



Fondement de l'Intelligence Artificielle



Abir CHAABANI

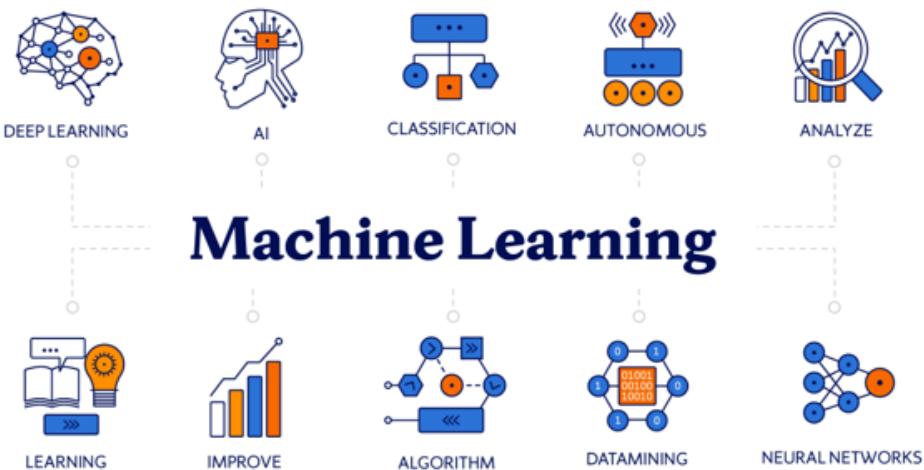
abir.chaabani@gmail.comabir.chaabani@enicar.ucar.tn

Université de Carthage - ENICARTHAGE

2 GINF 2024/2025

1

Chapitre 5



2

Plan Chapitre 5

Concepts de base de l'apprentissage automatique

- P1: Les concepts de l'apprentissage et de l'apprentissage automatique
- P2: Programmation Vs apprentissage automatique
- P3: Les applications de l'apprentissage automatique
- P4: Format des données en apprentissage automatique
- P5: Les approches d'apprentissage automatique (supervisé/non supervisé)

3

Partie I:

Les concepts de l'apprentissage et de l'apprentissage automatique

4

C'est quoi

L'Apprentissage

5

Apprentissage

Acquérir de nouvelles connaissances.

- On parle de l'apprentissage lorsque l'expertise humaine est absente

L'apprentissage désigne tout processus par lequel un être vivant (agent) acquiert des connaissances, compétences ou comportements à partir d'expériences, d'observations ou d'instructions.

Exemple : un enfant apprend à reconnaître un chat après en avoir vu plusieurs.

==> L'apprentissage est **un ensemble de mécanismes** menant à l'acquisition de **savoir-faire**, de **savoirs** ou **de connaissances**.

6

Problème d'apprentissage

Apprendre à classifier

l'apprenant commence par observer une série de champignons classés



Russula Aurea

(C)



Amanita Muscaria

(T)



Boletus Edulis

(C)



Amanita Virosa

(M)



Cortinarius Bolaris

(T)



Gyromitra Esculenta

(M)



Cantharellus Cibarius

(C)



Hygrocybe Coccinea

(N)

⋮

comestibles (C) non comestibles (N) toxiques (T) mortels (M).

7

Problème d'apprentissage

- l'apprenant cherche à classer les champignons d'une nouvelle série dont on ignore la classe d'appartenance



(C)



(M)

Amanita Phalloides



(T)



Amanita Pantherina

→ la performance de l'apprenant est mesurée par sa capacité à classifier correctement de nouveaux champignons.



(M)



(N)

Tricholoma Aurantium

8

Problème d'apprentissage

Ordonner des objets: Dans ce type de problème, l'objectif est d'apprendre à ordonner des éléments (par exemple, des films) selon les préférences d'un utilisateur.

- ✓ L'apprenant commence par observer un ensemble d'utilisateurs, chacun ayant fourni une liste de films ordonnée par préférence.
- ✓ L'algorithme apprend ainsi à capturer les critères implicites (tels que le genre, le réalisateur, la durée, etc.) de préférence afin de pouvoir, par la suite, prédire un ordre de préférence personnalisé pour de nouveaux films ou de nouveaux utilisateurs.

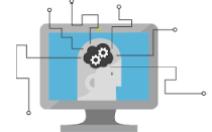


Les composantes de l'apprentissage

- La tâche à résoudre :** Il peut s'agir de classifier des champignons, ordonner des films selon les préférences des utilisateurs, etc.
- Les exemples de la tâche :** Ce sont des données annotées utilisées pour l'apprentissage, comme des champignons déjà classés ou des films ordonnés selon les préférences d'un utilisateur.
- La mesure de performance :** Elle permet d'évaluer la qualité de l'apprentissage, par exemple en mesurant le nombre d'erreurs, la distance entre ordres prédits et réels, ou d'autres métriques selon la tâche.

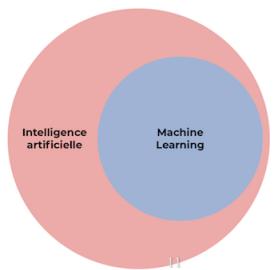
Apprentissage automatique

Apprentissage automatique=Apprentissage artificiel = Machine learning



Définition

- L'apprentissage automatique (en anglais Machine Learning, littéralement « l'apprentissage machine ») est une branche de l'IA consistant à apprendre par l'expérience ou par une base de données des règles implicites pour répondre à un problème donné.
- Elle concerne la conception, l'analyse, le développement et l'implémentation de modèles, permettant à une machine d'apprendre et de faire des prédictions à partir des données afin de remplir une tâche.
- Ce domaine s'oriente spécialement autour de l'analyse statistique de données d'entraînement.



Citations marquantes sur l'apprentissage automatique

- "A breakthrough in machine learning would be worth ten Microsofts."
— Bill Gates, Président de Microsoft
- "Machine learning is the next Internet."
— Tony Tether, Directeur de la DARPA
- "Machine learning is the hot new thing."
— John Hennessy, Président de Stanford
- "Machine learning is going to result in a real revolution."
— Greg Papadopoulos, CTO de Sun Microsystems

Citations marquantes sur l'apprentissage automatique

- Le terme "apprentissage automatique" (machine learning) a été introduit dès **1959** par **Arthur Lee Samuel**, pionnier de l'intelligence artificielle. Il l'a défini comme la capacité pour un programme à apprendre à partir de l'expérience sans être explicitement programmé.
- Un champ d'études qui donne aux ordinateurs la capacité d'apprendre des tâches pour lesquelles ils ne sont pas spécifiquement programmés
- Doter la machine d'un mécanisme d'apprentissage
- Concevoir des programmes pouvant s'améliorer automatiquement avec l'expérience

13

Citations marquantes sur l'apprentissage automatique

Le Machine Learning est un concept fondé sur le principe suivant :

1. Il existe des algorithmes génériques capables d'apprendre à partir des données ;
2. Ces algorithmes sont alimentés par des ensembles de données (données d'apprentissage) ;
3. Ils permettent d'automatiser l'apprentissage sans qu'il soit nécessaire de développer un code spécifique pour chaque tâche ;
4. À partir de ces données, les algorithmes peuvent construire leur propre logique et générer automatiquement des règles ou des modèles de prédiction.

14

Quelle différence entre l'intelligence artificielle (IA) et l'apprentissage automatique (Machine Learning)

- **Intelligence Artificielle (IA)** : Il s'agit de la science qui cherche à rendre les machines capables de se comporter de manière "intelligente", en imitant certaines capacités humaines comme la résolution de problèmes, la prise de décision, la reconnaissance vocale, etc. L'IA est un domaine très large qui inclut différentes approches et techniques, dont l'apprentissage automatique.

(Chappell, responsable du machine learning chez Intel)

- **Apprentissage Automatique (Machine Learning)** : C'est une méthode spécifique au sein de l'IA. Elle permet aux ordinateurs de "apprendre" à partir des données sans être explicitement programmés pour chaque tâche. En d'autres termes, au lieu de coder des règles précises pour accomplir une tâche, on permet au système d'apprendre par lui-même en analysant des exemples de données (comme des images, des textes, etc.).

