

The background is a dark, textured surface with faint, light-colored sketches of various scientific and mathematical concepts. These include a globe in the upper left, a large 'V' shape, a microscope, a stack of books, a cross, a percentage sign, and some geometric shapes like a triangle and a circle.

Classification et clustering

Présenté par :

Abderrahmane tahiri, Oussama Hdiddou, Ayoub Lamrabet ,
Khaoula Haddani , Ouidad Tarif et Maryeme Hassani

PLAN

- INTRODUCTION

- Classification
- Types d'algorithmes de classification
- Les applications de la classification
- Clustering
- Types d'algorithmes de clustering
- Les applications de clustering

- Processus de CAH

- L' Apprentissage
- L' Evaluation et selection

- Partitionnement – Clustering

- Hierarchique
- Non Hierarchique

- Conclusion

Introduction

- Les algorithmes d'apprentissage automatique sont généralement classés en fonction du type de variable de sortie et du type de problème à résoudre. Ces algorithmes sont largement divisés en trois types, à savoir la régression, le regroupement et la classification. La régression et la classification sont des types d'algorithmes d'apprentissage supervisé, tandis que le regroupement est un type d'algorithme non supervisé.
- Lorsque la variable de sortie est continue, il s'agit d'un problème de régression, alors que lorsqu'elle contient des valeurs discrètes, il s'agit d'un problème de classification. Les algorithmes de clustering sont généralement utilisés lorsque nous avons besoin de créer des clusters basés sur les caractéristiques des points de données.

La Classification

- La classification est un type d'algorithme d'apprentissage automatique supervisé. Pour toute entrée donnée, les algorithmes de classification aident à prédire la classe de la variable de sortie. Il peut y avoir plusieurs types de classifications, comme la classification binaire, la classification multiclass, etc. Cela dépend du nombre de classes dans la variable de sortie.
- Les techniques de classification permettent de faire des prédictions sur la catégorie des valeurs cibles à partir de n'importe quelle entrée fournie. Habituellement, le terme "classification" est utilisé pour décrire la modélisation prédictive dans laquelle l'annotation de l'échantillon est définie. En outre, vous pouvez utiliser un algorithme de classification pour attribuer chaque point de données à une classe particulière

Les algorithmes de la classification

Régression logistique :

C'est l'un des modèles linéaires qui peut être utilisé pour la classification.

Il

utilise la fonction sigmoïde pour calculer la probabilité qu'un certain événement se produise. Il s'agit d'une méthode idéale pour la classification de variables binaires.

K-Nearest Neighbours (kNN):

Il utilise des mesures de distance telles que la distance euclidienne, la distance de Manhattan, etc. pour calculer la distance entre un point de données et tous les autres points de données. Pour classer le résultat, il prend un vote majoritaire des k voisins les plus proches de chaque point de données.

L'arbre de décision

Il s'agit d'un modèle non linéaire qui surmonte certains des inconvénients des algorithmes linéaires comme la régression logistique. Il construit le modèle de classification sous la forme d'une structure arborescente qui comprend des nœuds et des feuilles. Cet algorithme implique de multiples instructions if-else qui aident à décomposer la structure en structures plus petites et fournissent finalement le résultat final. Il peut être utilisé pour les problèmes de régression et de classification

Les algorithmes de la classification

Random Forest :

- Il s'agit d'une méthode d'apprentissage d'ensemble qui implique plusieurs arbres de décision pour prédire le résultat de la variable cible. Chaque arbre de décision fournit son propre résultat. Dans le cas d'un problème de classification, on prend le vote majoritaire de ces multiples arbres de décision pour classer le résultat final. Dans le cas du problème de régression, il prend la moyenne des valeurs prédites par les arbres de décision.

Naïve Bayes :

- Il s'agit d'un algorithme basé sur le théorème de Bayes. Il suppose qu'une caractéristique particulière est indépendante de l'inclusion d'autres caractéristiques, c'est-à-dire qu'elles ne sont pas corrélées les unes aux autres. Il ne fonctionne généralement pas bien avec les données complexes en raison de cette hypothèse, car dans la plupart des ensembles de données, il existe une sorte de relation entre les caractéristiques.

Les applications de la classification

- Jusqu'à présent, il est connu que la classification des données est un processus d'exploration de données qui aide à catégoriser les éléments en les assignant à des catégories ou classes cibles. Par conséquent, dans toute circonstance où une énorme quantité de données doit être catégorisée, afin de faciliter toute tâche, la classification est appliquée.
- Autres domaines d'application :
 1. Détection des spams par email.
 2. La reconnaissance faciale.
 3. Identifier si le client va se désabonner ou non.
 4. Approbation de prêts bancaires

Clustering

- Le clustering est un type d'algorithme d'apprentissage automatique non supervisé. Il est utilisé pour regrouper les points de données ayant des caractéristiques similaires en clusters. Idéalement, les points de données d'un même cluster devraient présenter des propriétés similaires et les points de différents clusters devraient être aussi dissemblables que possible.
- Le clustering se divise en deux groupes: le hard clustering et le soft clustering. Dans le cas du hard clustering, le point de données est attribué à l'un des clusters uniquement, tandis que dans le cas du soft clustering, il fournit une probabilité qu'un point de données se trouve dans chacun des clusters.

Les algorithmes de clustering

K-Means Clustering :

KMC

- Il initialise un nombre prédéfini de k clusters et utilise des mesures de distance pour calculer la distance de chaque point de données par rapport au centroïde de chaque cluster. Il affecte les points de données à l'un des k clusters en fonction de leur distance.

Le clustering hiérarchique agglomérant

Approche ascendant

- Elle considère chaque point de données comme un cluster et fusionne ces points de données sur la base de la métrique de distance et du critère utilisé pour relier ces clusters.

Le clustering hiérarchique divisée

Approche descendante

- Elle s'initialise avec tous les points de données comme un seul cluster et divise ces points de données sur la base de la métrique de distance et du critère. Le clustering agglomérant et diviseur peut être représenté sous la forme d'un dendrogramme et le nombre de clusters à sélectionner en se référant au même

Les applications de clustering

- Les applications du clustering sont vastes par nature. Précisément dans l'exploration de données, le clustering est utilisé comme processus d'analyse pour déduire des images, des données et reconnaître des modèles sous-jacents dans celles-ci.
- Comme exemple :
 - services de streaming qui effectuent souvent une analyse de clustering pour identifier les téléspectateurs qui ont un comportement et des choix de visionnement similaires .
 - des équipes sportives utilisent souvent la méthode du clustering pour identifier les joueurs ayant des traits et des caractéristiques similaires. Ils regroupent ensuite ces joueurs pour constituer une équipe plus efficace.
 - Segmentation de la base de consommateurs sur le marché
 - Analyse des réseaux sociaux.
 - Systèmes de recommandation .

Différence entre classification et clustering

- **Type** : - Le clustering est une méthode d'apprentissage non supervisée alors que la classification est une méthode d'apprentissage supervisée.
- **Étapes** : - Le processus de classification comporte deux étapes : la formation et le test. Le processus de clustering n'implique que le regroupement des données.
- **Complexité** : - Comme la classification comporte un plus grand nombre d'étapes, la complexité des algorithmes de classification est plus élevée que celle des algorithmes de clustering dont le but est uniquement de regrouper les données.
- **Processus** : - Dans le clustering, les points de données sont regroupés en clusters en fonction de leurs similarités. Ainsi, ici, les instances sont classées en fonction de leur ressemblance et sans aucune étiquette de classe. La classification consiste à classer les données d'entrée dans l'une des étiquettes de classe de la variable de sortie. Par conséquent, elle peut être définie comme une approche de la classification des instances d'entrée sur la base de leurs étiquettes de classe associées.

Conclusion

- La méthodologie de la classification et du clustering est différente, et les résultats attendus de leurs algorithmes le sont également. En bref, la classification et le clustering sont tous deux utilisés pour résoudre des problèmes différents. Cet article a fourni une brève introduction à la classification et au clustering.
- Le clustering et la classification sont deux méthodes d'apprentissage automatique efficaces qui contribuent à l'amélioration des processus métier. Bien que ces processus soient identiques, vous pouvez les utiliser de manière unique pour comprendre les objectifs de vos clients et améliorer leur expérience d'achat. En examinant, en profilant et en ciblant vos consommateurs à l'aide de la classification et du clustering dans l'apprentissage automatique, vous préparerez une base de clients fiables et aurez un meilleur retour sur investissement.

The background of the slide is a dark grey collage of white line-art icons. These icons include a globe, a stack of books, a microscope, a test tube, a pencil, a ruler, a compass, a protractor, a calculator, and various geometric shapes like triangles and circles. The icons are scattered across the background, with some appearing more prominent than others.

