

Apprentissage automatique

Amira Barhoumi

amira.barhoumi@univ-grenoble-alpes.fr

Introduction

- Dans le **monde réel**, un **humain** est un être qui peut **apprendre** à partir des **expériences vécues** grâce à sa **capacité d'apprentissage**.
- Les ordinateurs et les machines fonctionnent selon des instructions prédéfinies.
- Une machine peut apprendre à partir des données comme le fait un être humain?



Introduction



C

Russula Aurea



T

Amanita Muscaria



C

Boletus Edulis



M

Amanita Virosa



T

Cortinarius Bolaris



M

Gyromitra Esculenta



C

Cantharellus Cibarius



N

Hygrocybe Coccinea

⋮

Apprendre à classifier

Dans ce problème, l'agent (**apprenant**) commence par observer une série de champignons classés comestibles (C), non comestibles (N), toxiques (T) ou mortels (M)

Introduction



C



Amanita Phalloides



T



Amanita Pantherina



M



Tricholoma Aurantium

Apprendre à classifier

Après la phase d'entraînement, la performance de l'agent est mesurée par sa capacité à classifier correctement de nouveaux champignons

Apprendre à classifier

Un problème d'apprentissage fait intervenir trois composants essentiels:

- la tâche à résoudre
- les exemples (le corpus, le jeu de données)
- la mesure de performance (nombre d'erreurs, taux de bonne classification, etc)

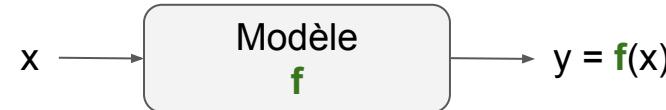
Introduction

- Différentes applications de tous les jours :
 - Google translate (traduction automatique des langues)
 - Assistant Google, Alexa, Siri (reconnaissance vocale)
 - Chatbot (assistant personnel virtuel)
 - Recommandation de produits sur les réseaux sociaux, youtube, etc
 - Filtrage de courriers indésirables et des logiciels malveillants
 - Marquage automatique des amis sur Facebook (reconnaissance d'images)
 - Voiture autonome (Tesla)

- Apprentissage automatique
 - Champ d'étude de l'intelligence artificielle (IA)
 - Développement d'algorithmes
 - Donner la capacité aux ordinateurs d'*apprendre* à partir des données
 - Construire des modèles
 - Améliorer les performances à résoudre des tâches
 - Minimiser l'erreur statistique moyenne
 - Apprentissage statistique
 - *Machine learning*

Définition

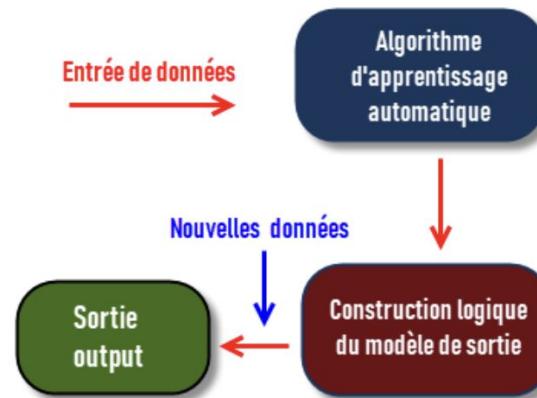
- **Apprentissage automatique** fait référence au **développement**, à l'**analyse** et à l'**implémentation** de méthodes qui permettent à une **machine** d'évoluer grâce à un processus d'apprentissage, et ainsi de remplir des tâches qu'il est difficile ou impossible de remplir par des moyens basiques
- “*L'apprentissage automatique permet à une machine d'apprendre automatiquement à partir des données, d'améliorer les performances à partir d'expériences et de fournir des prédictions sans être explicitement programmée.*” [Arthur Samuel, 1959]
- Formellement : déterminer une fonction mathématique **f** où $y = f(x)$ représente le résultat de **f** pour une observation **x** donnée.



Fonctionnement de l'apprentissage automatique

- Étapes d'apprentissage automatique :

- Préparation des données
- Construction du modèle
- Évaluation et/ou analyse des résultats



Types d'apprentissage automatique

- Les algorithmes d'apprentissage automatique peuvent se catégoriser selon le type en :
 - Apprentissage supervisé
 - Apprentissage non supervisé
 - Apprentissage semi-supervisé
 - Apprentissage par renforcement

Apprentissage supervisé

- Jeu de données/corpus décrivant les données alliant un **ensemble de descripteurs** et une **classe**
- Recherche d'une **fonction de classification** permettant d'**induire la classe** en fonction de l'ensemble de descripteurs

	var 1	...	var k	classe
obj				oui
...				...
obj				non

Apprentissage supervisé

- Jeu de données/corpus décrivant les données alliant un **ensemble de descripteurs** et une **classe**
- Recherche d'une **fondation de classification** permettant d'**induire la classe** en fonction de l'ensemble de descripteurs

	var 1	...	var k	classe
obj				oui
...				...
obj				non

	var1	...	var k	classe
new				?

Apprentissage non supervisé

- Jeu de données/corpus décrivant les données décrites par un ensemble de descripteurs **sans classes connues à priori**
- Recherche d'une **division** de ces données en **catégories**

	var 1	...	var k
obj			
...			
obj			

Identification des catégories

Apprentissage non supervisé

- Jeu de données/corpus décrivant les données décrites par un ensemble de descripteurs **sans classes connues à priori**
- Recherche d'une **division** de ces données en **catégories**

	var 1	...	var k
obj			
...			
obj			

Identification des catégories

	var1	...	var k
new			

Quelle catégorie?