15

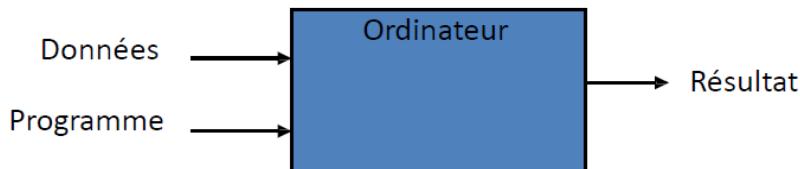
(l'Université de Stanford)

Partie II:

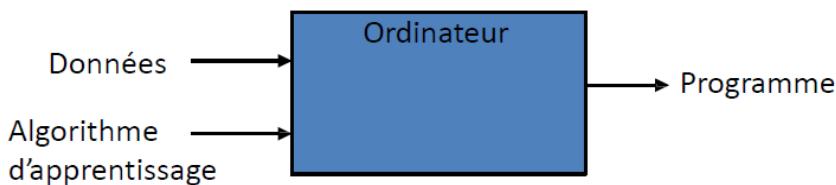
Programmation Vs apprentissage automatique

16

Programmation traditionnel



Apprentissage automatique



17

Problème: Estimation du prix d'une maison

Ecrire un programme qui permet d'estimer le prix d'une maison.



Prix ??

Prix ??

18

Programme estimation prix maison!!

```
Fonction estimation_prix_vente_maison(nombre_de_chambres, superficie,
quartier)
prix = 0
# A côté de chez moi, le prix moyen au m2 est de 1800 dinars
prix_par_metre_carre = 1800 if quartier == 'Ezzahra'
# mais quelques quartiers sont plus chers...
prix_par_metre_carre = 3000 elseif quartier== 'Lac2'
# ...et d'autres moins
prix_par_metre_carre= 500
# commençons avec un prix estimé à partir de la superficie du lieu
prix = prix_par_metre_carre * superficie
# ajustons maintenant notre prix en fonction du nombre de chambres
if nombre_de_chambres == 0:
# les studios sont moins chers (moins de murs)
prix = prix-20000
else:
# les lieux avec plus de chambres ont en général plus de valeurs
prix = prix + (nombre_de_chambres * 1000)
.....
.....
return prix
```

19

Programme estimation prix maison!!

- ⌚Le programmeur doit être connaisseur du domaine de vente des maisons
- ⌚La qualité de prédiction dépendra de l'expertise du programmeur
- ⌚Le nombre de paramètre à inclure dans le programme est réduit (exemple:vous ne pouvez pas spécifier tous les quartiers!!!)
- ⌚Si les prix changent, on reprogramme tout!!!L'année prochaine même estimation??

20

Programme estimation prix maison!!

==> Il faut d'apprendre à la machine la manière dont un agent immobilier détermine le prix d'une maison

Comment un agent immobilier peut-il déterminer le prix exacte d'une maison?

==> Un agent immobilier qui vend des maisons depuis longtemps, peut

Déterminer le prix d'une maison par similitude:

- ✓ Exemple: il sait qu'une maison dans le même quartier ayant les mêmes caractéristiques a été vendu à un tel prix. Donc la maison vaut le même prix

21

On doit reprendre ce principe pour la machine

Partie III:

Les applications de l'apprentissage automatique

22

Application classique de ML

- Les applications classiques de l'apprentissage automatique sont variées et couvrent de nombreux domaines. Voici quelques exemples :

1. Traitement du langage naturel (TLD) :

Par exemple, les chatbots, la traduction automatique, la reconnaissance de la parole, et l'analyse de sentiments.

2. Reconnaissance des formes :

Par exemple appliquée dans la reconnaissance d'images et de visages, la lecture de caractères manuscrits (OCR), et la détection d'objets dans des images ou des vidéos.

3. Diagnostic médical :

Permet de détecter des maladies à partir d'images médicales (radiographies, IRM), d'analyser les symptômes, et de prédire des diagnostics en fonction de données historiques des patients.

4. Finance (détectio[n] de fraude) :

Utilisé pour détecter des comportements suspects, comme les fraudes à la carte bancaire, en analysant les transactions pour repérer des anomalies.

5. Analyse des marchés boursiers :

Utilisé pour prédire les mouvements des marchés financiers, détecter des tendances et automatiser le trading, ainsi que pour l'analyse des risques.

Google RankBrain

• Un système d'IA développé par Google pour mieux comprendre les requêtes des utilisateurs et afficher des résultats plus pertinents.

• Lancé en 2015, c'est une partie de l'algorithme **Hummingbird** (le "moteur" principal de Google Search).

- Google traite plus de 3 milliards de recherches par jour.
- Seulement 15% sont des nouvelles recherches (chiffre de 2013)

→ Donc presque 85% des requêtes qui se répètent !!

- Google utilise une technologie de Machine Learning appelé RankBrain pour afficher les résultats les plus pertinents aux requêtes des internautes parmi des milliards de pages.



Autres applications

- ❖ Développement de véhicules autonomes (Ex: voiture autonome de Google)



- ❖ Classification des emails dans gmail



- ❖ La traduction en temps réel /La reconnaissance vocale Siri d'Apple



- ❖ Détection de fraude dans le monde de la banque

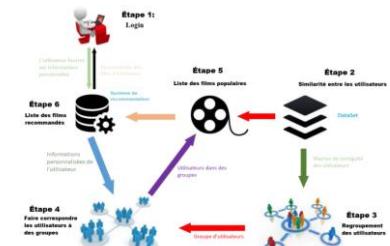


Autres applications

- ❖ Reconnaissance faciale



- ❖ Systèmes de recommandations en ligne comme (Ex: Netflix , Amazon)



- ❖ Analyse de sentiments (Sentiment Analysis (Avis, commentaires sur les réseaux sociaux, réponses à des sondages) de clients (Ex: comprendre quel type d'émotion leur marque suscite sur une population donnée)



26

Partie IV:

Format des données en apprentissage automatique

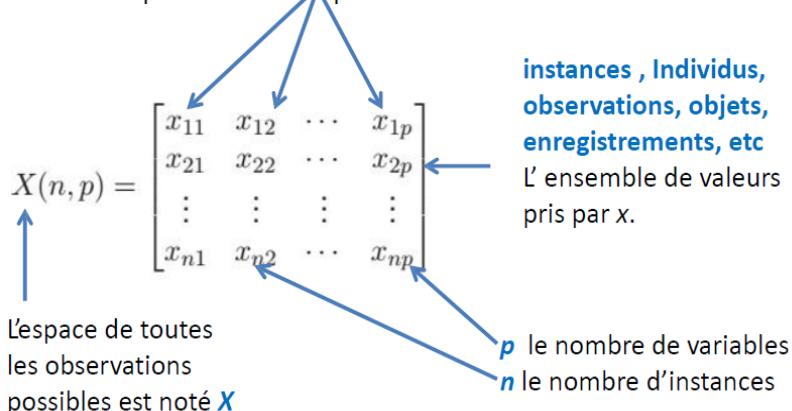
27

Format des données en apprentissage

- En apprentissage automatique, le format des données est crucial pour entraîner les modèles de manière efficace. Les données doivent être organisées d'une manière qui permet aux algorithmes d'apprentissage de comprendre et de détecter les relations entre les variables d'entrée (features) et les variables de sortie (cibles).

Variables, caractères, attributs, descripteurs, champs, etc.

décrit quantité ou une qualité observée d'une instance



28

Format des données en apprentissage

Format de données: exemple

X
• $n=6$
• $P=4$

Nom	Grade	Années	Titulaire
David	Assistant	3	non
Marie	Assistant	7	oui
Jean	Professeur	2	oui
Jim	Prof. Associé	7	oui
Pierre	Assistant	6	non
Anne	Prof associé	3	non

Variables

Instances

Ensemble d'enseignants

29

Format des données en apprentissage

On veut savoir qui sont les enseignants titulaires

Nom	Grade	Années	Titulaire
David	Assistant	3	non
Marie	Assistant	7	oui
Jean	Professeur	2	oui
Jim	Prof. Associé	7	oui
Pierre	Assistant	6	non
Anne	Prof associé	3	non

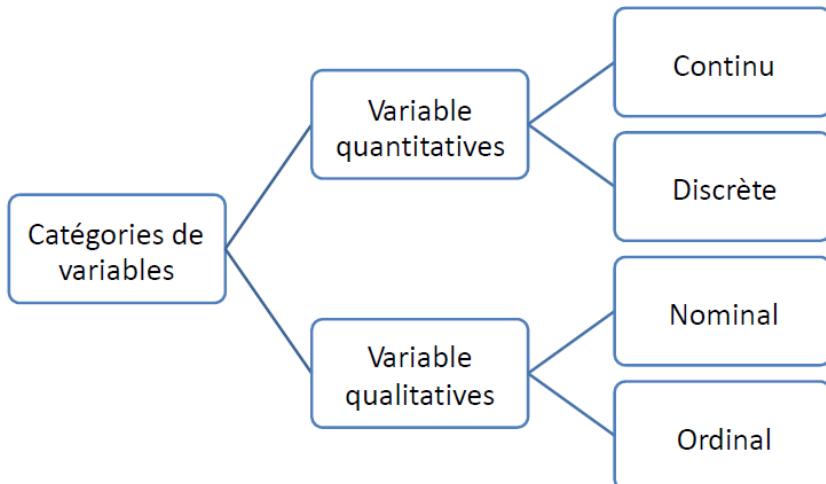
Variables prédictives
Descripteurs
Variables exogènes