Processus de CAH

- L' Apprentissage –algorithmes -
- L' Evaluation et selection

L'apprentissage -algorithms

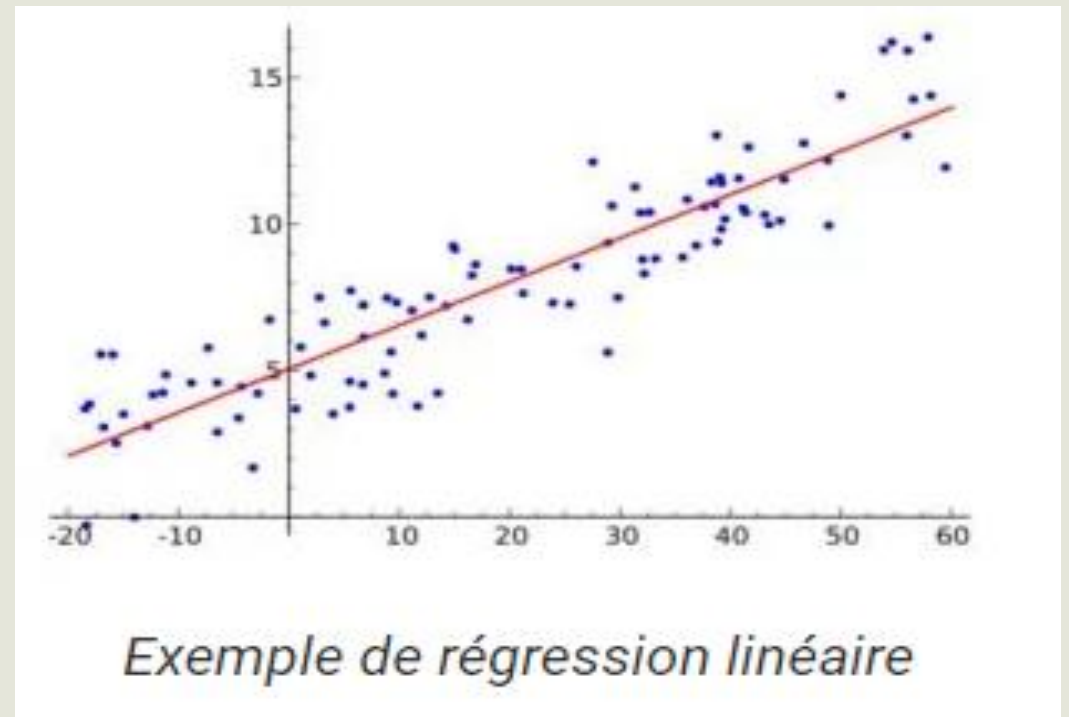
La régression linéaire

OBJECTIFS :

La régression linéaire est une technique statistique de modélisation des relations entre des différentes variables (dépendantes et indépendantes).

Utilisée pour décrire et analyser les valeurs ou les données, la régression linéaire a pour objectif de réaliser des prédictions ou des prévisions.

Alors on veut étudier l'influence d'une variable quantitative X sur une autre variable quantitative Y. La première est souvent appelée variable explicative (ou encore exogène) et la seconde est appelée variable expliquée (ou encore endogène).



L'apprentissage -algorithms

La régression linéaire

MÉTHODES DE TRAVAIL :

- Diagramme de régression
- La droite de régression
- Coefficient de corrélation linéaire

Exemple :

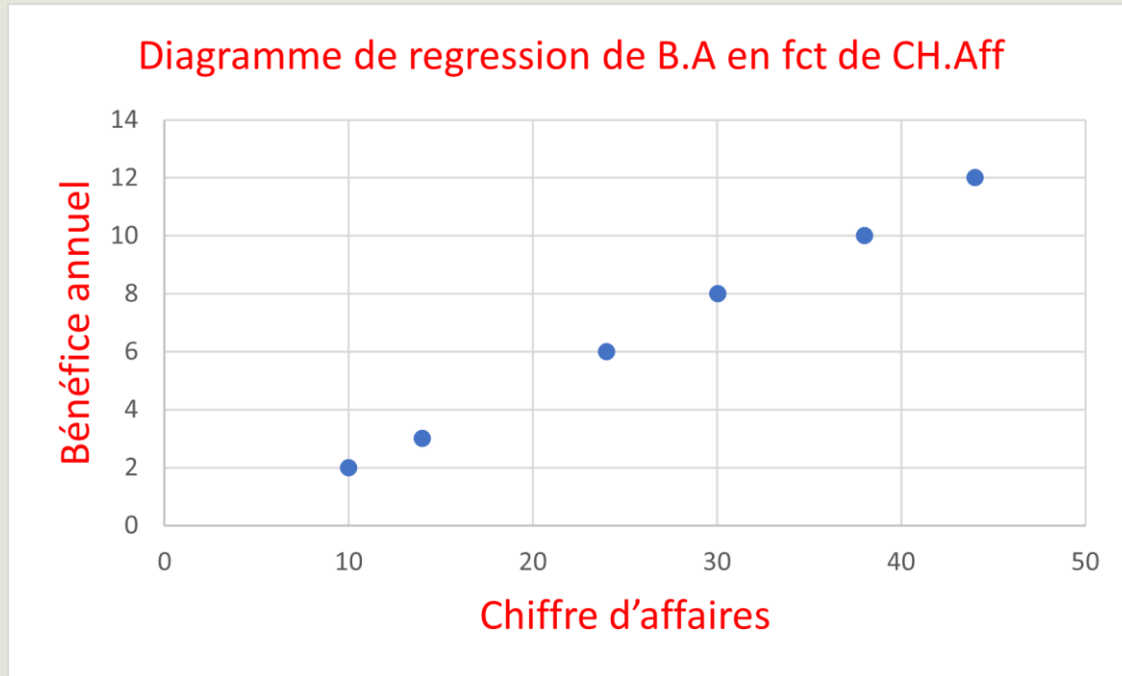
Chiffre d'affaires	10	14	24	30	38	44
Bénéfice annuel	2	3	6	8	10	12

On a réalisé une expérimentation qui consiste à étudier le chiffre d'affaires (en millions de DH) et le bénéfice annuel (en millions de DH).

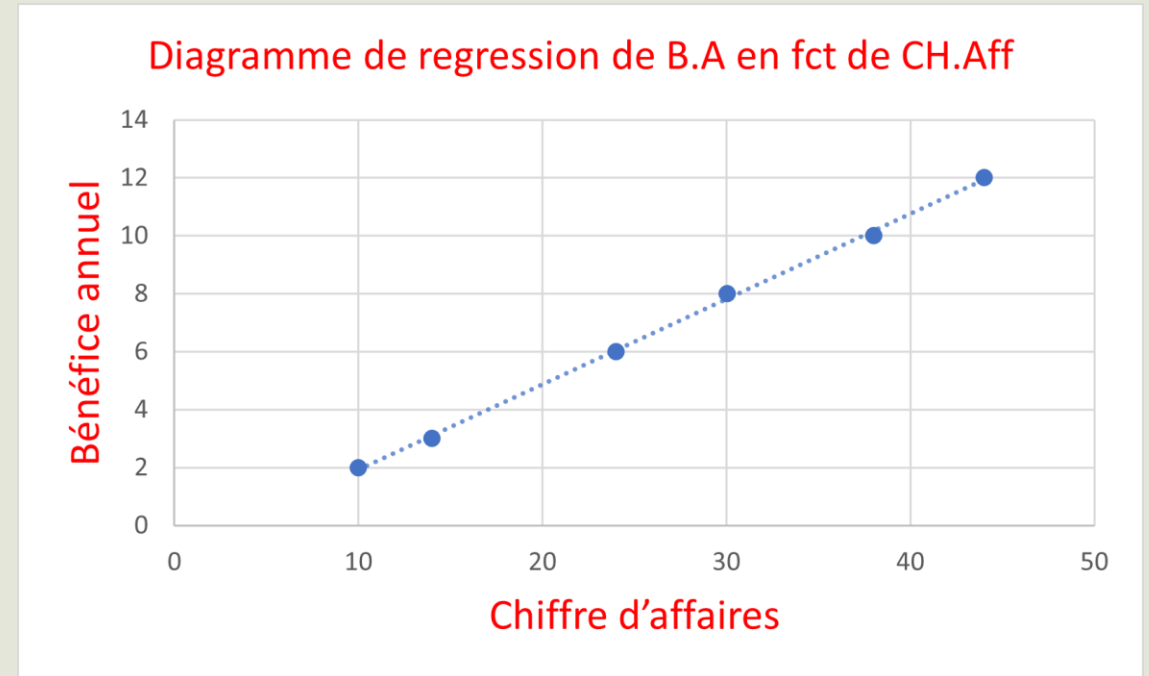
1. Existe-t-il une relation entre les deux variables ?
2. Quelle est la nature de cette relation ?
3. Quel est le degré de liaison entre les deux variables ?
4. Si on connaît le chiffre d'affaires peut on prévoir ou estimer le bénéfice annuel ?

