

Les principaux types d'apprentissages

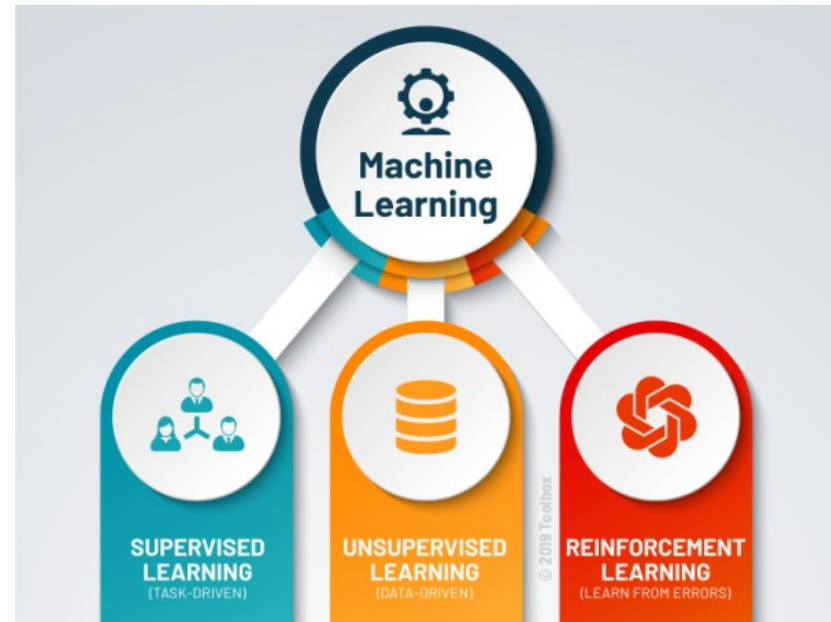
Présente par Nga

Sommaire

- Introduction de Machine Learning
- Apprentissage supervisé
- Apprentissage non-supervisé





Introduction de Machine Learning

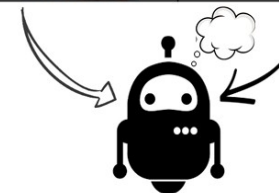
- une catégorie d'algorithme qui permet aux applications logicielles de prédire plus précisément les résultats sans être explicitement programmées
- Le principe de base
 - créer des algorithmes capables de recevoir des données d'entrée et
 - d'utiliser une analyse statistique
 - BUT : prédire une sortie tout en les mettant à jour à mesure que de nouvelles données deviennent disponibles.



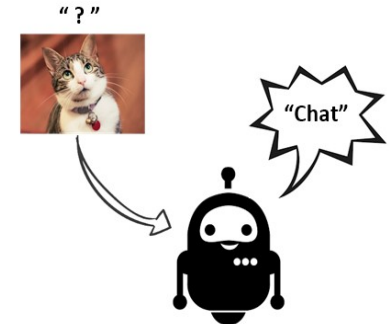
Apprentissage supervisé.1

- consiste en : $Y = f(X)$
 - des variables d'entrée (x)
 - une variable de sortie (Y).
- apprendre à faire une certaine tâche en étudiant des exemples de cette tâche :
 - Relation $f : X \rightarrow Y$
 - f relie x à y en ayant analysé des millions d'exemples d'associations
- Le processus d'algorithme ???

x	y
	"Chien"
	"Chien"
	"Chat"
	"Chien"