Variable à prédire
Attribut classe
Variable endogène

30

Format des données en apprentissage

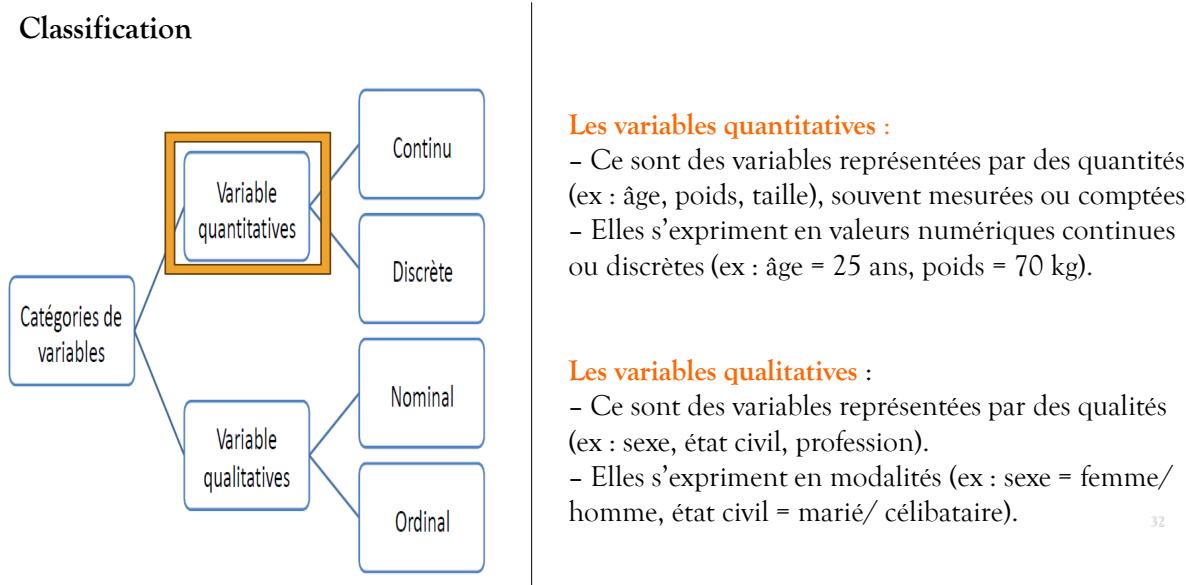
Classification



31

Format des données en apprentissage

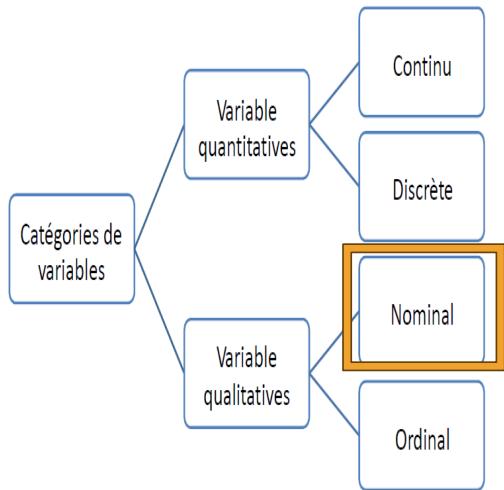
Classification



32

Format des données en apprentissage

Classification



Les variables qualitatives nominales :

- Il n'y a aucun ordre précis entre les modalités.
- Exemple :
 - ✓ Sexe : féminin, masculin
 - ✓ Saveur de crème glacée : chocolat, vanille, pistache

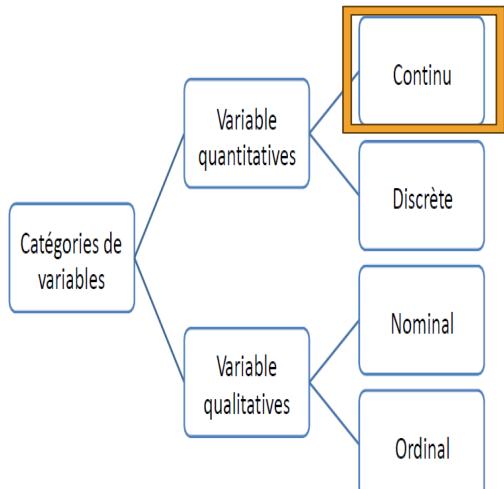
Les variables qualitatives ordinaires :

- Ce sont des variables qui contiennent un ordre entre les modalités.
- Exemple :
 - ✓ Degré de satisfaction : {insatisfait, satisfait, très satisfait}
 - ✓ Niveau d'étude : { primaire, secondaire, universitaire}

33

Format des données en apprentissage

Classification



Les variables quantitatives continues :

- Elles peuvent prendre n'importe quelle valeur numérique dans un intervalle donné.
- Exemple : Taille, Température

Les variables quantitatives discrètes :

- Elles ne peuvent prendre que des valeurs numériques spécifiques, mais ces valeurs ont une interprétation quantitative claire.
- Exemple : Nombre de plaintes de clients, Nombre d'heures supplémentaires

34

Format des données en apprentissage

Exercice : Les variables suivantes sont-elles quantitatives ou qualitatives ?

1. Les marques des voitures garées sur un parking de supermarché

Qualitative nominale (les marques sont des catégories sans ordre précis)

2. Les nationalités des touristes se rendant au festival de Carthage

Qualitative nominale (les nationalités sont des catégories sans ordre précis)

3. L'âge des auditeurs de Radio Mosaïque

Quantitative continue (l'âge est une valeur numérique pouvant varier sur un intervalle)

4. Les températures matinales relevées chaque jour sous abri à Nabeul

Quantitative continue (la température est une valeur numérique pouvant prendre une plage de valeurs continues)

35

Partie IV:

Approches d'apprentissage automatique

36

Approches d'apprentissage automatique

Données

Approches ML

Expériences

Apprentissage supervisé

- Il utilise un jeu de données étiquetées : chaque entrée est associée avec une sortie déjà connue qui peut être catégorique ou numérique. -> un mapping de l'entrée avec la sortie.
- Tente de trouver la règle permettant de relier l'entrée(x) à la sortie(y) -> **Phase d'apprentissage** dont le résultat est un **modèle prédictif**.

Objectif: Modéliser et prévoir

Apprentissage non supervisé

- Il utilise un ensemble de données non étiquetées et sont autorisés à agir sur ces données sans aucune supervision.
- Tente de trouver la structure sous-jacente d'un ensemble de données, **de regrouper ces données en fonction des similarités et des différences entre les données**

Objectif: Organiser et synthétiser

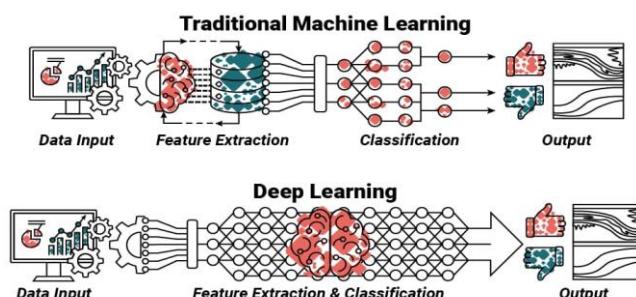
Apprentissage par renforcement

- Un agent apprend en interagissant avec un environnement.
- Il reçoit des récompenses ou pénalités et apprend à maximiser la récompense cumulée via essais/erreurs.

Objectif: Apprendre une stratégie (politique d'actions)¹⁷

Deep Learning

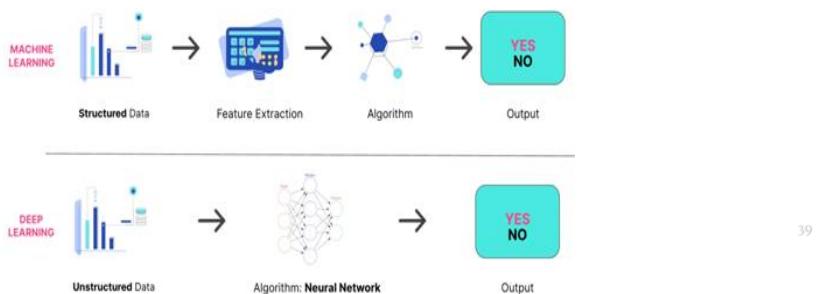
- Le **deep learning** (ou **apprentissage profond**) est une forme d'IA dérivée du Machine Learning
- Repose sur la capacité d'une technologie (machine) à apprendre par elle-même à partir de données brutes de quantités massives, grâce à des réseaux de neurones artificiels dont la structure imite celle du cerveau humain
- Plusieurs applications : reconnaissance faciale ou vocale, traduction des conversations orales en temps réel, voitures autonomes, modération automatique des réseaux sociaux, chatbots, traitement de texte, moteurs de recherches, etc.



