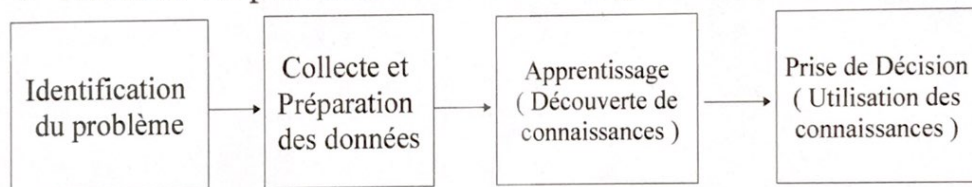


# Cours 1 - INTRODUCTION AU MACHINE LEARNING

## 1. Définition

### Machine Learning

- ♦ Extraction de connaissances à partir de données
  - connaissances : informations utiles pour la prise de décision
  - Données de grande quantité voire de très grande quantité (big data)
- ♦ Méthodes (techniques) automatisées
  - exécutées par une machine (machine learning)
  - à l'intersection de plusieurs disciplines
    - informatique (Intelligence artificielle, bases de données, programmation)
    - mathématiques (Probabilités, statistiques)
    - etc.
- ♦ Au cœur du processus de data mining



Le processus de data mining

- ♦ Domaines d'application
  - marketing
  - médecine,
  - biologie,
  - sciences sociales,
  - etc.

## 2. Les données utilisées

### Individu ( exemple ou instance)

Instance d'une entité faisant l'objet de l'étude

une personne, un objet, une organisation, etc.

Possède des caractéristiques correspondant à des variables

### Variable (attribut)

Caractérise un individu

Une variable  $X_j$  est une application

$$X_j : \Omega \rightarrow D_j$$

$$\omega_i \rightarrow X_j(\omega_i)$$

où

$\Omega$  : ensemble des individus

$D_j$  : domaine de la variable  $X_j$  (ensemble de ses valeurs possibles)

$X_j(\omega_i)$  : valeur de la variable  $X_j$  pour l'exemple  $\omega_i$ .

## Types de variables

### Variable qualitative

Domaine de la variable : ensemble fini d'éléments (les modalités de la variable)

Deux cas

Domaine ordonné : relation d'ordre entre les éléments

=> Variable qualitative ordinale

Exemple : variable Mention

Domaine = { passable, assez bien, bien, très bien }

Domaine non ordonné : pas de relation d'ordre entre les éléments => Variable qualitative nominale.

Exemple : variable Sexe

Domaine = { masculin, féminin }

### Variable quantitative

Domaine : ensemble infini de valeurs

Ensemble des nombres réels

=> variable quantitative continue

Ensemble des nombres entiers

=> variable quantitative discrète



## Tableau de données

Données regroupées dans un tableau individus – variables

	X1	X2	...	Xj	...	XP
$\omega_1$	$X1(\omega_1)$	$X2(\omega_1)$		$Xj(\omega_1)$		$XP(\omega_1)$
$\omega_i$	$X1(\omega_i)$	$X2(\omega_i)$	...	$Xj(\omega_i)$	...	$XP(\omega_i)$
$\omega_N$	$X1(\omega_N)$	$X2(\omega_N)$		$Xj(\omega_N)$		$XP(\omega_N)$

Individus  $\omega_1, \dots, \omega_N$

Variables  $X1, \dots, XP$  caractérisant ces individus

## 4. Les méthodes de machine Learning

### Deux catégories

méthodes supervisées

méthodes non supervisées

### Méthodes supervisées

Objectif

Trouver ou prédire la valeur d'une de variable (variable à expliquer) en fonction d'autres variables (variables explicatives)

Apprentissage supervisé

Tâches de prédiction

variable à expliquer qualitative => classification

variable à expliquer quantitative => régression

## Exemples de méthodes supervisées

méthode des K plus proches voisins

méthode naïve de Bayes

méthode des arbres de décision

méthode SVM (machines à vecteurs de support)

méthodes des réseaux de neurones

perceptron monocouche

perceptron multicouche (MLP)

réseaux de neurones convolutifs (CNN)

réseaux de neurones récurrents (RNN)

réseaux de neurones LSTM

Les réseaux de neurones convolutifs et les réseaux de neurones utilisent le Deep Learning (apprentissage profond), un cas particulier de Machine Learning avec des réseaux de neurones à plusieurs couches.