L'apprentissage -algorithms

- Diagramme de régression :



- La droite de régression :



Le nuage de points obtenus ressemble beaucoup a une droite. On peut donc ajuster une droite d'équation : $y = ax + b$.

L'apprentissage -algorithms

Le nuage de points obtenus ressemble beaucoup a une droite. On peut donc ajuster une droite d'équation : $y = ax + b$.

L'ajustement linéaire permet donc de déterminer la droite qui s'ajuste au mieux aux valeurs observées,

Cette droite est appelée droite de régression de Y en X .

Avec :

- X représente le chiffres d'affaires.
- Y représente Bénéfice annuel.

Méthode des moindres carrés :

C'est une méthode d'ajustement qui consiste à minimiser les erreurs des différences entre les valeurs observées y et les valeurs estimées \hat{Y} .

Alors, il faut déterminer les constantes a et b

On a :

L'apprentissage -algorithms

On a :

$$a = \frac{Cov(xy)}{V(x)}$$

$$Cov(xy) = \frac{1}{N} \sum_{i=1}^k x_i y_i - \bar{x} \bar{y}$$

$$b = \bar{y} - a \bar{x}$$

cherchons a :

Xi	Yi	Xi*Yi	Xi ²	Yi ²
10	2	20	100	4
14	3	42	196	9
24	6	144	576	36
30	8	240	900	64
38	10	380	1444	100
44	12	528	1936	144
160	41	1354	5152	357

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k x_i$$

$$\bar{x} = \frac{1}{6} \times 160$$

$$\bar{x} = 26,67$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^k y_i$$

$$\bar{y} = \frac{1}{6} \times 41$$

$$\bar{y} = 6,83$$

$$Cov(xy) = \frac{1}{N} \sum_{i=1}^k x_i y_i - \bar{x} \bar{y}$$

$$Cov(xy) = \left(\frac{1}{6} \times 1354 \right) - (26,67 \times 6,83)$$

$$Cov(xy) = 225,67 - 182,16$$

$$Cov(xy) = 43,51$$

L'apprentissage -algorithms

$$\text{Et } \text{Var}(X) = \frac{1}{N} \sum_{i=1}^n (X_i - \bar{X})^2 = \overline{X^2} - (\bar{X})^2$$

$$\text{A.N } \text{Var}(X) = 858.66 - 26.67^2$$

$$\underline{\text{Var}(X) = 147.37}$$

$$\text{D'où } : \quad \underline{\mathbf{a = 0.29}}$$

Cherchons b :

$$\text{On a : } \bar{Y} = a\bar{X} + b$$

$$\text{Alors : } b = \bar{Y} - a\bar{X}$$

$$\text{A.N: } b = 6.83 - 0.29 * 26.67$$

$$\underline{\mathbf{b = -0.90}}$$

Finalement l'équation de la droite de régression est :

$$\underline{\mathbf{Y = 0.29 * X - 0.90}}$$

L'apprentissage -algorithms

Coefficient de corrélation linéaire :

Le coefficient de corrélation linéaire a pour objet de détecter et décrire la relation entre deux variables, et de mesurer l'intensité de la liaison linéaire entre les deux variables X et Y.

On a :

$$r = \frac{\sum_{i=1}^k x_i y_i - N \bar{x} \bar{y}}{\sqrt{\sum_{i=1}^k \bar{x}_i^2 - N \bar{x}^2} \sqrt{\sum_{i=1}^k y_i^2 - N \bar{y}^2}}$$
$$r = \frac{1354 - (6 \times 26,67 \times 6,83)}{\sqrt{5152 - (6 \times 26,67^2)} \sqrt{357 - (6 \times 6,83^2)}}$$
$$r = \frac{1354 - 1092,94}{29,74 \times 8,78}$$
$$r = 0,99$$

Le coefficient de corrélation linéaire possède le même signe que la covariance et il est toujours entre -1 et 1.

Le signe de coefficient de corrélation linéaire indique le sens de la relation entre X et Y :

$r > 0$: veut dire que les deux variables X et Y varient dans le même sens. On parle de Corrélation positive.

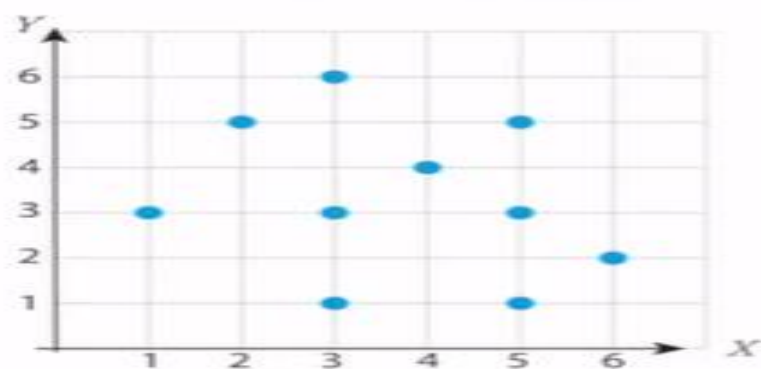
$r < 0$: veut dire que les deux variables X et Y varient en sens inverse. On parle de Corrélation négative.

$r = 1$: tous les points se trouvent sur une même droite.

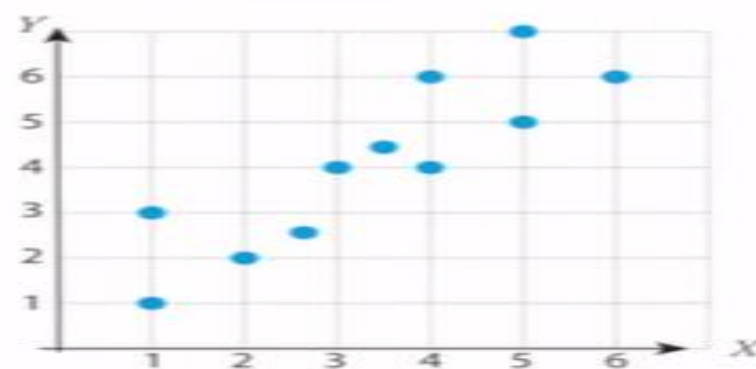
On parle de Corrélation linéaire parfaite.

$r = 0$: veut dire qu'il n'y a aucune dépendance linéaire entre les deux variables. On parle de Corrélation nulle.