38

Deep Learning

- les algos de ML traitent des données quantitatives et structurées (des valeurs numériques).
- Les algos de DL traiteront des données non-structurées, comme le son, le texte brut, l'image et toutes sortes de données non structurées.
- Méthode de **ML** particulièrement adaptée pour analyser des images ou du texte en grande quantités, comme en disposent les **GAFAM**, pour identifier des objets, des visages dans les photos, du sentiment ou des sujets dans du texte, etc.



Deep Learning vs Machine Learning

Objectif partagé : extraction d'insights facilitant la prise de décision. ML, à l'instar des statistiques, adopte une approche permettant la compréhension de la structure des données. Le DL combiner des avancées technologiques (puissance calcul et réseaux de neurones artificiels) afin d'apprendre des patterns complexes au sein de quantités massives de données.

- Le ML se base sur l'extraction de caractéristiques (feature extraction-dépendant de l'expertise humaine) **ALORS** Pas de features extraction en Deep Learning

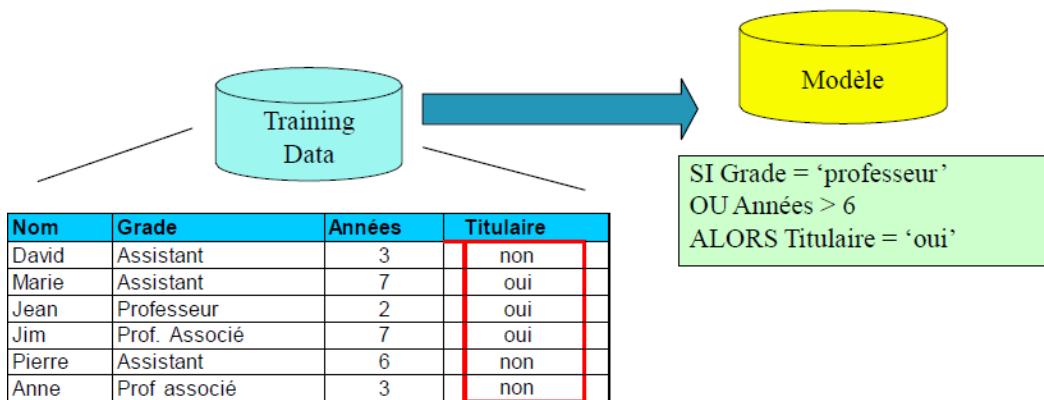
1. Les réseaux de neurones **apprennent d'eux-mêmes** les features directement à partir des données brutes
2. Les couches cachées (**hidden layers**) des réseaux de neurones **auto-optimisent** les features: Premières couches détectent des motifs simples (bords, couleurs). Et les couches profondes combinent ces motifs pour reconnaître des concepts complexes (visages, objets).

Apprentissage Supervisé

41

Apprentissage supervisé

- On cherche à produire automatiquement des règles à partir d'une base de données d'apprentissage contenant des «exemples» (en général des cas déjà traités et validés).



42

Apprentissage supervisé

Population Ω :
 (objet de l'étude)
 Ensemble de paires (x_i, y_i) ,



• $X: X = (x_1 | \dots | x_p)$ variables exogènes
 • Y : variable à prédire (endogène)

On veut construire une fonction de prédiction telle que
 $y_i = f(x_i)$

L'objectif de l'apprentissage:

Trouver une bonne approximation h de f en utilisant un échantillon de Ω telle que l'on minimise l'erreur théorique.

43

Apprentissage supervisé

Population Ω :
 (objet de l'étude)
 Ensemble de paires (x_i, y_i) ,



• $X: X = (x_1 | \dots | x_p)$ variables exogènes
 • Y : variable à prédire (endogène)

On veut construire une fonction de prédiction telle que
 $y_i = f(x_i)$

L'objectif de l'apprentissage:

Trouver une bonne approximation h de f en utilisant un échantillon de Ω telle que l'on minimise l'erreur théorique.

44

Evaluation de l'apprentissage

Pourcentage de classification correcte (PCC)

$$PCC = \frac{\text{Nombre d'objets correctement classés}}{\text{Nombre total des objets tests}}$$

Ensemble test

Âge	Fumeur	Taille	...	Cible (Atteint du cancer : 1 Pas atteint : 0)	Prédiction (Ce que notre modèle prédisait)
68	Oui	1m79		1	1
43	Oui	1m82		0	1
47	Non	1m54		1	1
58	Qui	1m67		1	0
33	Non	1m91		0	0
41	Non	1m62		0	0
74	Oui	1m74		1	1
59	Oui	1m77		1	1
27	Non	1m68		0	0

$$PCC = 7/9 = 77,78\%$$

$$\text{Taux d'erreur} = 22,22\%$$

45

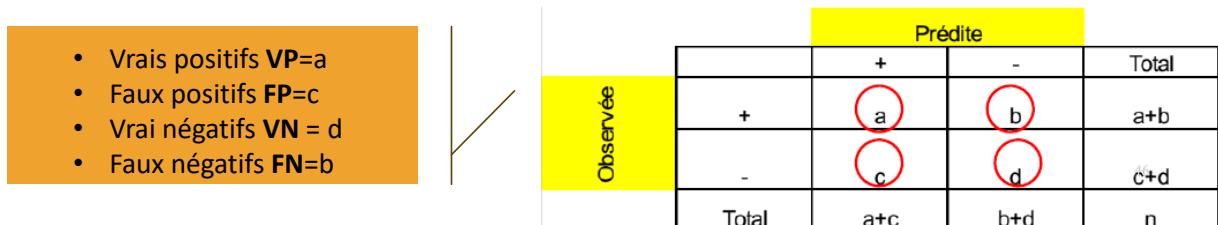
Evaluation de l'apprentissage

Matrice de confusion

Objectif : Le principe de la matrice de confusion est de "confronter la vraie valeur avec la prédiction". En d'autres termes, elle permet de visualiser comment votre modèle a classifié les données par rapport à la réalité. Elle résume les résultats d'une classification en montrant le nombre de prédictions correctes et incorrectes pour chaque catégorie

Structure de la matrice :

- Les lignes représentent les valeurs observées (la vérité du terrain).
- Les colonnes représentent les valeurs prédites par votre modèle.



		Prédite		Total
		+	-	
+	+	a	b	a+b
	-	c	d	c+d
Total		a+c	b+d	n

Evaluation de l'apprentissage

Matrice de confusion

Les quatre indicateurs principaux (VP, FP, VN, FN) résultant de cette matrice:

Observée	Prédite		Total
	+	-	
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n

(a) : **VRAIS POSITIFS (VP)** : Le modèle a prédit "+" et la valeur réelle était "+". Ce sont les cas correctement identifiés comme positifs.

==> C'est une mesure de la proportion de cas positifs que le modèle a correctement identifiés.

Taux de Vrais Positifs (VP)= $VP/(VP+FN)$

Dans cet exemple a correspond au nombre de Vrais Positifs (VP)= $a/(a+b)$ \rightarrow **Sensibilité=Rappel**

47
VP est connue sous deux noms principaux dans le domaine de ML

Evaluation de l'apprentissage

Matrice de confusion

Les quatre indicateurs principaux (VP, FP, VN, FN) résultant de cette matrice:

Observée	Prédite		Total
	+	-	
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n

(c) : **FAUX POSITIFS (FP)** : Le modèle a prédit un résultat positif (« + »), alors que la valeur réelle était négative (« - »), indiquant une fausse identification d'un cas comme positif

\rightarrow C' est la proportion de Faux Positifs parmi tous les cas réellement négatifs

Taux de Faux positifs (FP)= $FP/(FP+VN)$

Dans cet exemple a correspond au nombre de Faux positifs (FP)= $c/(c+d)$

48

Evaluation de l'apprentissage

Matrice de confusion

Les quatre indicateurs principaux (VP, FP, VN, FN) résultant de cette matrice:

Observée	Prédite		Total
	+	-	
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n

(b) : FAUX NÉGATIFS (FN) : des éléments étaient réellement des "+" (Observée +) mais que ton modèle a incorrectement prédit qu'ils étaient des "-" (Prédite -).