Apprentissage semi-supervisé

- Idée: L'étiquetage des données est souvent coûteux, alors qu'en général il est peu coûteux d'obtenir les données sans label
- Apprentissage semi-supervisé : apprendre avec peu de données étiquetées et beaucoup de données non-étiquetées

	var 1	...	var k	classe
obj				oui
...				...
obj				non

	att 1	...	att m
obj			
...			
obj			

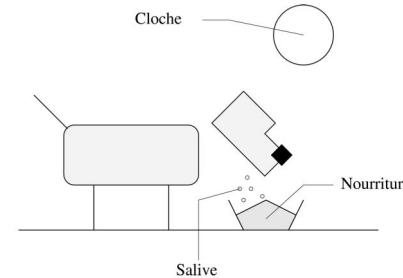
Apprentissage semi-supervisé

- Méthodes d'apprentissage semi-supervisé:
 - **Auto-apprentissage** (self-training) consiste à :
 - entraîner un classifieur avec les données étiquetées (DL).
 - utiliser le classifieur pour étiqueter les données incomplètes (DU)
 - ajouter les données étiquetées avec un score de confiance élevé au corpus DL
 - ré-entraîner le classifieur sur le nouveau corpus DL
 - répéter jusqu'à satisfaire un critère d'arrêt

Apprentissage semi-supervisé

- Méthodes d'apprentissage semi-supervisé:
 - **Co-apprentissage (*co-training*)** consiste à entraîner 2 classifieurs sur 2 projections indépendantes d'un même espace de données de manière à étiqueter de manière identique la même donnée
- Principe:*
- diviser l'ensemble d'attributs en 2 sous-ensembles indépendants
 - entraîner 2 classifieurs avec le corpus DL en utilisant pour chacun un sous-ensemble d'attributs différent
 - étiqueter DU avec ces classifieurs
 - ajouter les données étiquetées avec une bonne confiance au corpus DL
 - ré-entraîner sur le nouveau DL
 - lorsque l'apprentissage est terminé, les 2 classifieurs sont combinés

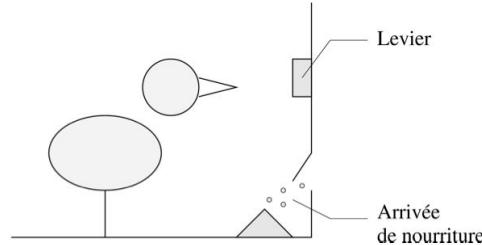
- **Introduction**



1- Chien de Pavlov :

- Présenter de la nourriture à un chien provoque sa salive
- Expérience répétée: juste après avoir présenté la nourriture, on fait sonner une cloche
- le chien finit par saliver au simple son de la cloche

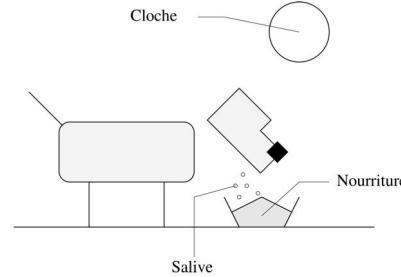
- **Introduction**



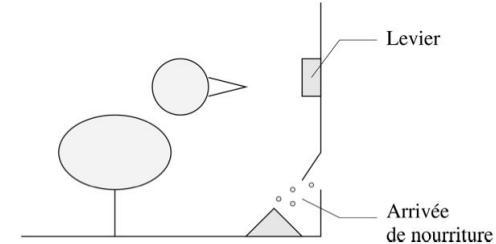
2- boîte de Skinner :

- Appuyer sur le levier permet de délivrer de la nourriture
- Le pigeon apprend par essais-erreurs à appuyer sur le levier pour de la nourriture soit délivrée
- On peut aussi compliquer le dispositif pour conditionner la délivrance de nourriture par signal sonore ou lumineux

Apprentissage par renforcement



Chien de Pavlov

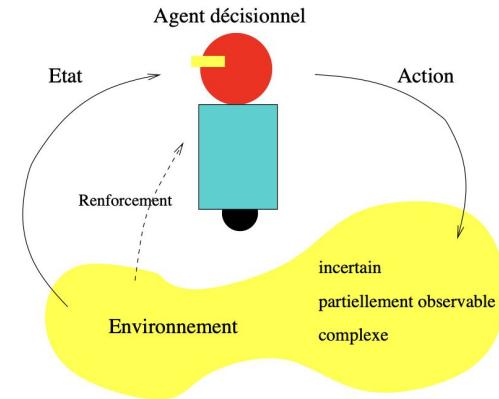


Boîte de Skinner

- **Conditionnement** (classique ou par essais/erreurs) du comportement de l'animal
- Idée de l'apprentissage par renforcement :
 - reproduction artificielle du conditionnement de l'animal
 - comment faire apprendre un comportement à une machine en lui introduisant des récompenses

Apprentissage par renforcement

- Apprentissage par renforcement : apprendre à un agent à atteindre un but à partir de son expérience.
- Objectifs de l'apprentissage par renforcement
 - Acquérir automatiquement des compétences pour la prise de décisions (**actions**) en milieu complexe et incertain.
 - Apprendre par l'expérience une stratégie comportementale (**politique**) en fonction des échecs ou succès constatés (**renforcements** ou **récompenses**).
- Exemples: robotique mobile autonome, gestion de portefeuille, recherche opérationnelle, jeux, ...



Apprentissage par renforcement

- Apprentissage par renforcement se concentre sur l'apprentissage de la manière de prendre des séquences de décisions afin de maximiser un signal de récompense.
 - **Agent** : apprend à interagir avec un environnement pour atteindre un objectif.
 - **Environnement** : fournit un retour d'information à l'agent sous forme de récompenses.
 - **Objectif** : apprendre une politique qui associe des états à des actions afin de maximiser la récompense cumulative attendue.