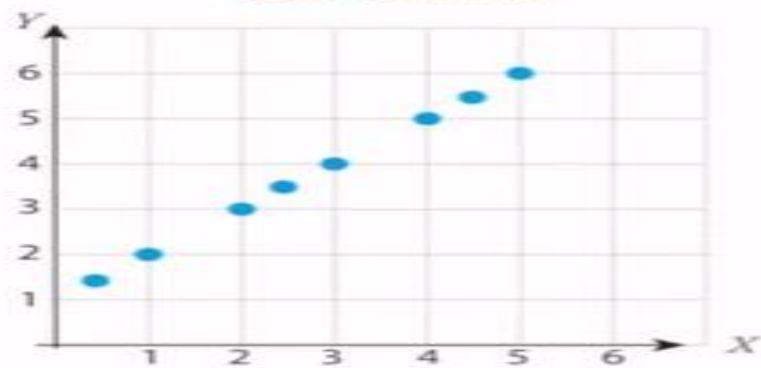
L'apprentissage -algorithms



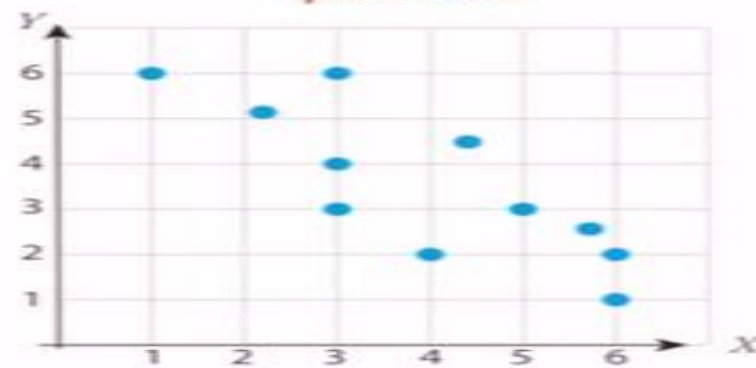
Absence de corrélation



Corrélation positive



Corrélation positive parfaite

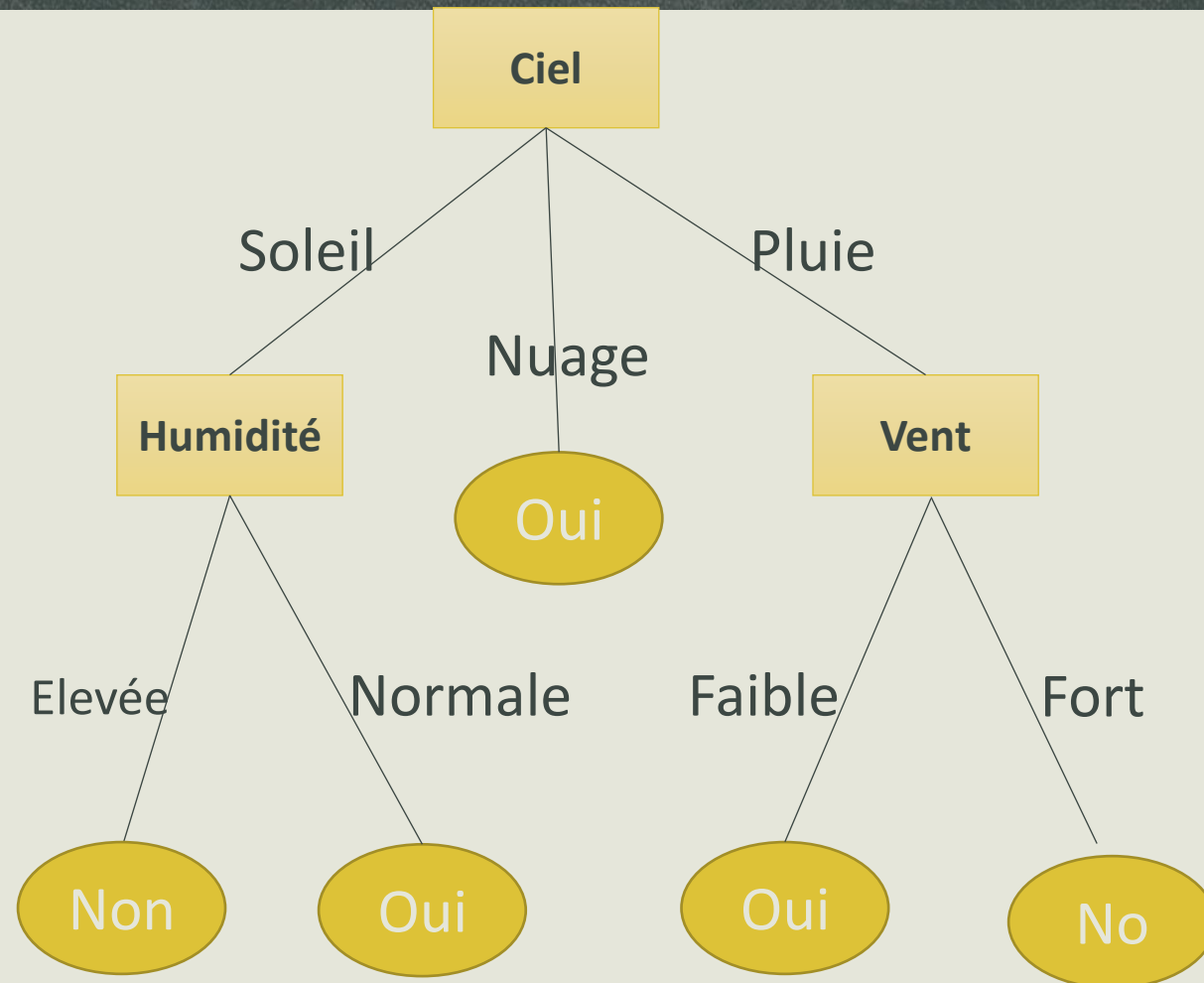


Corrélation négative

On veut estimer une décision (jouer football ou non) en se basant sur 4 caractéristiques: temps, température, humidité et vent. On va construire un arbre de décision en se basant sur les données suivants

Jour	Ciel	Température	Humidité	Vent	Jouer
J1	Soleil	Chaude	Elevée	Faible	Non
J2	Soleil	Chaude	Elevée	Fort	Non
J3	Nuage	Chaude	Elevée	Faible	Oui
J4	Pluie	Douce	Elevée	Faible	Oui
J5	Pluie	Froid	Normale	Faible	Oui
J6	Pluie	Froid	Normale	Fort	Non
J7	Nuage	Froid	Normale	Fort	Oui
J8	Soleil	Douce	Elevée	Faible	Non
J9	Soleil	Froid	Normale	Faible	Oui
J10	Pluie	Douce	Normale	Faible	Oui
J11	Soleil	Douce	Normale	Fort	Oui
J12	Nuage	Douce	Elevée	Fort	Oui
J13	Nuage	Chaude	Normale	Faible	Oui
J14	Pluie	Douce	Elevée	Fort	?

On veut estimer une décision (jouer football ou non) en se basant sur 4 caractéristiques: temps, température, humidité et vent. On va construire un arbre de décision en se basant sur les données suivants



EXEMPLE :Accorder un crédit pour acheter un PC

ID	X1:Age	X2:revenu	X3:Etudiant	X4:Credit Rating	Y:Achat d'un PC
1	Jeune	Elevé	Non	Bon	Non
2	Jeune	Elevé	Non	Mauvais	Non
3	Moyen	Elevé	Non	Bon	Oui
4	Agé	Moyen	Non	Bon	Oui
5	Agé	Faible	Oui	Bon	Oui
6	Agé	Faible	Oui	Mauvais	Non
7	Moyen	Faible	Oui	Mauvais	Oui
8	Jeune	Moyen	Non	Bon	Non
9	Jeune	Faible	Oui	Bon	Oui
10	Agé	Moyen	Oui	Bon	Oui
11	Jeune	Moyen	Oui	Mauvais	Oui
12	Moyen	Moyen	Non	Mauvais	Oui
13	Moyen	Elevé	Oui	Bon	Oui
14	Agé	Moyen	Non	Mauvais	Non
15	Moyen	Moyen	Oui	Bon	?

- # L'apprentissage –algorithms

L'arbre de décision

Définition:

- **L'arbre de décision** est une méthodes de [prise de décision](#) la plus efficace. Elle permet non seulement de présenter visuellement des informations mais aussi de les hiérarchiser. C'est un outil qui facilite grandement nos décisions et limite le sentiment de [surcharge informationnelle](#).
- Le nœud le plus haut représente la racine de l'arbre de décision
- Les nœuds interne (les nœuds qui ont des descendants (ou *enfants*))représente les attribues
- La branche (l'arc) représente une règle de décision (les modalités)
- La feuille (Nœuds *terminaux* : nœuds qui n'ont pas de descendant) représente la décision

Les étapes pour construire l'arbre de décision

L'algorithme général de création d'un arbre de décision:

- Déterminer la meilleure caractéristique dans l'ensemble de données d'entraînement.
- Diviser les données d'entraînement en sous-ensembles contenant les valeurs possibles de la meilleure caractéristique.
- Générez de manière récursive de nouveaux arbres de décision en utilisant les sous-ensembles de données créés.
- Lorsqu'on ne peut plus classifier les données, on s'arrête.

- **L'apprentissage –algorithms**
L'arbre de décision

Algorithme ID3

- ID3 construit l'arbre de décision récursivement
- A chaque étape de la récursivité , on calcule parmi les attributs restant pour la branche en cours, celui qui maximisera le gain d'information
- Le calcul se fait à base de l'entropie de Shannon déjà présentée.
- L'algorithme suppose que tous les attributs sont catégoriels ; Si des attributs sont numériques, ils doivent être discrétisés pour pouvoir l'appliquer.

- L'apprentissage –algorithms
L'arbre de décision

Notion d'entropie

L'entropie est la quantité d'information de la distribution S qu'elle peut apporter. Elle est donnée par l'équation ([Entropie de Shannon](#)) :

$$E(Y) = \sum_{i=1}^k -p_i \log_2(p_i) = \sum_{i=1}^k \frac{-p_i \ln(p_i)}{\ln(2)}$$

Avec:

k : le nombre de classe c_i de l'attribue cible Y

p_i : la probabilité de la classe c_i ($p_i = \frac{c_i}{n_i}$)

➤ En générale on a :

$$E = - \sum_{i=1}^k \frac{c_i}{n_i} \log_2 \left(\frac{c_i}{n_i} \right)$$

Nombre de classes dans notre cas 2

Nombre d'éléments de chaque classe

Nombre total de valeurs

- L'apprentissage –algorithms
L'arbre de décision

- Soit un échantillon de taille N pour une modalité a_i , composé de N_1 «Oui» et N_2 «Non». $N=N_1+N_2$

on pose $p=N_1/N$, et $q=N_2/N$, $p+q=1$

$$E(a_i) = E(N_1, N_2) = \frac{-p \ln(p) - q \ln(q)}{\ln(2)}$$

- On utilise la notion d'entropie pour calculer l'homogénéité d'un échantillon
Si l'échantillon est homogène ($N_1=0$ ou $N_2=0$) : $E(N_1, N_2)=0$
Si l'échantillon est divisé en 2 sous-échantillon égaux ($N_1=N_2$) : $E(N_1, N_2)=1$
- $0 \leq E(N_1, N_2) \leq 1$