Taux de Faux négatifs (FN)=FN/(FN+VP)

Dans cet exemple c correspond au nombre de Faux positifs (FP)= b/(b+a)

49

Evaluation de l'apprentissage

Matrice de confusion

Les quatre indicateurs principaux (VP, FP, VN, FN) résultant de cette matrice:

Observée	Prédite		Total
	+	-	
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n

(d) : VRAI NÉGATIFS (VN) : Ces éléments étaient réellement des "-" (Observée -) et que ton modèle a correctement prédit qu'ils étaient des "-" (Prédite -).

Taux de Vrai Négatifs (VN)=VN/(VN+FP)

Dans cet exemple a correspond au nombre de Faux positifs (FP)= d/(d+c)

50

Evaluation de l'apprentissage

Matrice de confusion: exemple

		REEL	
		Si le patient est atteint ou non	
		Est atteint	N'est pas atteint
PREDICTION <i>Ce que notre modèle prédisait</i>	Est atteint	Nombre de Vrai positif	Nombre de Faux positif
	N'est pas atteint	Nombre de Faux négatif	Nombre de Vrai négatif

- **Vrai positif (VP)**: le patient est atteint du cancer, et le modèle l'a également prédit.
- **Vrai négatif (VN)** : Le modèle prédit que le patient n'a pas le cancer, et effectivement, il ne l'a pas.
- **Faux positif (FP)** : Le patient n'a pas le cancer mais le modèle prédit qu'il l'a.
- **Faux négatif (FN)** : Le modèle ne prédit pas que le patient a le cancer, alors qu'il en est réellement atteint.

51

Evaluation de l'apprentissage

Matrice de confusion: exemple

- ❖ **Matrice de confusion :** Il y a 4 indicateurs déductibles de la matrice de confusion déterminant les performances du modèle :

- ❖ L'exactitude (**Accuracy**)
- ❖ La précision (valeur de prédiction positive) (**Precision**)
- ❖ Le rappel (taux positif réel) (**Recall**)
- ❖ Le score F-1 (**F-measure**)

52

Evaluation de l'apprentissage

1. L'exactitude (**Accuracy**): Le nombre de vraies valeurs correctement prédites sur le total

$$\text{Exactitude} = \frac{\text{Nombre de prédictions correctes}}{\text{Total des prédictions}}$$

$$\text{Exactitude} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FN} + \text{FP} + \text{VN}} = \frac{a + d}{(a + b + c + d)}$$

➔ Doit être prise en compte lorsque les vrais positifs et les vrais négatifs sont les plus importants, et que l'ensemble de données est équilibré (qu'il y a à peu près le même nombre d'exemples pour chaque classe (par exemple, autant de cas positifs que de cas négatifs)).

53

Evaluation de l'apprentissage

2. Précision (**Precision**): Le nombre de valeurs positives réelles sur le total des valeurs positives prévues.

$$\text{Précision} = \frac{\text{Nombre de valeurs positives réelles (parmi les prédites positives)}}{\text{Totoal des valeurs positives prévues}}$$

$$\text{Précision} = \frac{\text{VP}}{\text{VP} + \text{FP}} = \frac{a}{(a + c)}$$

En fonction du problème, choisir précision chaque fois que **le taux de faux positif a un impact plus important**. Si un Faux Positif a un impact très négatif (financier, humain, etc.), tu veux minimiser le risque que ton modèle fasse cette erreur.

==> la Précision est la métrique à regarder en priorité lorsque le coût ou l'impact de se tromper en disant "positif" (Faux Positif) est plus important que le coût de se tromper en disant "négatif" (Faux Négatif)

Evaluation de l'apprentissage

3. Rappel (Recall) : le nombre de valeurs correctement prédictes comme positives sur le total des valeurs positives réelles.

$$\text{Rappel} = \frac{\text{Nombre de cas positifs correctement prédicts}}{\text{Total des cas réellement positifs}}$$

$$\text{Rappel} = \frac{\text{VP}}{\text{VP} + \text{FN}} = \frac{a}{(a + b)}$$

- Un Rappel élevé signifie que parmi tous les éléments qui étaient réellement positifs, ton modèle en a identifié une grande proportion

==> En fonction du problème, choisir rappel chaque fois **que le taux de faux négatif est important**

55

Evaluation de l'apprentissage

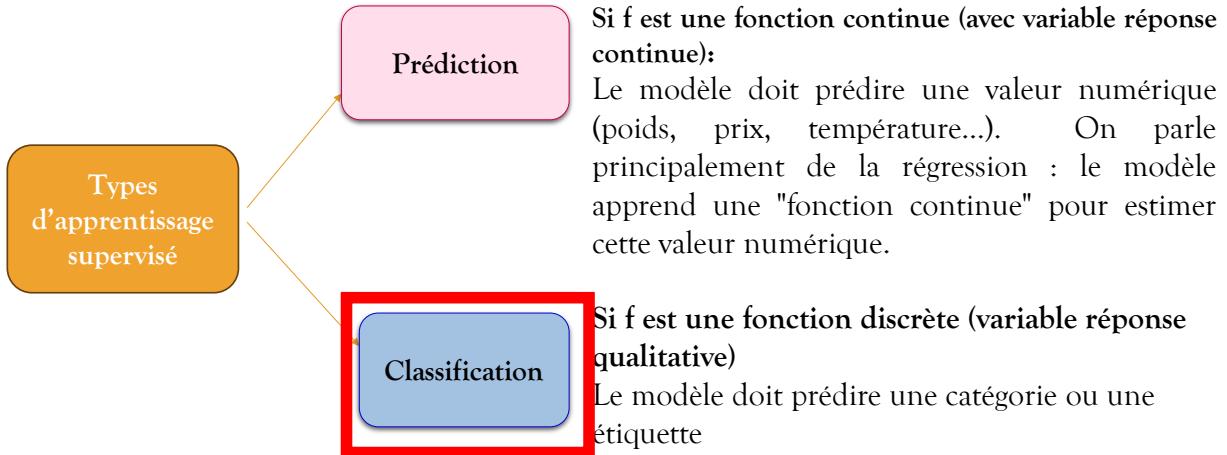
4. F1-Score /F-measure : Combiner précision et rappel pour obtenir un résultat reflétant le plus possible la réalité.

$$\text{F1-Score} = 2 * \frac{(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})}$$

- ➔ Ce score est à utiliser lorsque le faux positif et le faux négatif sont tous deux d'égale importance. Cela permet au modèle de considérer à la fois la précision et le rappel de manière égale en utilisant un seul score.
- ➔ F1-Score varie entre 0 et 1. Un F1-Score élevé indique que le modèle a à la fois une bonne Précision et un bon Rappel. Si l'une des deux métriques (Précision ou Rappel) est très faible, le F1-Score sera également faible, car il pénalise les performances déséquilibrées.

56

Type d'apprentissage supervisé



57

Classement

Principe : C'est l'opération qui consiste à prendre chaque "individu" ou élément étudié (par exemple, une image, un client, une transaction) et à lui attribuer une **catégorie** ou une **classe** parmi un choix de classes qui sont déjà définies à l'avance (par exemple, "chat", "chien", "spam", "non-spam"). Cette attribution se fait en se basant sur les **caractéristiques** de cet individu, que l'on appelle les "variables explicatives- features" (ce sont les informations dont on dispose sur l'individu).

- Pour savoir à quelle classe attribuer un individu en fonction de ses caractéristiques, on utilise généralement une **formule**, un **algorithme** ou un ensemble de **règles**. C'est ce qu'on appelle **le modèle de classification**. Ce modèle n'est pas connu au départ ; il faut le "découvrir" ou l'apprendre à partir d'un ensemble de données où l'on connaît déjà les caractéristiques des individus et à quelle classe ils appartiennent réellement (c'est la phase d'apprentissage supervisé).

58

Processus de Classement

Étape 1 : Construction du modèle à partir de l'échantillon d'apprentissage: C'est la phase principale d'**entraînement** du modèle.