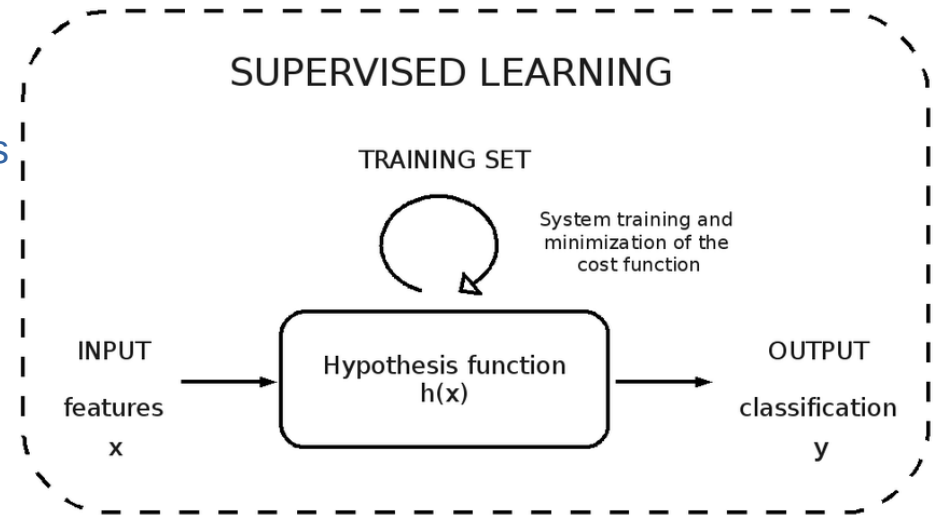
Apprentissage Supervisé



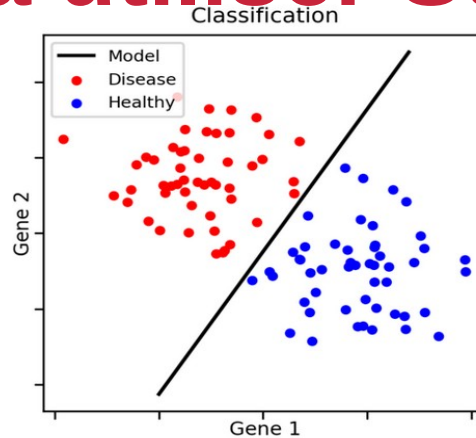
Utilisation finale

Processus d'algorithme Supervisé

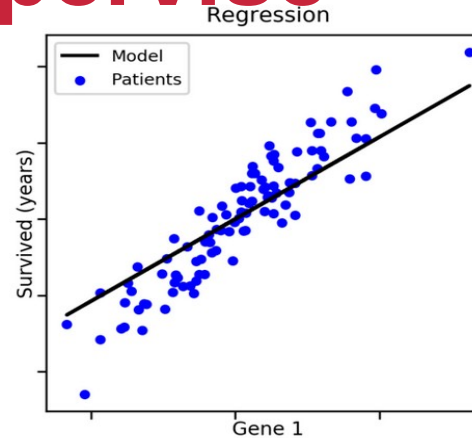
- 1) tiré de l'ensemble de données de formation (training set)
- 2) l'algorithme effectue des prédictions itératives sur les données d'apprentissage et est corrigé par l'enseignant = le modèle
- 3) L'apprentissage s'arrête lorsque l'algorithme atteint un niveau de performance acceptable.



Contexte à utiliser Supervisé



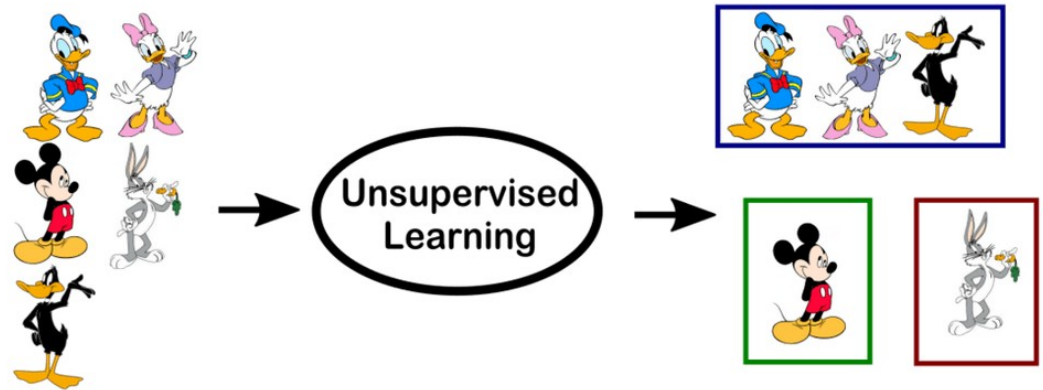
- **Classification: variable de sortie est une catégorie**
- En finance/ secteur bancaire : détection de la fraude par carte de crédit
- Détection de courrier électronique indésirable
- Domaine du marketing : l'analyse du sentiment de texte
- En médecine : prédire si un patient a une maladie ou non.