- L'apprentissage –algorithms
L'arbre de décision

- **Gain d'information**

Formule mathématique:

$$\text{Gain}(X_j) = E(Y) - E(Y, X_j)$$

Remarque :

La construction d'un arbre de décision consiste à trouver les attributs qui donnent le gain d'information le plus élevés (les branches les plus homogènes)

- **Calcul de l'entropie**

Pour construire un arbre de décision, on doit calculer 2 types d'entropie

Entropie 1 : en utilisant le tableau de fréquence d'un seul attribut (Y) on l'appelle entropie de la cible notée $E(Y)$

Entropie 2 : en utilisant le tableau de fréquence de deux attributs (tableau de contingence) on l'appelle entropie bivariée notée $E(Y, X)$

- L'apprentissage –algorithms
L'arbre de décision

- **Entropie 1** : en utilisant le tableau de fréquence d'un seul attribut (Y)

$$p_1 = p = \frac{N_1}{N} = \frac{9}{14} = 0,64$$

$$p_2 = p = \frac{N_1}{N} = \frac{5}{14} = 0,36$$

$$E(Y) = E(9,5) = \frac{-p \ln(p) - q \ln(q)}{\ln(2)} = 0,94$$

Notation : On note aussi : $E(Y) = \ln fo(Y)$

Achat d'un PC	
C1=Oui	C2=Non
9	5

- **Entropie 2** : en utilisant le tableau de fréquence de deux attributs

Soit Y la variable cible

X_i : l'attribut explicatif ayant k modalités

i : le compteur des modalités

a_i : la modalités de la variable X_i

N_i : le nombre d'observation ayant a_i dans l'échantillon

$$E(Y, X_i) = \sum_{i=1}^k p(a_i) E(a_i)$$

$$p(a_i) = \frac{N_i}{N}$$

- L'apprentissage –algorithms

L'arbre de décision

Y : Acheter un PC

X₁ : Age ayant 3 modalités

a₁: Jeune ; a₂: Moyen ; a₃:Agé

	Modalités	Y=Oui	Y=Non	Total	P(ai)	E(ai)
X1(Age)	a ₁ : Jeune	2	3	5	5/14	E(2,3)=0,971
	a ₂ :Moyen	4	0	4	4/14	E(4,0)=0
	a ₃ :Agé	3	2	5	5/14	E(3,2)=0,971
Totale		9	5	14	1	

$$\begin{aligned}
 E(Y, X_1) &= P(a_1)E(a_1) + P(a_2)E(a_2) + P(a_3)E(a_3) \\
 &= P(\text{Jeune})E(\text{Jeune}) + P(\text{Moyen})E(\text{Moyen}) + P(\text{Agé})E(\text{Agé}) \\
 &= \frac{5}{14}0,971 + \frac{4}{14}0 + \frac{5}{14}0,971 = 0,639 \quad 1 \\
 \text{Gain}(X_1=\text{Age}) &= E(Y) - E(Y, X_1) = 0,94 - 0,639 = 0,247
 \end{aligned}$$

- L'apprentissage –algorithms
L'arbre de décision

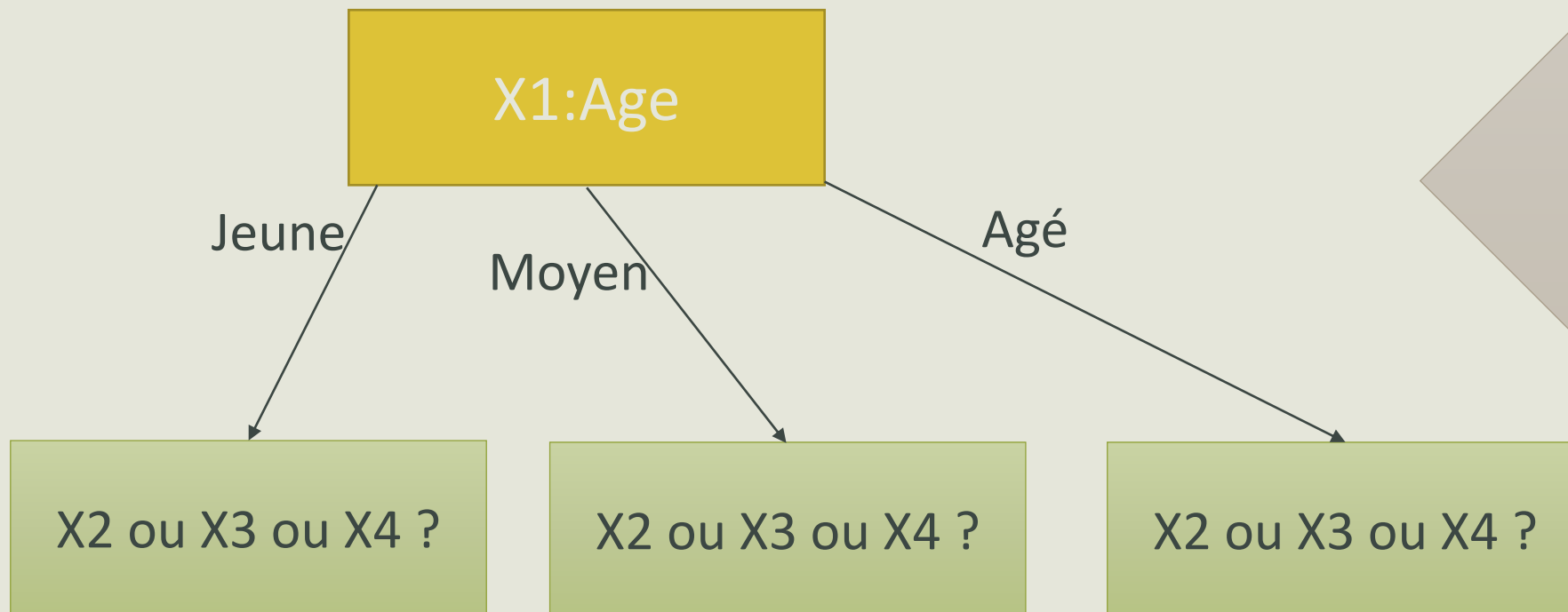
Tableau de calcul des entropies et des gains correspondants a chaque attribut explicatif

Attributs(X_j)	Modalités(a_i)	Achat d'un PC		Total	$P(a_i)$	$E(a_i)=E(n_1, n_2)$	$E(Y, X_i)$	Gain(X_i)
		Oui(n_1)	Non(n_2)					
X1:Age	Jeune	2	3	5	5/14	0,971	0,694	0,248
	Moyen	4	0	4	4/14	0		
	Agé	3	2	5	5/14	0,971		
X2:Revenu	Elevé	2	2	4	4/14	1	0,914	0,029
	Moyen	4	2	6	6/14	0,918		
	Faible	3	1	4	4/14	0,811		
X3:Etudiant	Oui	6	1	7	7/14	0,592	0,791	0,151
	Non	3	4	7	7/14	0,985		
X4:Credit Rating	Bon	6	2	8	8/14	0,811	0,895	0,048
	Mauvais	3	3	6	6/14	1		
Totale		N1=9	N2=5	14	1	$E(Y)=0.94$		