Étape 2 : Tester le modèle : vérifier la performance du modèle sur un échantillon test

Une fois le modèle construit, il est crucial de l'évaluer sur des données qu'il n'a **jamais vues pendant l'apprentissage**. C'est le rôle de « l'échantillon test » (ou "test set"). En appliquant le modèle à cet échantillon test et en comparant ses prédictions avec les vraies valeurs

Étape 3 : Validation du modèle (facultative) : vérification sur un troisième échantillon

Étape 4 : Utilisation du modèle : utiliser le modèle dans le classement de nouvelles données

59

Processus de Classement

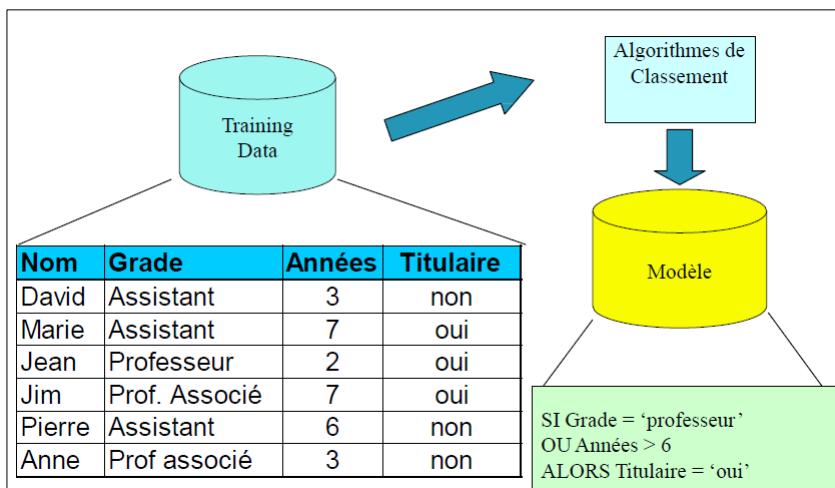
Étape 1 : Construction du modèle à partir de l'échantillon d'apprentissage: C'est la phase principale d'**entraînement** du modèle.

- On utilise un ensemble de données appelé "échantillon d'apprentissage" (ou "**training set**") pour que l'algorithme apprenne **les motifs et les relations entre les variables explicatives et la variable réponse**.

- ✓ Chaque élément (instance) étudié est supposé appartenir à l'une des classes prédéfinies.
- ✓ La classe à laquelle appartient une instance est définie par un attribut spécifique (l'attribut "classe").
- ✓ L'ensemble des données d'apprentissage est utilisé pour construire le modèle de classification.
- ✓ Ce modèle peut prendre diverses formes, **comme des règles de classement, des arbres de décision ou des formules mathématiques.**

Processus de Classement

Étape 1 : Construction du modèle à partir de l'échantillon d'apprentissage: C'est la phase principale d'entraînement du modèle.



61

Processus de Classement

Étape 2 : Tester le modèle : vérifier la performance du modèle sur un échantillon test

- **Vérifier le modèle :** On évalue le modèle qui vient d'être construit (entraîné) en l'appliquant à un autre ensemble d'individus. Pour ces individus, on connaît déjà leur vraie classe (leur classement réel).
- **Estimer le taux d'erreur du modèle :** Pour mesurer la performance du modèle :
 - ✓ On compare la classe que le modèle a prédite pour chaque individu de l'échantillon de test avec sa vraie classe connue.
 - ✓ Le Taux d'erreur est le pourcentage d'individus de l'échantillon de test qui ont été classés incorrectement par le modèle.

62

C'est l'inverse de l'Exactitude (Accuracy) : Taux d'erreur = 1 - Exactitude.

Processus de Classement

Étape 2 : Tester le modèle : vérifier la performance du modèle sur un échantillon test

Estimation de taux d'erreur (Partitionnement)

On divise l'ensemble total des données en **deux sous-ensembles indépendants** : un pour l'apprentissage (entraîner le modèle) et un pour le **test** (évaluer le modèle final).

Exemple: 2/3 des données pour l'apprentissage et le 1/3 restant pour le test.

Cette méthode est simple et rapide, mais son estimation de performance peut varier si la division initiale n'est pas représentative. Elle est souvent un premier pas ou utilisée lorsque la validation croisée est trop coûteuse en temps de calcul pour de très grands datasets.

63

Evaluation de l'apprentissage

Étape 2 : Tester le modèle : vérifier la performance du modèle sur un échantillon test

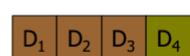
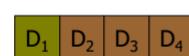
Estimation de taux d'erreur (Cross Validation)

- Utilisée pour les ensembles de données modérés.

- Principe**

- Partition de l'ensemble d'apprentissage T en n ensembles disjoints (T_1, T_2, \dots, T_n) de même taille $|T_i|$.

- Pour chaque $i = 1, 2, \dots, n$
 - On fait l'apprentissage sur $T - \{T_i\}$
 - On teste sur T_i
 - On calcule le PCC sur T_i
- On fait la moyenne des PCC.



Evaluation de l'apprentissage

Étape 2 : Tester le modèle : vérifier la performance du modèle sur un échantillon test

Estimation de taux d'erreur (Bootstrapping)

- Utilisé particulièrement pour les ensembles de données de taille réduite.

Principe: Crédation d'échantillons "artificiels" :

1. **Tirage avec remise** : On sélectionne aléatoirement des observations du jeu de données original en autorisant les répétitions.
2. **Taille identique** : On s'arrête dès qu'on a atteint le même nombre d'observations que l'original.

=> L'idée est de créer plusieurs nouveaux ensembles d'apprentissage en tirant des échantillons avec remplacement à partir de ton ensemble de données original. Chaque nouvel ensemble a la même taille que l'original, mais certains individus apparaissent plusieurs fois, et d'autres n'apparaissent pas du tout.

- Entraîner et tester le modèle plusieurs fois sur différents tirages.
- Le taux d'erreur final est la **moyenne** des erreurs obtenues sur tous les tests.

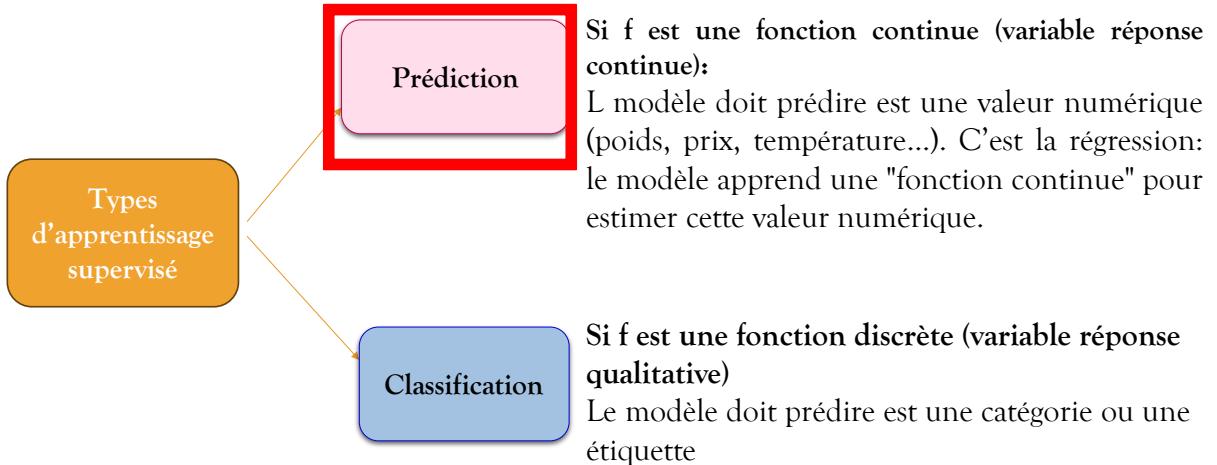
65

Algorithmes de classement

- **Les réseaux de neurones**, très utilisés dans les années 80-90, se sont inspirés à l'origine de la manière dont le cerveau humain fonctionne.
- **Les arbres de décision** construisent le modèle d'apprentissage comme une suite de décisions (critères) organisées en une structure de type arbre.
- **L'algorithme des K plus proches voisins (KNN)** classe un nouvel individu en se basant sur les classes des K individus les plus similaires ("plus proches") dans les données d'apprentissage, mesurés par une distance.
- **Les Machines à Vecteurs de Support (SVM)** reposent sur l'idée de trouver la meilleure séparation possible (une frontière ou un plan) entre les différentes classes dans l'espace des données.

66

Type d'apprentissage supervisé



67

Prédiction

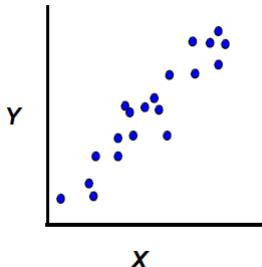
- La méthode de prédiction la plus courante est la **régression**. Son objectif est de trouver une relation entre des variables afin de prédire une valeur continue (un nombre).
- Il existe plusieurs types de régression :
 - ✓ La **régression linéaire** (simple si une seule variable explicative, multiple si plusieurs).
 - ✓ La **régression non linéaire**.
- Pour visualiser la relation entre deux variables (X et Y), on peut la représenter graphiquement par un **nuage de points**, où chaque point correspond à une observation des deux variables.

