Intelligence Artificielle et Analyse de données

TP3 : Apprentissage par renforcement

L'objectif de ce TP est d'apprendre une loi de commande stabilisante pour un pendule inverse.

Une évaluation, sous forme de présentation orale 'par groupe de 2), aura lieu lors de séance suivante.

1. Comprendre le modèle :
   Modifier la cellule *tester l'environnement* pour comprendre : quel est l'état du système, la ou les actions, les fonctions principales de l'environnement (*reset, step*, …).

2. Discrétisation du modèle :

   L'algorithme de Q-Learning tabulaire, est adapté à un état discret.

   La cellule *fonction de discrétisation* permet de discrétiser l'état.

   A quoi correspond chacune des 4 valeurs de *lower_bounds*, et *upper_bounds* ?

   Comprendre comment choisir la finesse de discrétisation (i.e. le nombre d'état discret).

   Quels avantages / inconvénients de prendre une discrétisation très fine ?

3. Algorithme de Q-learning (tabulaire) :

   Compléter la cellule *apprentissage : version tabulaire :*

   - Implémenter la version ϵ-greedy pour le choix des actions.
   - Compléter pour afficher la somme des récompenses obtenues à chaque épisode.
   - Implémenter des visualisations qui peuvent aider à voir l'évolution de l'apprentissage / les fonctions de valeurs obtenues.
   - Implémenter la cellule *évaluation* pour évaluer la performance de votre apprentissage (tester votre commande sans exploration).

4. Algorithme de Q-learning (reseaux de neurons) :
   Implémenter l'algorithme avec un réseau de neurones MLP.