L'apprentissage –algorithms

L'arbre de décision

Le choix du nœud racine

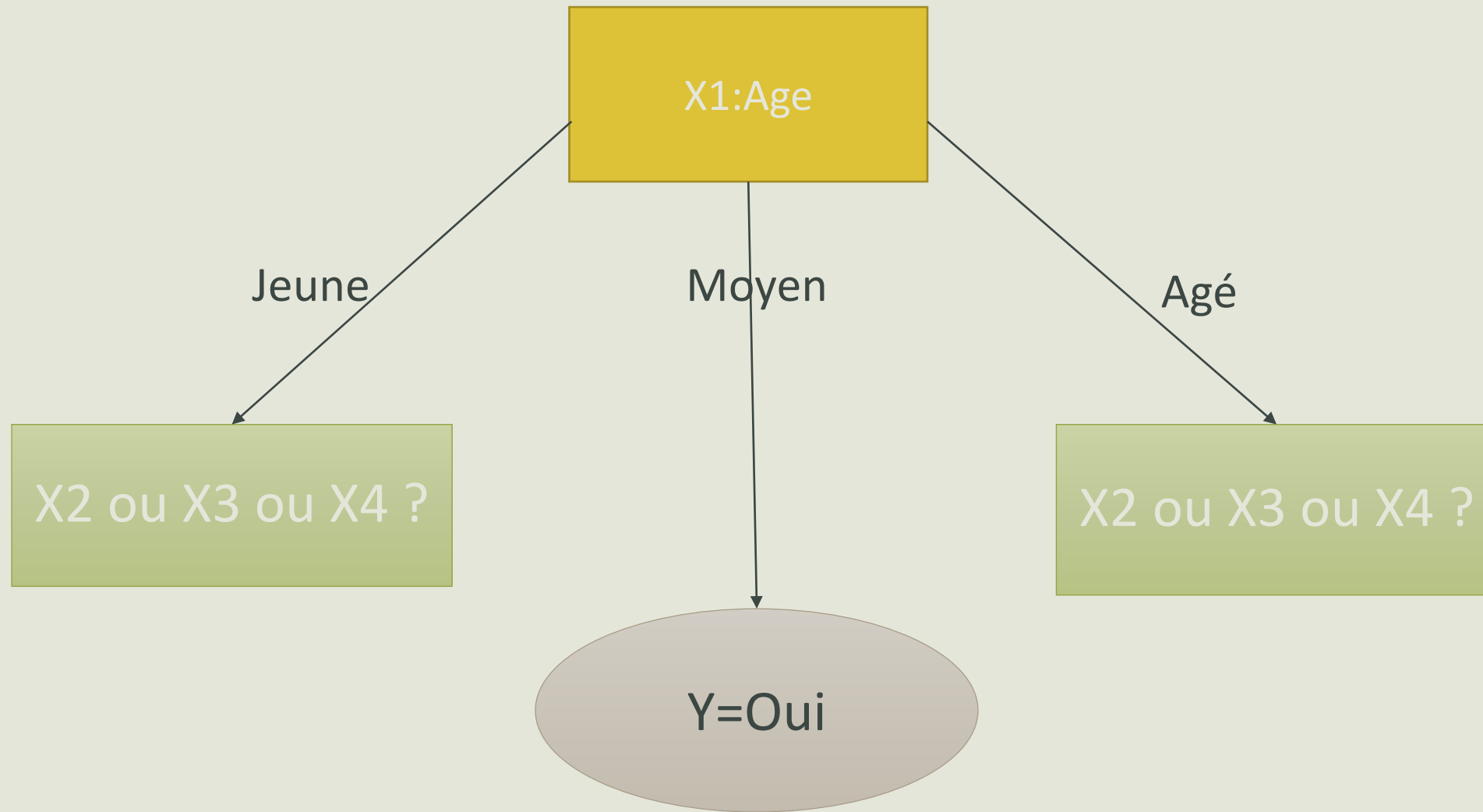


L'attributs X1=Age est choisi comme nœud racine de l'arbre de décision puisque il détient le gain le plus élevé par rapport au autres attributs

Tableau de calcul des entropies et des gains correspondants a chaque attribut explicatif

Age=Jeune				Age=Moyen				Age=Agé			
X2	X3	X4	Y	X2	X3	X4	Y	X2	X3	X4	Y
Elevé	Non	Bon	Non	Elevé	Non	Bon	Oui	Moyen	Non	Bon	Oui
Elevé	Non	Mauvais	Non	Faible	Oui	Mauvais	Oui	Faible	Oui	Bon	Oui
Moyen	Non	Bon	Non	Moyen	Non	Mauvais	Oui	Faible	Oui	Mauvais	Non
Faible	Oui	Bon	Oui	Elevé	Oui	Bon	Oui	Moyen	Oui	Bon	Oui
Moyen	Oui	Mauvais	Oui					Moyen	Non	Mauvais	Non

Positionner la feuille (Moyen) vue que tous les Y sont de classe «Oui»



- L'apprentissage –algorithms
L'arbre de décision

Tableau de calcul des entropies et des gains correspondants a chaque attribut explicatif sachant que $X1(\text{Age})=\text{Jeune}$

Achat d'un PC / $X1=\text{Jeune}$	
C1=Oui	C2=Non
N1=2	N2=3
$E(Y)=E(2,3)=0,971$	

- L'apprentissage –algorithms
L'arbre de décision

Tableau de calcul des entropies et des gains correspondants a chaque attribut explicatif sachant que $X1(\text{Age})=\text{Jeune}$

Attributs(X_j)	Modalités(a_i)	Achat d'un PC		Total	$P(a_i)$	$E(a_i)=E(n_1, n_2)$	$E(Y, X_i)$	Gain(X_i)
		Oui(n_1)	Non(n_2)					
X2:Revenu	Elevé	0	2	2	2/5	0	0,4	0,571
	Moyen	1	1	2	2/5	1		
	Faible	1	0	1	1/5	0		
X3:Etudiant	Oui	2	0	2	2/5	0	0	0,971
	Non	0	3	3	3/5	0		
X4:Credit Rating	Bon	1	1	2	2/5	1	0,895	0,048
	Mauvais	2	1	3	3/5	0,918		
Totale		N1=2	N2=3	5	1	$E(Y)=0,971$		

- L'apprentissage –algorithms
L'arbre de décision

Tableau de calcul des entropies et des gains correspondants a chaque attribut explicatif sachant que $X1(\text{Age})=\text{Agé}$

Achat d'un PC / $X1=\text{Agé}$	
C1=Oui	C2=Non
N1=3	N2=2
$E(Y)=E(3,2)=0,971$	

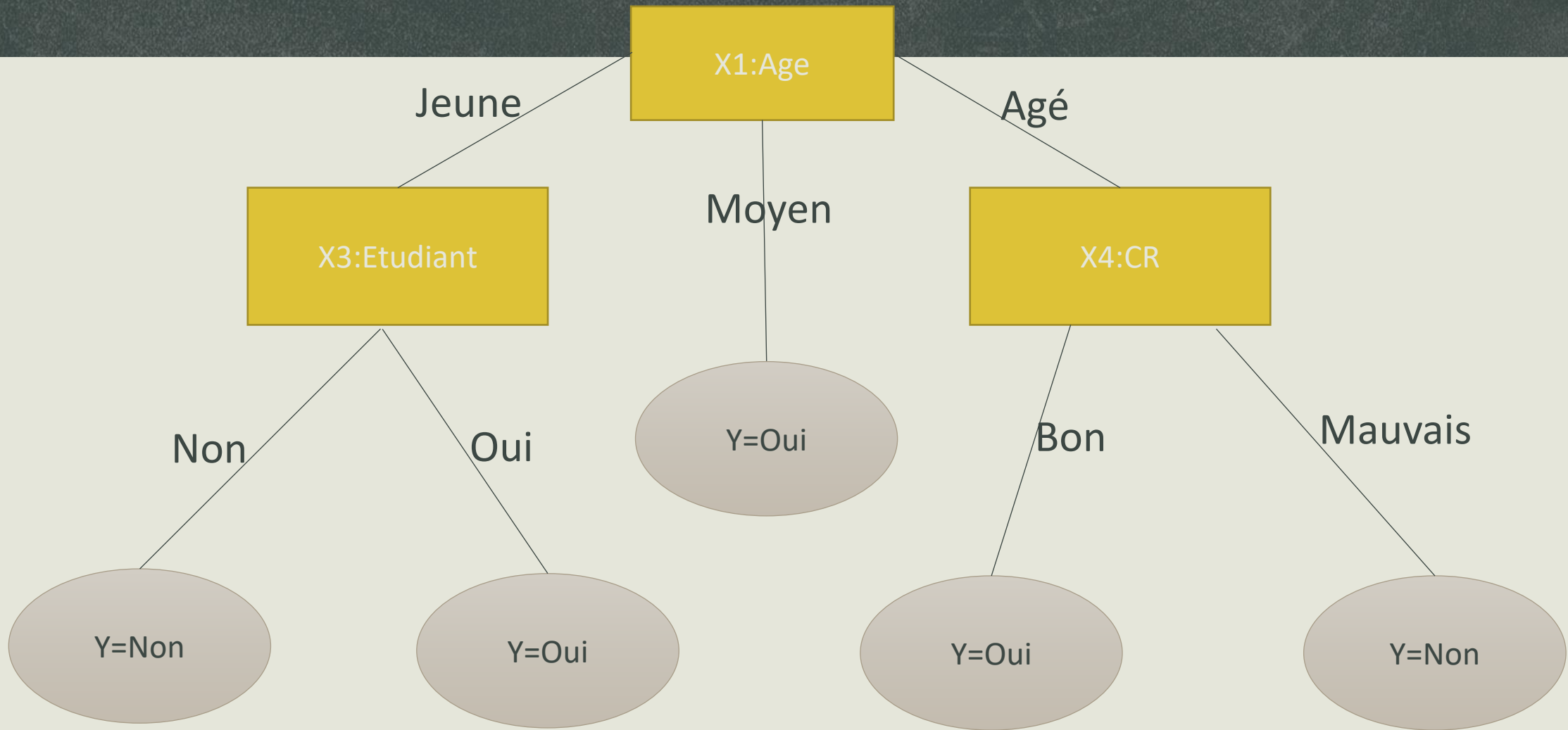
- L'apprentissage –algorithms

L'arbre de décision

Tableau de calcul des entropies et des gains correspondants a chaque attribut explicatif sachant que $X1(\text{Age})=\text{Agé}$