68

Régression

Comment représenter ce nuage de points ou diagramme de dispersion par une droite représentative qui décrit les rapports entre ces deux variables?

NUAGE DE POINTS:

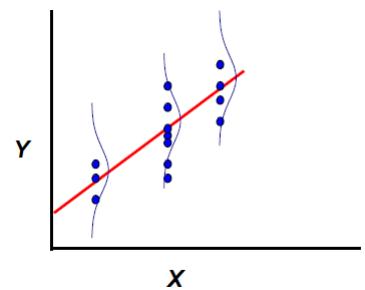


=> Cette droite s'appelle **droite d'ajustement**: Cette droite est choisie pour être la plus "représentative" de la tendance générale des points, décrivant ainsi la relation linéaire entre les deux variables.

Régression linéaire

- Concerne la relation entre deux variables : une variable explicative (X) et une variable réponse (Y).
- On essaie de modéliser cette relation par une droite.
- L'équation de cette droite est typiquement : $Y = f(x) = aX + b$

Deux paramètres à estimer
a la pente et b l'ordonnée à l'origine



✓ X=variable **explicative** («indépendante»), **contrôlée**

Utilisée pour **expliquer** ou **prédirer**

l'influence va de X vers Y, et non l'inverse

✓ Y=variable **expliquée** (dépendante), **aléatoire**

la variable que l'on essaie de comprendre, d'expliquer, ou de **prédirer**

dont la valeur dépend de l'entrée (X) mais aussi d'une part de hasard /incertitude (en général)

Propriété d'une droite de régression

- Une fois que l'on a trouvé l'équation de la droite de régression, on peut l'utiliser pour **prédir** la valeur de Y pour n'importe quelle valeur de X.
- La droite de régression **passe toujours par le point moyen** des données observées (le point ayant la moyenne de X « x.moy » et la moyenne de Y « Y.moy »).
- L'**ordonnée à l'origine** (**souvent notée b**) représente la valeur prédictive de Y lorsque la valeur de X est égale à zéro.
- La **pente** (**souvent notée a**) mesure comment la variable Y change en moyenne lorsque la variable X augmente d'une unité. Cette pente peut être positive (relation directe), négative (relation inverse) ou nulle (pas de relation linéaire).

71

Régression linéaire

Estimation des paramètres

- Il existe différentes méthodes pour calculer les valeurs des paramètres d'un modèle (comme les coefficients 'a' et 'b' en régression linéaire) :
 - ✓ La Méthode des Moindres Carrés ("Least Squares").
 - ✓ La Méthode du Maximum de Vraisemblance.
 - ✓ Des méthodes cherchant à obtenir le Meilleur Estimateur Linéaire Non Biaisé (qui décrit une propriété souhaitable de l'estimateur).
- La **Méthode des Moindres Carrés** est la plus couramment utilisée. Une raison importante de sa popularité est qu'elle permet d'obtenir des estimateurs (des valeurs calculées pour les paramètres) qui sont **non biaisés**, c'est-à-dire qu'en moyenne, ces estimations ne s'écartent pas systématiquement de la vraie valeur du paramètre.

72

Types des techniques de classement et de prédiction

- Les techniques de prédiction et de classement peuvent être classées. Parmi elles, on trouve:
 - 1. les techniques inductives**, qui se déroulent en deux étapes principales :
 - ✓ **La Phase d'apprentissage (ou d'induction)** : C'est l'étape où l'on utilise les données disponibles pour construire un modèle. Ce modèle a pour but de découvrir et résumer les relations et les schémas présents entre les différentes variables.
 - ✓ **La Phase de déduction** : Une fois le modèle créé, on l'utilise en l'appliquant à de nouvelles données (jamais vues pendant l'apprentissage). Le modèle utilise alors les relations qu'il a apprises pour faire une déduction, c'est-à-dire assigner un classement (une catégorie) ou fournir une prédiction (une valeur).

Régression Linéaire, Arbres de Décision SVM, Réseaux de Neurones

- 2. Les techniques transductives** fonctionnent en une seule fois. Pour classer un nouvel élément, elles ne construisent pas de modèle général, mais le classent directement en le comparant aux autres éléments déjà classés dans les données d'apprentissage. Elles n'impliquent pas la création d'un modèle avec des paramètres spécifiques.

- C'est pourquoi ces méthodes transductives sont dites non paramétriques.

73

K-NN

Qualités attendues

- 1. La précision** le taux d'erreur, proportion d'individus mal classés doit être le plus bas possible.
- 2. La robustesse:** le modèle doit dépendre peu que possible de l'échantillon d'apprentissage utilisé et bien se généraliser à d'autres échantillons
- 3. La concision :** les règles du modèle doivent être aussi simples et aussi peu nombreuses que possible
- 4. Les résultats explicites:** les règles du modèle doivent être accessibles et compréhensibles
- 5. La diversité des types de données manipulées :** tous les algorithmes ne sont pas aptes à manipuler les données qualitatives, discrètes, continues, et manquantes.
- 6. La rapidité de calcul du modèle :** le modèle doit donner des résultats rapidement, surtout pour les applications en temps réel.
- 7. La possibilité de paramétrage :** selon les techniques de classement ou de prédiction, il faut donner la possibilité de changer les paramètres du modèle (par exemple, pondérer les erreurs).

74

Critères de comparaison des classifieurs

1. **Taux de précision:** capacité à prédire correctement
2. **Temps de calcul:** temps nécessaire pour apprendre et tester le modèle
3. **Robustesse:** précision en présence de bruit
4. **Volume de données :** efficacité en présence de données de grande taille
5. **Compréhensibilité:** Niveau de compréhension et de finesse

N.B.

- ✓ Critères 1 et 2 “mesurables”
- ✓ Critère 4 important pour l’extraction de connaissances dans les bases de données (ECBD)
- ✓ Critères 3 et 5 “laissés à l’appréciation” de l’utilisateur- analyste

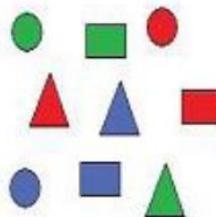
75

Apprentissage Non Supervisé

76

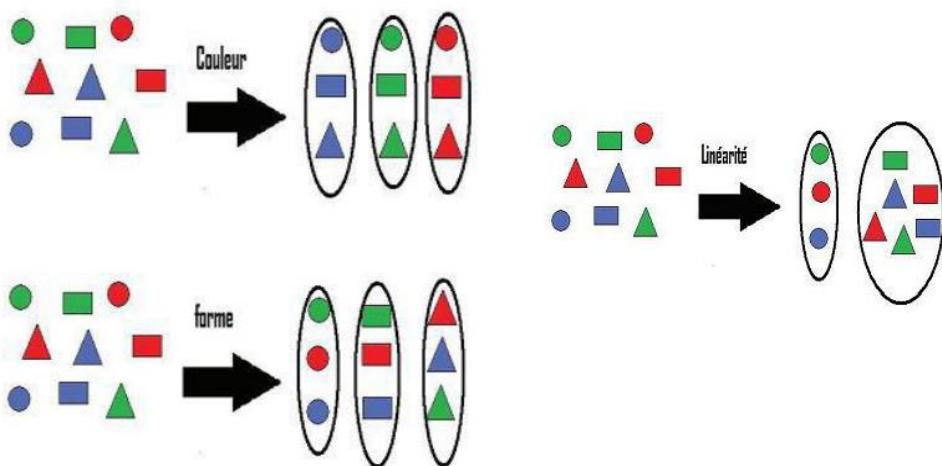
Apprentissage non supervisé

- Étant donné un ensemble d'objets, caractérisés par un certain nombre de variables



- La classification automatique (clustering) cherche à trouver des groupes (des sous-ensembles) parmi ces objets.
- L'objectif est que les objets à l'intérieur d'un même groupe soient très homogènes (très similaires entre eux)
- Et que les groupes soient bien séparés les uns des autres (les objets d'un groupe soient différents de ceux d'un autre groupe).

Apprentissage non supervisé



- ✓ Chaque sous-ensemble d'objets trouvé est appelé un cluster, un groupe ou une partition.
- ✓ Les objets qui font partie d'un même groupe (cluster) doivent donc se ressembler entre eux.
- ✓ Inversement, deux objets appartenant à des groupes différents doivent être différents l'un de l'autre.

78

Format des données en apprentissage non supervisé

Object-by-variable structure: Il s'agit d'une structure de données organisée sous forme de **matrice**, notée ici $X(n,p)$.

