

# Machines Learning Introduction

En 2017 : apprentissage automatique avec les réseaux de neurones et le deep learning

L'Apprentissage Automatique Ou *Machine Learning*

Définition : L'apprentissage automatique peut-être vu comme l'ensemble des techniques permettant à une machine d'apprendre à réaliser une tâche sans avoir à la programmer explicitement pour cela.

# Machines Learning Introduction

Le type de tâches traitées consiste généralement en des problèmes de classification de données:

## Exemples

- ❑ Voici 50 photos de ma fille, voici maintenant toutes les photos de mon album, retrouve celles où se trouve ma fille
- ❑ Voici 10000 personnes avec leurs caractéristiques (âge, localisation, taille, profession...), rassemble-les en 12 groupes cohérents partageant les mêmes points communs

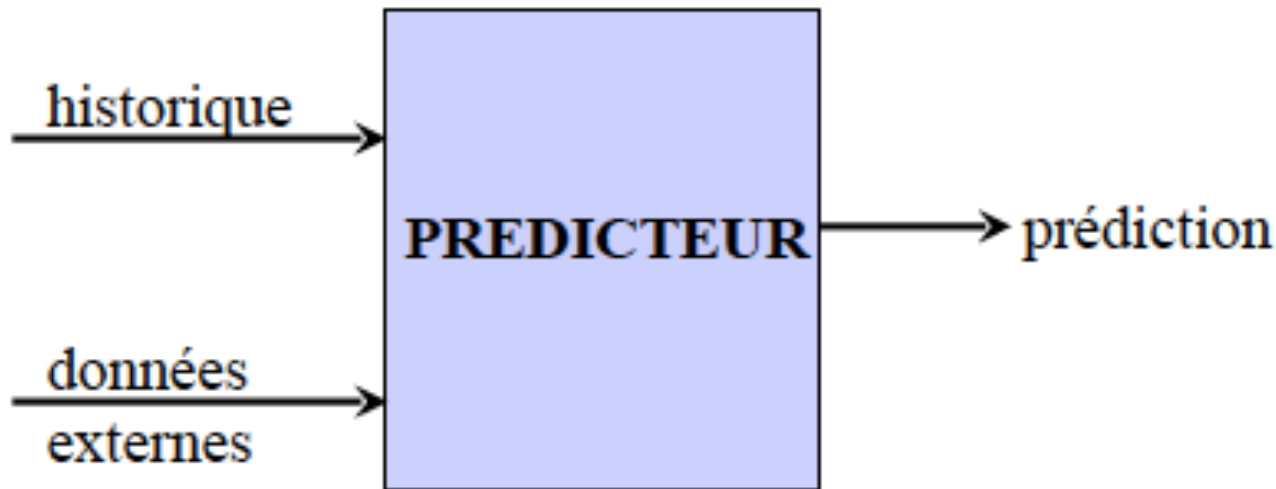
# Machines Learning Introduction

## Exemples suite

- ❑ Voici les paramètres de vol des avions de ma flotte avec les pannes survenues à chaque vol, quand il s'en produit. Maintenant voici les derniers vols réalisés par cet avion, indique-moi quand arrivera la prochaine panne et sur quel élément elle se produira.
- ❑ Voici les livres écrits par Victor Hugo, voici une nouvelle dont nous recherchons l'auteur, est-elle de lui ?
- ❑ Voici des mesures scientifiques réalisées sur un échantillon de données, elles semblent répondre à une loi Mathématique, essaye de l'approximer par des fonctions Mathématiques puis prédit les valeurs des prochaines données

# Machines Learning Introduction

- **Agent « prédicteur »**



- **Performance visée : minimiser erreur de prédiction**
- **Moyen mis en œuvre :**  
utiliser des données expérimentales pour trouver un modèle  
 $\text{prédiction} = f(\text{historique}, \text{données externes})$  le plus correct possible

# Machines Learning Introduction

## 3 Types d'apprentissage :

- ❑ Supervisé : production automatique de règles à partir d'une base de donnée d'apprentissage contenant des exemples.
- ❑ Non-supervisé : division d'un groupe hétérogène de données, en sous-groupes de manière que les données considérées comme les plus similaires soient associées au sein d'un groupe homogène et qu'au contraire les données considérées comme différentes se retrouvent dans d'autres groupes distincts ; l'objectif étant de permettre une extraction de connaissance organisée à partir de ces données
- ❑ Par renforcement : considérer un agent autonome, plongé au sein d'un environnement, et qui doit prendre des décisions en fonction de son état courant. En retour, l'environnement procure à l'agent une récompense, qui peut être positive ou négative..