- **Régression: variable de sortie est une valeur réelle**
- Prédire le prix de l'immobilier
- Prédire le cours de bourse

Apprentissage non supervisé.1

- ne disposer que de données d'entrée (X) et pas de variables de sortie correspondantes.
- l'algorithme doit opérer à partir d'exemples non annotés.
- Objectif : de modéliser la structure ou la distribution sous-jacente dans les données **afin d'en apprendre davantage sur les données.**
- Le résultat dépend beaucoup de la modélisation du problème et de la distance ...
- catégories d'algorithmes:
 - Regroupement
 - association



Clustering/Regroupement

- séparer ou à diviser un ensemble de données en un certain nombre de groupes, de sorte que les ensembles de données appartenant aux mêmes groupes se ressemblent davantage que ceux d'autres groupes.
- L'objectif est de séparer les groupes ayant des traits similaires et de les assigner en grappes.

algorithme très souvent
utilisé dans le clustering
est le K-moyen

- Eloigner des échantillons différents



- Qu'est-ce qu'un bon clustering ?
- Problème mal posé
- L'objectif dépend du problème considéré

Association

- à découvrir des relations intéressantes entre des variables dans de grandes bases de données ; trier et regrouper les données qui peuvent être liées grâce à certaines caractéristiques.
- Il découvre la probabilité de cooccurrence d'éléments dans une collection sans qu'il s'agisse néanmoins d'objets identiques.
- L'algorithme très souvent utilisé dans l'association est l'algorithme A-priori**

Milk, eggs, sugar,
bread



Customer1

Milk, eggs, cereal,
bread



Customer2

Eggs, sugar



Customer3

- Qu'est-ce que le client achète ?
- Quels les produits a-il-acheté l'ensemble ?

Aim: Find **associations** and **correlations** between the different items that customers place in their shopping basket

Semi-supervisé

- Avoir une grande quantité de données d'entrée (X)
 - seules certaines données sont étiquetées (Y)
 - **Objectif** : utiliser les données non-étiquetées (non-annotées) pour compléter l'apprentissage supervisé.
- ▶ Apprentissage semi-supervisé
- Avoir une grande quantité de données d'entrée (X)
 - seules certaines données sont étiquetées (Y)
 - **Objectif** : utiliser les données non-étiquetées (non-annotées) pour compléter l'apprentissage supervisé.
 - **Exemple** : une archive de photos dans laquelle seules certaines images sont étiquetées (chien, chat, personne, par exemple) et la plupart ne le sont pas.