Attributs(X_j)	Modalités(a_i)	Achat d'un PC		Total	$P(a_i)$	$E(a_i)=E(n_1, n_2)$	$E(Y, X_i)$	Gain(X_i)
		Oui(n_1)	Non(n_2)					
X2:Revenu	Elevé	0	0	0	0/5	0	0,95	0,02
	Moyen	2	1	3	3/5	0,918		
	Faible	1	1	2	2/5	1		
X3:Etudiant	Oui	2	1	3	3/5	0,918	0,95	0,02
	Non	1	1	2	2/5	1		
X4:Credit Rating	Bon	3	0	3	3/5	0	0	0,971
	Mauvais	0	2	2	2/5	0		
Totale		N1=3	N2=2	5	1	$E(Y)=0,971$		

Arbre de décision final



- L'apprentissage –algorithms

k -Nearest Neighbors

Objectif :

L'algorithme K-plus proche voisin ou K-NN est un discriminant d'apprentissage supervisé non paramétrique, crée essentiellement une frontière imaginaire pour classer les données ; on calculant la probabilité que le donné de test appartient aux classes des données d'apprentissage "K" .

Comment fonctionne le K-NN ?

Le fonctionnement de K-NN peut être expliqué sur la base de l'algorithme ci-dessous :

Étape 1 : Sélectionner le nombre K de voisins.

Étape 2 : Calculer la distance euclidienne du nombre K de voisins.

Étape 3 : Prendre les K voisins les plus proches selon la distance euclidienne calculée.

Étape 4 : Parmi ces K voisins, compter le nombre de points de données dans chaque catégorie.

Étape 5 : Attribuez les nouveaux points de données à la catégorie pour laquelle le nombre de voisins est maximal.

• L'apprentissage –algorithms

k -Nearest Neighbors

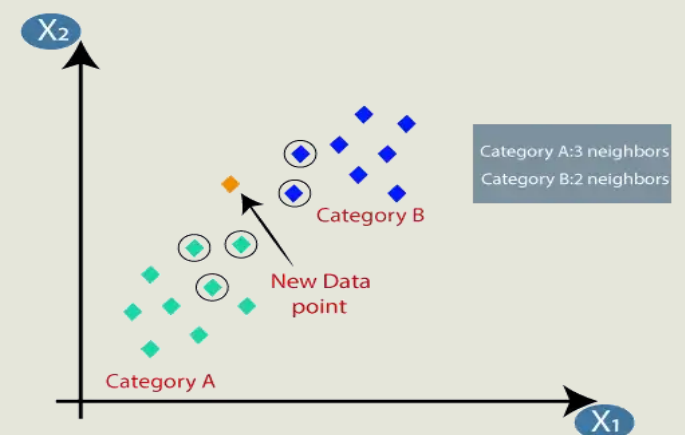
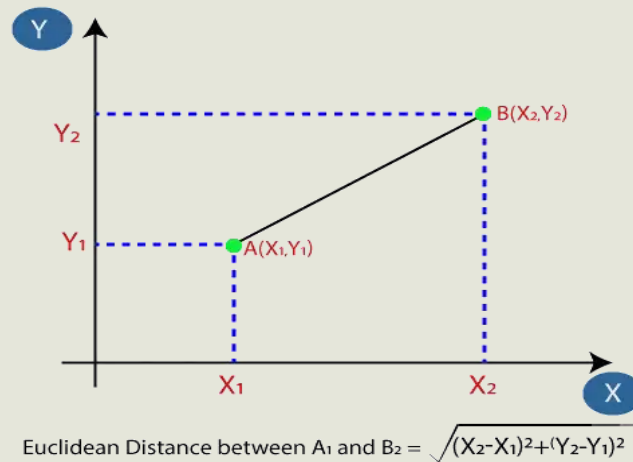
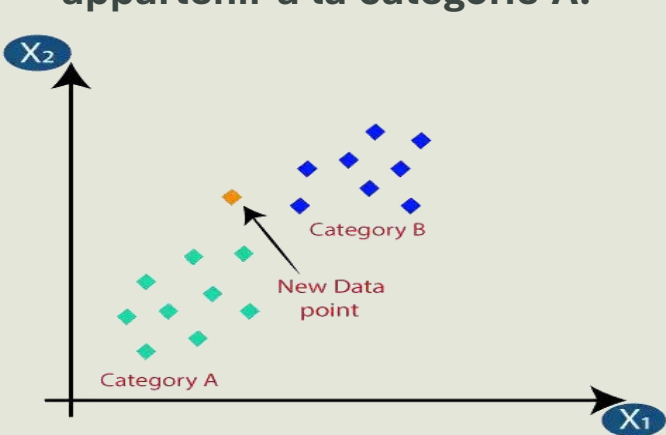
Supposons que nous ayons un nouveau point de données et que nous devons le placer dans la catégorie requise. Considérons la représentation ci-dessous :

-Tout d'abord, nous allons choisir le nombre de voisins, nous choisirons donc le $k=5$.

-Ensuite, nous allons calculer la distance euclidienne entre les points de données.

- En calculant la distance euclidienne, nous obtenons les plus proches voisins, soit trois plus proches voisins dans la catégorie A et deux plus proches voisins dans la catégorie B. Considérons l'image ci-dessous :

Comme nous pouvons le voir, les 3 plus proches voisins sont de catégorie A, donc ce nouveau point de données doit appartenir à la catégorie A.



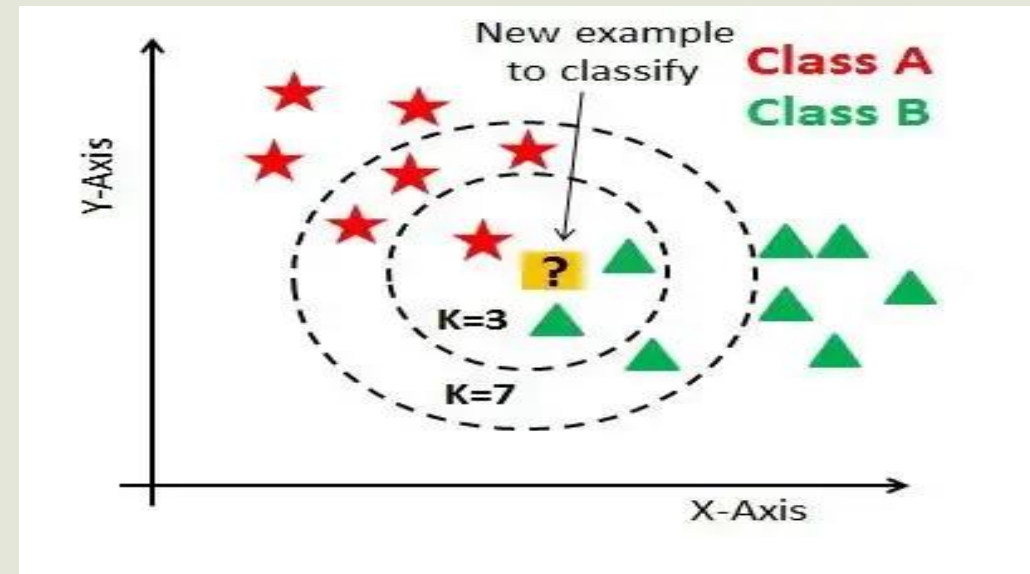
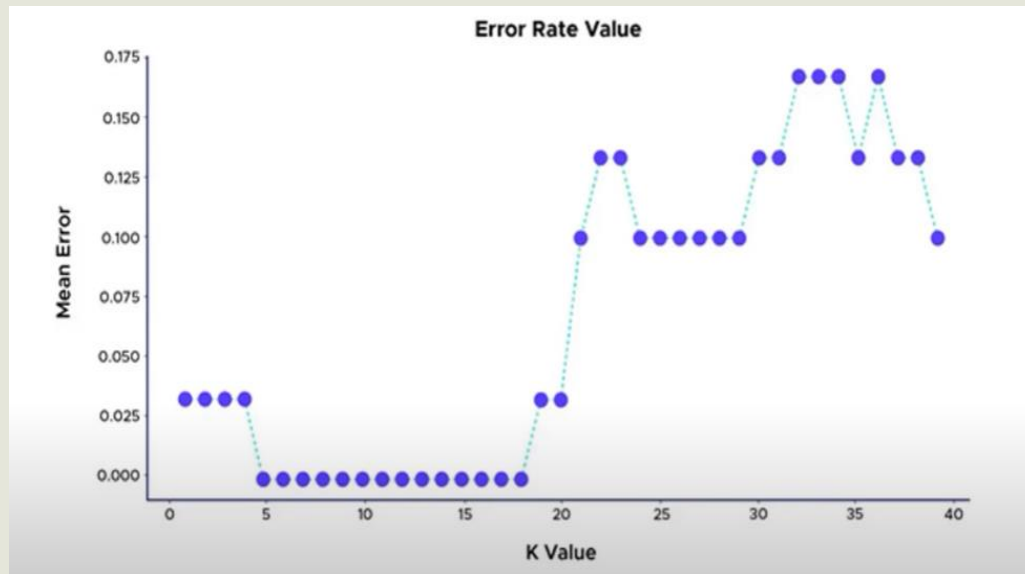
- L'apprentissage –algorithms

- k -Nearest Neighbors

Alors comment choisir la valeur optimale de K ?