# Machines Learning Introduction

Chaque type d'apprentissage peut s'appuyer sur différents algorithmes :

- régression linéaire
- régression logistique
- arbre de décision
- forêts d'arbres/arbes aléatoires
- machines à vecteur de support (SVM)
- classification naïve bayésienne
- Plus proches voisins
- Q-Learning
- Réseaux de Neurones
- ...

# Machines Learning Introduction

Apprentissage Supervisé : permet de répondre à des problématiques de classification et de régression.

L'idée consiste à associer un label à des données sur lesquelles vous possédez des mesures :

- ❑ Si les labels sont discrets (des libellés ou valeurs finies) on parlera de classification.
- ❑ Si au contraire les labels sont continus (comme l'ensemble des nombres réels), on parlera de régression.

# Machines Learning Introduction

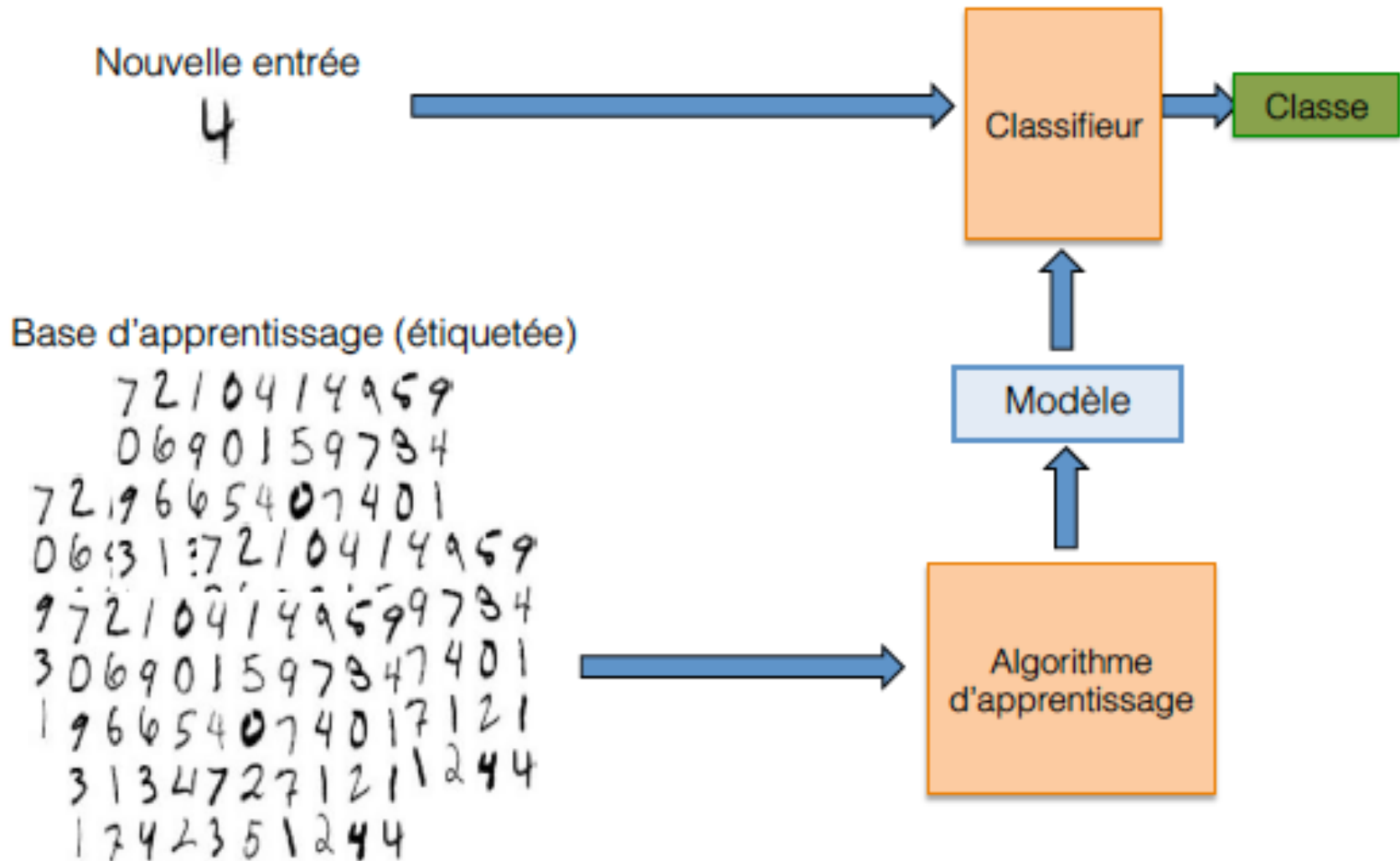
## *Classification*

- La classification consiste à donner des étiquettes à ses données:
- Vous disposez d'un ensemble de données connues que vous avez déjà classé (photos, plantes, individus...)
- Vous souhaitez, à partir de cette première classification, dite connaissance, classer de nouveaux éléments.

