.1$; $\gamma = 0.1$

Étape 1 – Calcul de la valeur cible : À partir de l'état S_2, on regarde les 4 actions possibles :

- Q(S2, gauche) = 0.5
- Q(S2, droite) = 0.75
- Q(S2, haut) = -0.5
- Q(S2, bas) = 1.5 ⇐ maximum

Donc :

$$Q_{\text{cible}} = -1 + 0.1 \cdot 1.5 = -0.85$$

Étape 2 – Calcul du TD Error :

$$TD = -0.85 - 1 = -1.85$$

Étape 3 – Mise à jour de la valeur :

$$Q_{\text{nouveau}}(S_1, \text{droite}) = 1 + 0.1 \cdot (-1.85) = 0.815$$

Q-table après mise à jour (valeur mise à jour en gras) :

Q	gauche	droite	haut	bas
S1	-0.5	*0.815*	2.1	1.3
S2	0.5	0.75	-0.5	1.5
S3	-1.2	1.2	0.7	1.7

—

Partie 4-2 : Mise à jour de Q(S2, droite)

Q-table avant mise à jour :

Q	gauche	droite	haut	bas
S1	-0.5	0.815	2.1	1.3
S2	0.5	0.75	-0.5	1.5
S3	-1.2	1.2	0.7	1.7

Étape 1 – Calcul de la valeur cible depuis S3 :

- Q(S3, gauche) = -1.2
- Q(S3, droite) = 1.2
- Q(S3, haut) = 0.7
- Q(S3, bas) = 1.7 $\Leftarrow$ maximum

$$Q_{\text{cible}} = -1 + 0.1 \cdot 1.7 = -0.83$$

Étape 2 – TD Error :

$$TD = -0.83 - 0.75 = -1.58$$

Étape 3 – Mise à jour :

$$Q(S_2, \text{droite}) = 0.75 + 0.1 \cdot (-1.58) = 0.592$$

Q-table après mise à jour (valeur mise à jour en gras) :

Q	gauche	droite	haut	bas
S1	-0.5	0.815	2.1	1.3
S2	0.5	*0.592*	-0.5	1.5
S3	-1.2	1.2	0.7	1.7

—

Partie 4-3 : Mise à jour avec environnement stochastique (glissement 20%)

Hypothèse : 80% de probabilité d'aller dans la direction choisie, 20% d'aller ailleurs.

Q(S1, droite) :

$$Q_{\text{cible}} = -1 + 0.1 \cdot (0.8 \cdot 1.5 + 0.2 \cdot 0.25) = -1 + 0.1 \cdot 1.25 = -0.875$$

$$TD = -0.875 - 1 = -1.875 \quad \Rightarrow \quad Q(S_1, \text{droite}) = 1 - 0.1875 = 0.8125$$

Q(S2, droite) :

$$Q_{\text{cible}} = -1 + 0.1 \cdot (0.8 \cdot 1.7 + 0.2 \cdot 0.233) = -0.8594$$

$$TD = -0.8594 - 0.75 = -1.6094 \quad \Rightarrow \quad Q(S_2, \text{droite}) = 0.75 - 0.1609 = 0.5891$$

Q-table avec environnement stochastique :

Q	gauche	droite	haut	bas
S1	-0.5	*0.8125*	2.1	1.3
S2	0.5	*0.5891*	-0.5	1.5
S3	-1.2	1.2	0.7	1.7