## Méthodes non supervisées

Pas de variable à expliquer

Objectif1 : Identifier des groupes homogènes d'individus (classes ou clusters)

=> classification non supervisée (ou clustering)

Exemples de méthodes de classification non supervisée

- méthode des KMoyennes (KMeans)
- Classification Ascendante Hiérarchique (CAH)

Objectif 2 : Trouver des associations entre les valeurs des variables

=> Extraction de règles d'association

Exemple de méthode d'extraction de règles d'association

- méthode A Priori



## Cours 2 – GENERALITES SUR LES METHODES D'APPRENTISSAGE SUPERVISEES

### Position du problème

Soit le tableau de données suivant

	X1	X2	...	Xj	...	XP	Y
$\omega_1$	$X1(\omega_1)$	$X2(\omega_1)$	...	$Xj(\omega_1)$	...	$XP(\omega_1)$	$Y(\omega_1)$
$\omega_2$	$X1(\omega_2)$	$X2(\omega_2)$	...	$Xj(\omega_2)$	...	$XP(\omega_2)$	$Y(\omega_2)$
...	...	...	...	...	...	...	...
$\omega_i$	$X1(\omega_i)$	$X2(\omega_i)$	...	$Xj(\omega_i)$	...	$XP(\omega_i)$	$Y(\omega_i)$
...	...	...	...	...	...	...	...
$\omega_N$	$X1(\omega_N)$	$X2(\omega_N)$	...	$Xj(\omega_N)$	...	$XP(\omega_N)$	$Y(\omega_N)$

Chaque exemple  $\omega_i$ , pour  $i = 1..N$ , est caractérisé par

$X1(\omega_i), X2(\omega_i), \dots, Xj(\omega_i), \dots, XP(\omega_i), Y(\omega_i)$

ou

$X(\omega_i), Y(\omega_i)$  si on note  $X = (X1, X2, \dots, XP)$

### Problème de l'apprentissage

Trouver dans les données une relation de dépendance entre la variable Y et les autres variables X1, X2, ..., XP

Autrement dit, trouver une fonction  $\phi$  telle que

$$Y = \phi(X1, X2, \dots, XP)$$

ou

$$Y = \phi(X)$$

## Objectif

Utiliser  $\phi$  comme un modèle pour prédire ou trouver  $Y$  pour un nouvel individu  $\omega$  si on connaît ses attributs

$X_1(\omega), X_2(\omega), \dots, X_P(\omega)$

c'est à dire

Pour un nouvel individu  $\omega$

prédire (ou trouver)  $Y(\omega)$

étant donnés  $X_1(\omega), X_2(\omega), \dots, X_P(\omega)$

$\omega$	$X_1(\omega)$	...	$X_j(\omega)$	...	$X_P(\omega)$	?
----------	---------------	-----	---------------	-----	---------------	---

$Y(\omega) = \phi(X(\omega))$

Les données sont dites données d'apprentissage

L'apprentissage est dit supervisé

Le modèle de prédiction  $\phi$  est appris (découvert) dans des données qui contiennent des exemples  $\omega_i$  pour lesquels on connaît  $Y(\omega_i)$

## Deux cas

variable à expliquer  $Y$  qualitative

les valeurs de  $Y$  sont des classes à prédire

tâche de prédiction appelée **classification**

variable à expliquer  $Y$  quantitative

tâche de prédiction appelée **régression**



## **Méthode supervisée**

méthode d'apprentissage supervisée

méthode de classification supervisée

méthode de régression

## **Notions de sur-apprentissage et de sous-apprentissage**

Le modèle appris dans les données d'apprentissage peut être

- un modèle trop complexe

construit avec un apprentissage trop poussé

S'ajuste parfaitement aux données d'apprentissage

prend trop en compte les caractéristiques propres aux données d'apprentissage

sur-apprentissage ou sur-ajustement (overfitting en Anglais)