Apprentissage automatique renforcement

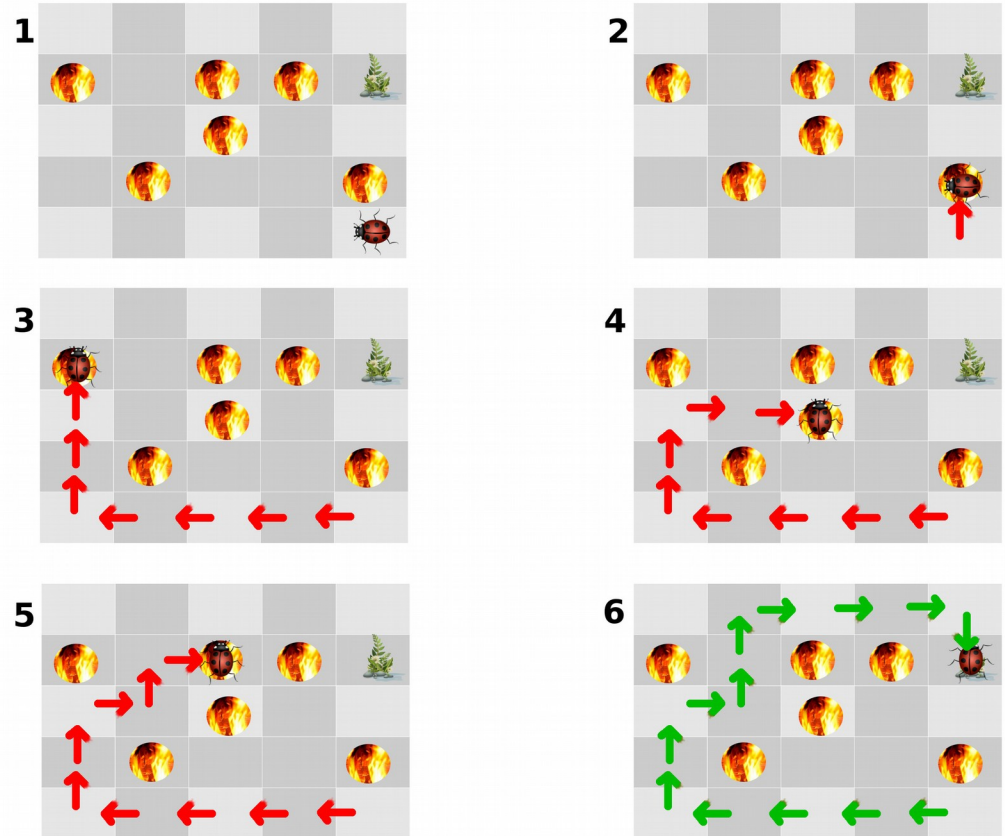
- Un système autonome qui s'améliore en fonction de la combinaison de données étiquetées et d'interactions avec les données entrantes (tout au long de séquences contiguës d'essais et d'échecs)
- s'adapte aux instances d'informations disponibles limitées ou incohérentes
- La mécanique est simple
 - l'action a lieu, les conséquences sont observées
 - l'action suivante considère les résultats de la première action
- Est utilisé pour amplifier et ajuster le traitement du langage naturel (PNL) et la génération de dialogues pour les chatbots afin de:
 - imiter le style d'un message d'entrée
 - élaborer des réponses plus engageantes et informatives
 - trouver des réponses pertinentes en fonction de la réaction de l'utilisateur

Apprentissage automatique renforcement

L'apprentissage par renforcement ne requiert pas un ensemble initial de données, mais plutôt un environnement (ou un contexte).

Les données sont collectées par l'algorithme d'apprentissage en explorant l'environnement.

L'algorithme tente d'optimiser un processus de décision vers un objectif.



Type d'algorithmes ML

Top 3 types of Machine Learning algorithms

```
graph TD; A[Top 3 types of Machine Learning algorithms] --> B[Supervised Learning]; A --> C[Unsupervised Learning]; A --> D[Reinforcement Learning];
```

Supervised Learning

- Well defined goals
- Reverse engineering
- Example- Fraud/Non-fraud transactions, inventory management

Unsupervised Learning

- Outcome is based only on inputs
- Outcome- typically Clustering or segmentation