- Cette matrice contient la description de **n objets** (ou individus, observations, échantillons).
- Chaque ligne de la matrice correspond à un objet distinct.
- La description de chaque objet est faite à l'aide de **p variables** (ou caractéristiques, attributs, features). Chaque colonne de la matrice correspond à une variable distincte.

$$X(n, p) = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

79

Format des données en apprentissage non supervisé

Object-by-object structure: Cette structure de données est représentée par une **matrice carrée**, souvent notée M , de dimensions $n \times n$, où n est le nombre total d'objets (ou individus); **chaque ligne représente un objet ET chaque colonne représente un objet**

La valeur située à l'intersection de la ligne i et de la colonne j de cette matrice $d(i,j)$ représente une mesure de la **ressemblance** ou de la **différence** (la **distance** ou la **similarité**) entre l'objet numéro i et l'objet numéro j .

$$M(n, n) = \begin{bmatrix} 0 & d(1, 2) & \cdots & d(1, n) \\ d(2, 1) & 0 & \cdots & d(2, n) \\ \vdots & \vdots & \ddots & \vdots \\ d(n, 1) & d(n, 2) & \cdots & 0 \end{bmatrix}$$

80

Tâches associées à l'apprentissage non supervisé

Clustering (segmentation, regroupement) : Cette tâche consiste à construire des groupes (ou "classes") automatiquement en analysant les observations disponibles. L'apprentissage non supervisé est très souvent considéré comme synonyme de clustering, car c'est une tâche centrale de ce domaine..

Règles d'association : L'objectif ici est d'analyser les données pour découvrir des relations ou des associations intéressantes entre les différentes variables (par exemple, identifier des habitudes d'achat conjointes).

Re-description des données : Cette tâche fait référence à des méthodes pour résumer les caractéristiques principales des données ou pour en décrire la structure fondamentale, sans chercher à prédire une variable cible ou à former des groupes distincts, mais simplement à mieux comprendre les données elles-mêmes.

81

Clustering

- Les techniques de clustering cherchent à décomposer un ensemble d'individus en plusieurs sous ensembles les plus homogènes possibles
- Il existe un grand nombre de méthodes de clustering, et on distingue différents types :

1. Méthodes de partitionnement vs Méthodes hiérarchiques :

- ✓ Les méthodes de partitionnement divisent directement les données en un **nombre fixe de clusters**. Exemple : *K-means*.
- ✓ Les méthodes hiérarchiques construisent une structure arborescente (appelée **dendrogramme**) qui représente les relations d'inclusion entre les clusters

2. Méthodes avec recouvrement vs Méthodes sans recouvrement :

- ✓ Méthodes avec recouvrement : un même individu peut appartenir à **plusieurs clusters simultanément**. Exemple : *clustering flou* (fuzzy clustering)..
- ✓ Dans les méthodes sans recouvrement, chaque individu est affecté à **un seul cluster unique..**

82

Clustering

3. Méthodes incrémentales vs Méthodes non incrémentales :

- ✓ Les méthodes incrémentales sont capables de mettre à jour les clusters **dynamiquement (en temps réel)** à mesure que de nouvelles données arrivent.
- ✓ Les méthodes non incrémentales nécessitent de réanalyser l'ensemble des données pour intégrer de nouvelles observations.

83

Fonctions attendues du clustering

1. Prise en compte de données de différents types: catégorique, numérique, matrice des similarités, données mixtes
 Exemple: Comment comparer des objets caractérisés à la fois par des attributs numériques (tailles, poids) et catégoriques(couleur des yeux, type de véhicule)?

2. Capacité à découvrir des formes arbitraires de clusters :

- ✓ De nombreuses méthodes traditionnelles (comme K-means) favorisent la découverte de **formes sphériques**.
- ✓ Un bon algorithme doit pouvoir détecter des **clusters de formes complexes** (courbes, allongées, en anneaux, etc.).

3. Minimisation du nombre de paramètres à fixer :

- ✓ Certaines méthodes nécessitent de spécifier à l'avance le nombre de clusters (ex : K-means).

4. Interprétation et utilisation des résultats: Les utilisateurs doivent pouvoir donner un sens aux classes découvertes

5. Résistance au bruit et aux anomalies: La recherche des points isolés (outliers) est un sujet de recherche en soi

6. Problème en grande dimensions: Certaines méthodes ne sont applicables que si les objets ou individus sont décrits sur deux ou trois dimensions

7. Passage l'échelle: Certaines méthodes ne sont pas applicables sur de gros volumes de données

84

Clustering: Méthodes de partitionnement

- **L'objectif:** Construire une partition en classes d'un ensemble d'objets ou d'individus
- **Principe:**
 - ✓ Création d'une partition initiale de K classes,
 - ✓ Puis itération d'un processus qui optimise le partitionnement en déplaçant les objets d'un cluster à l'autre
- **Solutions**
 - Méthode exacte: Par évaluation de toutes les partitions possibles(non réalisable)
 - Méthodes approchées/heuristiques:
 - ✓ K-moyennes(centres mobiles): un cluster est représenté par son centre de gravité
 - ✓ K-medoids: un cluster est représenté par son élément le plus représentatif

85

Clustering: Méthodes hiérarchiques

- **Organisation en structure d'arbre :**
Les clusters sont organisés sous forme d'un **arbre** illustrant les relations entre les différents groupes.
- **Choix d'un critère de distance et d'une stratégie d'agrégation ou de division :**
La qualité du regroupement dépend du **critère de distance** (ex : distance euclidienne, de Manhattan...) et de la **stratégie de liaison** choisie (ex : linkage simple, linkage complet, linkage moyen).
- **Complexité algorithmique :**
Tous les algorithmes de classification hiérarchique présentent **au moins une complexité de $O(N^2)$** , où N est le nombre d'observations.
- **Résultat : un dendrogramme :**
Le résultat est un **dendrogramme**, un arbre représentant **les relations d'inclusion** entre les clusters, permettant de visualiser les différentes étapes de fusion ou de division. Les **nœuds** représentent des **fusions** (ou **divisions**) de groupes d'objets

Clustering: Méthodes hiérarchiques

Les principales stratégies de liaison en classification hiérarchique :

Lien simple (single linkage) :

- La distance entre deux clusters est définie comme la distance entre les deux objets les plus proches.
- Tendance à former des chaînes (effet "chaining"), ce qui peut produire des clusters allongés.

Lien complet (complete linkage) :

- La distance entre deux clusters est définie comme la distance entre les deux objets les plus éloignés.
- Produit des clusters compacts et bien séparés, mais est sensible aux anomalies.

Lien des moyennes (average linkage ou UPGMA) :

- La distance entre deux clusters est la moyenne des distances entre tous les couples d'objets (un dans chaque cluster).
- Peut lisser les effets extrêmes, mais a une faible résistance aux anomalies.

Lien moyen:

Distance moyenne de toutes les distances entre un membre d'un cluster et un membre de l'autre (Plus difficile à calculer)

Clustering: Méthodes hiérarchiques

Deux grandes familles d'approches hiérarchiques :

•Approche ascendante (agglomérative) :

- On commence avec chaque objet dans un cluster séparé.
- À chaque étape, on fusionne les deux clusters les plus proches.
- Le processus continue jusqu'à obtenir un seul cluster contenant tous les exemples.

•Approche descendante (divisive) :

- Tous les objets sont initialement regroupés dans un unique cluster.
- À chaque étape, le cluster est divisé en sous-groupes.
- Le processus continue jusqu'à atteindre un nombre suffisant de niveaux ou des clusters contenant un seul élément.

Clustering: méthodes à base de densité

Principe général :

- Utilisation de la densité au lieu de la distance :

- Les clusters sont définis comme des **régions** de l'espace contenant une **forte densité** de points.

- Définition d'un point dense :

- Un point est considéré comme **dense** si le **nombre de ses voisins** dépasse un **seuil** prédéfini.

- Définition de voisinage :

- Un point est considéré comme **voisin** d'un autre point s'il est **situé à une distance inférieure** à une **valeur fixée** (par exemple, un rayon ϵ).

89

Clustering: méthodes à base de grille

Principe général :

- Découpage de l'espace d'observation :

- ✓ L'espace est divisé en cellules suivant une grille régulière.
- ✓ *Exemple* : Pour des données 2D (poids/taille), chaque cellule fait $10\text{kg} \times 5\text{cm}$.

- Travail sur les cellules :

- ✓ On compte combien de points tombent dans chaque cellule.
- ✓ On ignore les cellules vides.

- Construction des clusters :

- ✓ Les clusters sont formés en assemblant les cellules voisines selon un critère de proximité (distance entre cellules).

90