

# = APPRENTISSAGE PAR RENFORCEMENT =

## = INTRODUCTION AU COURS =

### ► Principe distinct du Machine Learning

- Pas d'apprentissage « supervisé » (classification ou régressions avec utilisation de labels)
- Pas d'apprentissage « non supervisé » (sans utilisations de labels)

### ► Principe général

- Définit un **agent** (par exemple un robot) et l'**environnement** dans lequel cet agent évolue
- L'agent effectue certaines **actions** dans l'environnement (par exemple les mouvements du robot)
- Ces actions sont **récompensées** (en bien ou en mal) en fonction de certains critères
- Le but est que l'agent gagne un **maximum de récompenses**

### ► Équations de Bellman

- Le comportement optimum est défini par les **équations de Bellman**

# = APPRENTISSAGE PAR RENFORCEMENT =

## = INTRODUCTION AU COURS =

### Intelligence Artificielle



Algorithmes qui imitent l'intelligence humaine, afin de résoudre des problèmes.

### Machine Learning

Algorithmes qui utilisent des jeux de données pour apprendre à prendre les bonnes décisions.

Apprentissage supervisé

Apprentissage non supervisé

Apprentissage par  
renforcement

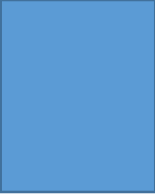


Deep  
Reinforcement  
Learning

### Deep Learning

Réseaux de neurones profonds  
capables d'approximer n'importe  
quelle fonction.

Utilisé pour la détection d'objets, la  
reconnaissance vocale, l'analyse  
d'image ...



# = APPRENTISSAGE PAR RENFORCEMENT =

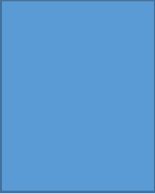
## = INTRODUCTION AU COURS =

### ► Ce que nous allons traiter

- Les **concepts** utilisés dans l'apprentissage par renforcement
- Comprendre le principe de fonctionnement des **processus décisionnels de Markov**
- Mettre en œuvre et résoudre les **équations de Bellman**
- Méthode **basée sur un modèle parfaitement connu de l'environnement** :
  - Programmation dynamique (**Dynamic Programming – DP**) : Modèle parfaitement connu
- Méthodes **non basées sur un modèle parfait de l'environnement**:
  - **Monte Carlo** et Différence Temporelle (**Temporal Difference – TD**) : Modèle partiellement connu
  - **Q-Learning** (stratégie suivie n'intervient pas pour le calcul des récompenses – « **Off-Policy** »)
  - **SARSA** (utilise la stratégie – « **On-Policy** »)

### ► Apprentissage en ligne (**ONLINE RL**)

- L'agent interagit avec son environnement, effectue des actions, observe comment ces actions influencent l'environnement et découvre des stratégies optimales



# = APPRENTISSAGE PAR RENFORCEMENT =

## = INTRODUCTION AU COURS =

### ► Apprentissage hors ligne (OFFLINE RL) – Domaine de la recherche

- L'apprentissage ne se base que sur des données statiques qui ont été par exemple collectées en suivant certaines stratégies : **méthodes de type Batch**
- L'agent peut apprendre les stratégies qui ont été utilisées pour collecter les données et les reproduire
- L'agent peut tenter d'apprendre des stratégies plus optimisées (plus difficile)
- Méthodes OPE (Off-Policy Evaluation) ou HCOPE (High-Confidence Off-Policy Evaluation)
- Ces méthodes peuvent être basées sur un modèle ou non :
  - On utilise les données collectées afin d'estimer le modèle
  - On utilise la méthode Q-Learning comme si les données étaient collectées en temps réel