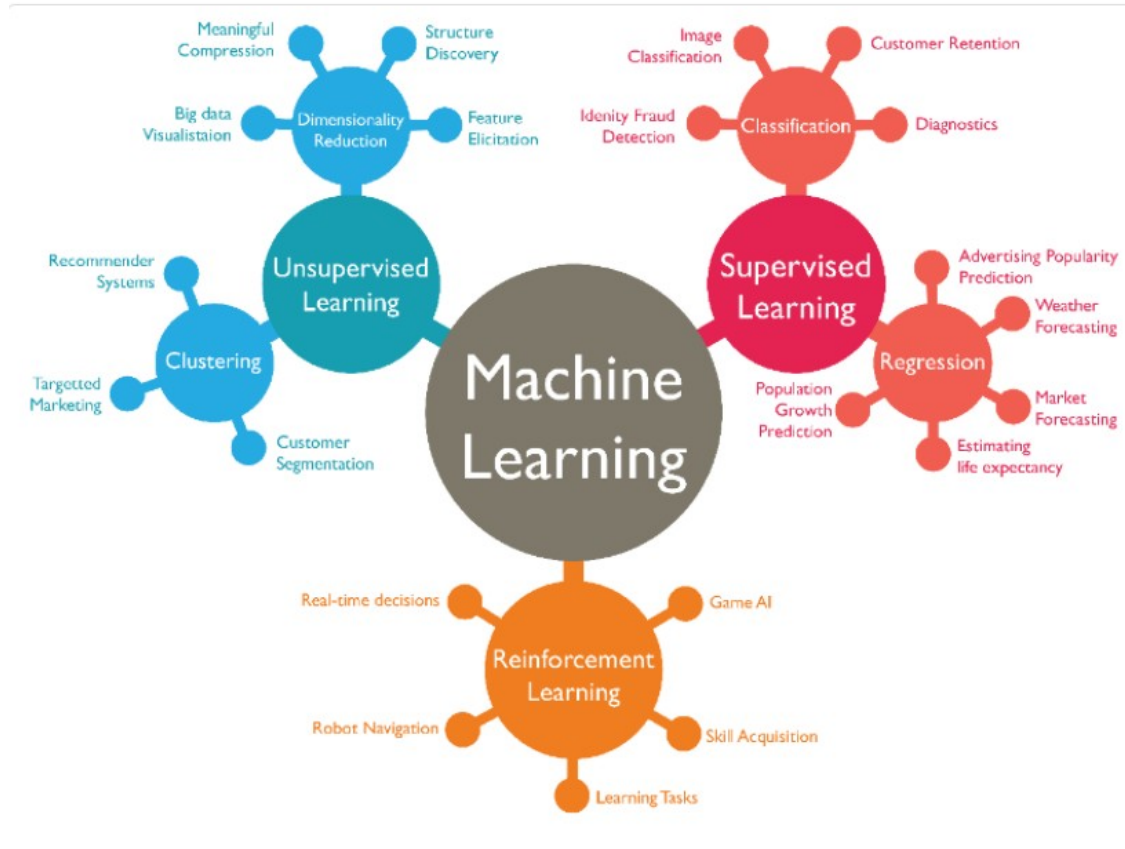
Reinforcement Learning

- Start state and end state are defined
- The agent discovers the path and the relationships on its own

Supervisé vs Non-supervisé

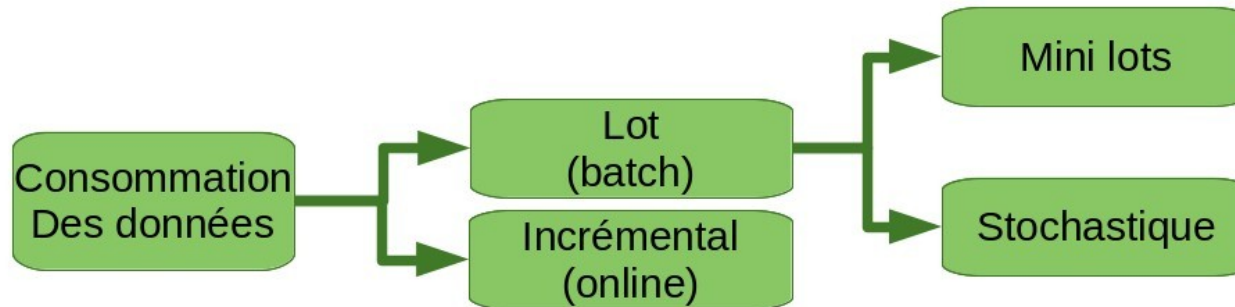
	Apprentissage supervisé	Apprentissage non supervisé
Données d'entrée	Données connues en entrée	Données inconnues en entrée
Complexité informatique	Complexe	Moins complexe
Domaines d'activités	Classification et régression	Exploitation de règles de <u>clustering</u> et d'association
Précision	Produit des résultats précis	Génère des résultats modérés