Erreur d'apprentissage faible

Erreur de prédiction avec les exemples des données d'apprentissage

Erreur de généralisation élevée

Erreur de prédiction avec de nouveaux exemples

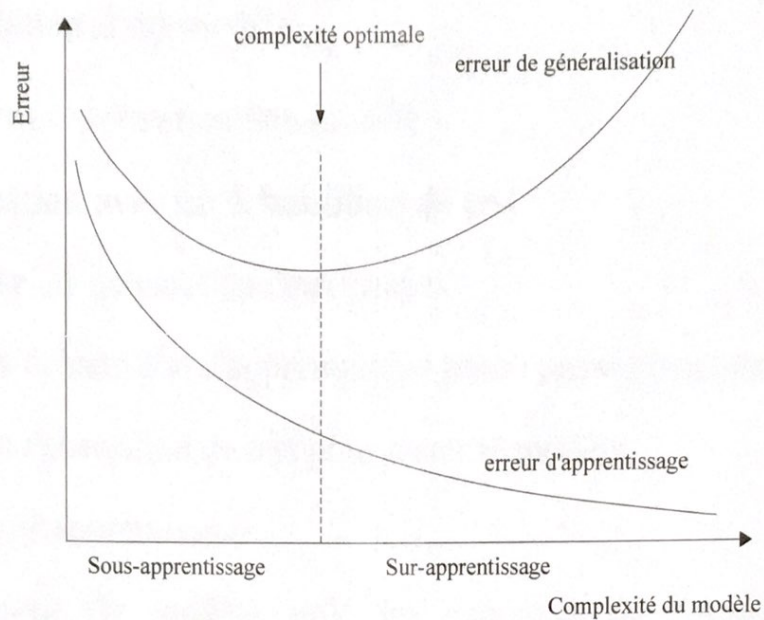
- un modèle trop simple

construit avec un apprentissage pas assez poussé

sous-apprentissage ou under-fitting

Erreur d'apprentissage élevée

Erreur de généralisation élevée



Relation entre complexité et d'un modèle et erreurs  
d'apprentissage et de généralisation

Un bon modèle

compromis entre complexité et bonne généralisation

## **Validation d'un modèle**

Objectif : trouver un bon modèle

### **Validation avec un échantillon de test**

Diviser les données en deux parties

- un échantillon d'apprentissage pour construire le modèle

- un échantillon de test pour tester le modèle

Erreur d'apprentissage

- erreur du modèle avec les exemples de l'échantillon d'apprentissage

Erreur de test

- erreur du modèle avec les exemples de l'échantillon de test

- une estimation de l'erreur de généralisation

Bon modèle

- une erreur d'apprentissage faible

- une erreur de test faible



## Calcul des erreurs d'apprentissage et de test

Même formule

$$E(\varphi) = \frac{1}{N} \sum_{i=1}^N d(Y(\omega_i), \varphi(X(\omega_i)))$$

N est la taille de l'échantillon d'apprentissage ou de test

### Cas de la classification

$$\begin{aligned} d(Y(\omega_i), \varphi(X(\omega_i))) &= 0 \text{ si } Y(\omega_i) = \varphi(X(\omega_i)) \\ &= 1 \text{ si } Y(\omega_i) \neq \varphi(X(\omega_i)) \end{aligned}$$

donc

$$E(\varphi) = (\text{Nombre d'exemples mal classés}) / N$$

Exprimée souvent en %

=>taux d'erreur de classification

### Cas de la régression

1) Erreur quadratique moyenne ( Mean Squared Error MSE)

$$d(Y(\omega_i), \varphi(X(\omega_i))) = (Y(\omega_i) - \varphi(X(\omega_i)))^2$$

$$E(\varphi) = \frac{1}{N} \sum_{k=1}^N (Y(\omega_i) - \varphi(X(\omega_i)))^2$$

2) Erreur absolue moyenne (Mean Absolute Error MAE)

$$d(Y(\omega_i), \varphi(X(\omega_i))) = |Y(\omega_i) - \varphi(X(\omega_i))|$$

$$E(\varphi) = \frac{1}{N} \sum_{k=1}^N |Y(\omega_i) - \varphi(X(\omega_i))|$$