Ce type de classification permet de répondre à de nombreux problèmes d'identification : reconnaissance de plantes, de personnes, de produits, reconstitution de valeurs manquantes (en remplacement d'une interpolation), etc...Il peut utiliser différents types d'algorithmes, comme *les plus proches voisins, les Réseaux de Neurones, les machines à vecteurs de supports, les arbres décisionnels, ...*



# Machines Learning Introduction



# Machines Learning Introduction

L'apprentissage non supervisé répond au même besoin de classification de données.

Mais contrairement à l'apprentissage supervisé, vous ne possédez pas de données déjà classées/connues servant de base à la prédiction.

On parle alors de *clustering* c'est à dire de *regroupement*.

# Machines Learning Introduction

## Clustering

Avec ce type d'apprentissage il n'y a qu'une phase de prédiction. Vous alimentez l'algorithme de toutes vos données et lui demandez de les répartir en N groupes.

L'algorithme tentera alors de créer des groupes pour lesquels les paramètres de chaque donnée sont les plus similaires.

# Machines Learning Introduction

Quelques exemples d'applications:

- Classer les cultures d'une région : Vous disposez des métriques sur les parcelles agricoles d'une région (teneur en nitrates, phosphates, salinité, surface, haies, ...) et vous savez qu'il y a 12 cultures différentes.  
Vous allez demander à votre classifieur de créer 12 groupes en les classant selon des critères de *ressemblance*.
- *Vous possédez des statistiques sur une population (salaire, localité, âge, profession, nombre d'enfants), vous souhaitez les regrouper en différentes catégories.*

Les algorithmes des plus proches voisins, arbres de décision, les réseaux de neurones, la propagation par affinité sont généralement utilisés pour ce type de traitement.

# Machines Learning Introduction

## Apprentissage par Renforcement

Ce type d'apprentissage s'applique plus à des problèmes d'optimisations. L'idée étant de faire prendre des décisions à un système pour obtenir un résultat qui soit le meilleur possible.

Ce type d'algorithme est très inspiré d'études du comportement en biologie animale ou psychologie

Pour cela l'algorithme va appliquer des règles sur son environnement pour arriver au résultat attendu.

Il dispose de la faculté de mesurer l'impact de la règle sur l'environnement.

il peut donc se constituer une base de connaissances des gains réalisés par chaque action qui l'aidera à améliorer ses décisions et ainsi trouver les *meilleures manières d'atteindre son but*.

# Machines Learning Introduction

Les algorithmes ne garantissent pas forcément d'obtenir le meilleur résultat, mais de s'en approcher. Là où ils commencent à se complexifier c'est que l'obtention de petits gains immédiats ne doit pas empêcher de chercher des gains plus forts qui ne s'obtiennent qu'après plusieurs actions ayant entraîné une suite de pertes.

Différents algorithmes sont disponibles pour ce type d'apprentissage comme Q-Learning et Réseaux de Neurones

Pour l'IA il est surtout utile que le langage permette :

- La « récursivité »

Aujourd'hui quasiment tous les langages le permettent

- La programmation orientée objets

Là encore, la plupart des langages modernes couramment utilisés en disposent. Python offre toute la mécanique requise: héritage, surcharge, méthodes virtuelles ; même s'il est moins rigoureux que d'autres (pas de visibilité private/public/protected).

- Le paradigme fonctionnel

Python permet ce paradigme de programmation même s'il est moins "naturel" que dans d'autres langages comme Haskell ou Lisp.

Clairement, Python n'est pas vraiment au dessus des autres langages pour l'intelligence artificielle. Mais il s'y prête bien et sa syntaxe concise et facile permet d'y progresser très certainement plus aisément que dans d'autres langages. Il dispose aussi de quelques librairies spécialisées en IA qui lui permettent de s'initier à cette discipline.

En outre, concernant le machine learning, la problématique la plus difficile est souvent de disposer de bons jeux de données. Sur ce plan, Python est particulièrement bien outillé avec des librairies comme Numpy ou Pandas, par exemple.

Enfin, concernant le domaine de l'apprentissage automatique Python se distingue tout particulièrement en offrant une pléthore de librairies couvrant tous les types d'apprentissages