Cas d'utilisation de ML



Taxonomie par modes d'approvisionnement en données d'apprentissage

- les techniques d'apprentissage automatique ne devraient pas garder en mémoire l'ensemble des données sur lesquelles elles ont été entraînées
- nombreuses techniques d'apprentissage ne pourront pas, sur de nouvelles données, ajuster davantage une fonction de prédiction déjà déterminée, tout en la maintenant cohérente avec son apprentissage précédent



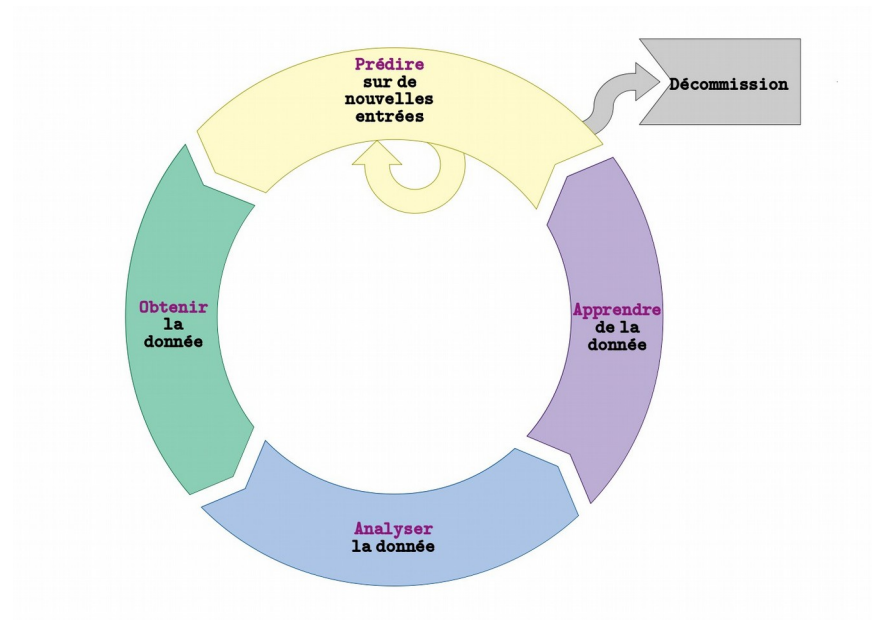
Apprentissage par lots

- Tous les exemples doivent être fournis pendant la phase d'apprentissage
- La “fonction de prédiction” résultant de l'apprentissage est ensuite utilisée en production et l'apprentissage est définitivement terminé pour cette fonction de prédiction.
- Les méthodes out-of-core :
 - Peuvent aborder le problème au niveau de l'accès aux données et de leur gestion, en permettant par exemple de charger en mémoire uniquement le sous-ensemble de données nécessaires à un calcul à un instant t .



Apprentissage incrémental (en-ligne)

- peut être alimentée progressivement avec de nouveaux éléments et modifie ses représentations en conséquence, même lors de son utilisation en production.
- utilise la fonction de prédiction courante comme input supplémentaire, en plus des données.



Conclusion

- Plusieurs axes sont possibles pour réaliser une taxonomie des algorithmes d'apprentissage automatique (machine learning).
 - les apprentissages supervisé
 - Semi-supervisé
 - non supervisé
 - par renforcement.
- Une possibilité consiste à différencier les algorithmes d'apprentissage automatique par la manière dont ils consomment les données d'apprentissage
 - méthodes d'apprentissage par lots
 - méthodes d'apprentissage incrémental.