Il n'existe pas de méthodes statistiques prédéfinies pour trouver la valeur la plus favorable de K.

On initialise une valeur K aléatoire et on commence à calculer. A condition, Qu' on Choisissez une valeur de K qui présente un taux d'erreur minimal.



- L'apprentissage –algorithms

k -Nearest Neighbors

Calcul de distance :

Il existe plusieurs méthodes pour calculer cette distance, dont les plus connues sont la distance euclidienne, la distance de Manhattan (pour les données continues)...

- Distance euclidienne : $d(\mathbf{p}, \mathbf{q}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$

- Distance de Manhattan : $d(\mathbf{p}, \mathbf{q}) = \left(\sum_{i=1}^m |p_i - q_i| \right)$

- L'apprentissage –algorithms

k -Nearest Neighbors

Exercice d application :

'k-NN en classification' Avec k=4

PERSONNE	NBR DE CIGARETTE	POIDS	PROP.CRISE CARDIAQUE	DISTANCE EUCLIDIENNE
A	7	70	affecté	$\sqrt{(3 - 7)^2 + (70 - 70)^2} = 4.00$
B	7	40	affecté	$\sqrt{(3 - 7)^2 + (70 - 30)^2} = 30.27$
C	3	40	non affecté	$\sqrt{(3 - 3)^2 + (70 - 40)^2} = 30.00$
D	1	40	non affecté	$\sqrt{(3 - 1)^2 + (70 - 40)^2} = 30.07$
E	3	70	?	

- L'apprentissage –algorithms

k -Nearest Neighbors

Exercice d application :

'k-NN en régression ' Avec k= 3

AGE	LOAN	HOUSE PRICE INDEX	DISTANCE	.
25	40.000\$	135	102000	
35	60.000\$	256	82000	
45	80.000\$	231	62000	
20	20.000\$	267	122000	
35	120.000\$	139	22000	2
52	18.000\$	150	124000	
23	95.000\$	127	47000	
40	62.000\$	216	80000	
60	100.000\$	139	42000	3
48	220.000\$	250	78000	
33	150.000\$	264	8000	1
48	142.000\$?		

- L'apprentissage –algorithms

k -Nearest Neighbors

Conclusion :

Nous avons appris que :

- K-NN est facile à mettre en œuvre car il nécessite peu d'hyperparamètres
- K-NN étant un algorithme paresseux, il occupe plus de mémoire et de stockage de données
- K-NN est plus enclin à l'ajustement excessif, afin que le choix de K est crucial pour le modèle .

- L'apprentissage –algorithms

La forêt aléatoire

LA FORÊT ALÉATOIRE

- Objectifs :

La forêt aléatoire est une technique utilisée dans la modélisation des prédictions et l'analyse du comportement est construite sur des arbres de décision. Elle contient de nombreux arbres de décision qui représentent une instance distincte de la classification des données entrées dans la forêt aléatoire.

- Fonctionnalité :

La forêt aléatoire est un algorithme d'apprentissage supervisé. La forêt qu'il construit est un ensemble d'arbres de décision, généralement formés avec la méthode « d'ensachage ». L'idée générale de la méthode d'ensachage est qu'une combinaison de modèles d'apprentissage augmente le résultat global.

En termes simples : la forêt aléatoire crée plusieurs arbres de décision et les fusionne pour obtenir une prédiction plus précise et plus stable.

L'un des grands avantages de la forêt aléatoire est qu'elle peut être utilisée à la fois pour les problèmes de classification et de régression, qui constituent la majorité des systèmes d'apprentissage automatique actuels.

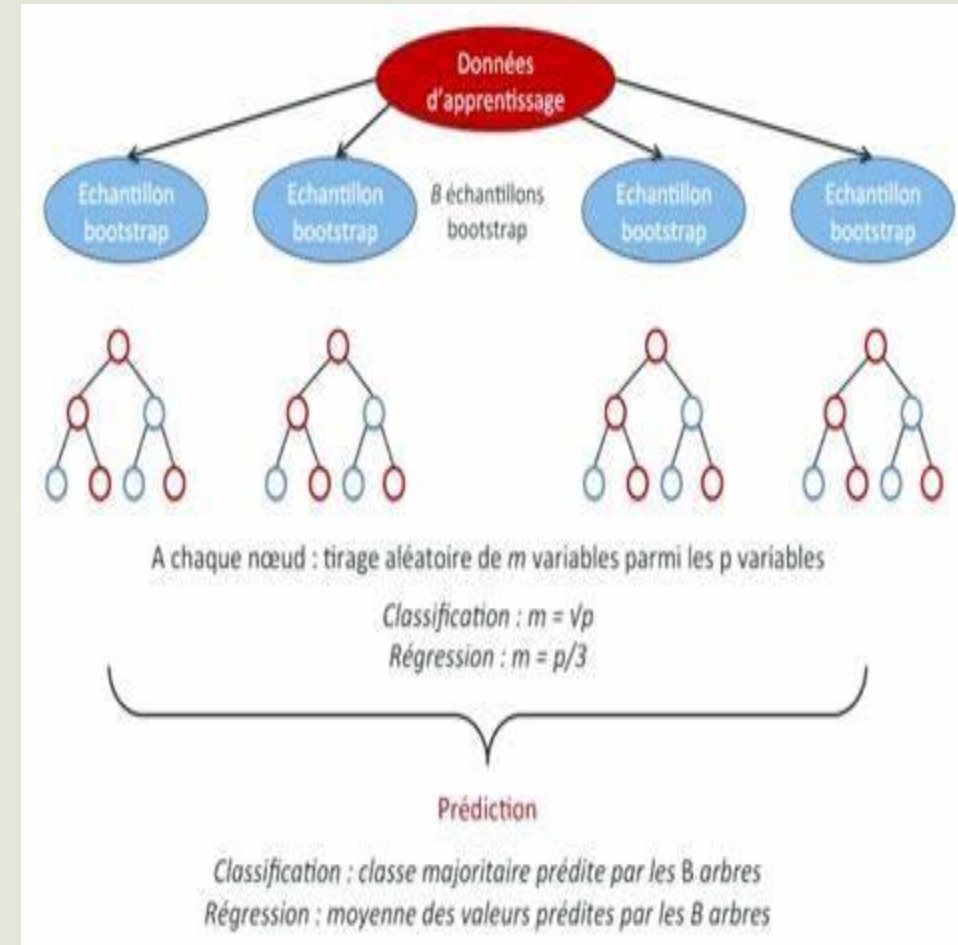
• L'apprentissage –algorithms

La forêt aléatoire

Le Bootstrap est une méthode d'estimation de la distribution d'échantillonnage fondée sur le prélèvement de plusieurs échantillons avec remise à partir d'un échantillon unique. On appelle ces échantillons répétés des rééchantillonnages. Chaque rééchantillonnage est de la même taille que l'échantillon d'origine.

- Différence entre les arbres de décision et les forêts aléatoires :

la différence principale entre l'arbre de décision et la forêt aléatoire est qu'un arbre de décision est un graphique qui utilise une méthode de branchement pour illustrer chaque résultat possible d'une décision, tandis qu'une forêt aléatoire est un ensemble d'arbres de décision qui donne le résultat final en fonction des résultats de tous ses arbres de décision



• L'apprentissage –algorithms

La classification naïve bayésienne

- Objectifs :

La classification naïve bayésienne est une technique d'apprentissage automatique qui peut être utilisée pour prédire à quelle catégorie appartient un cas de données spécifique

- Notions de base :

- La probabilité bayésienne :

$$P(A | B) = \frac{P(B|A) P(A)}{P(B)}$$

$$P(B|A) = P(B_1|A) P(B_2|A) \dots P(B_k |A)$$

$$Y = \text{Argmax}_{\pi_i} P(x_i | A) P(A)$$

- La méthode de la classification bayésienne :

- Déterminer la probabilité de chaque classe.
 - Déterminer la probabilité de chaque attribue du cas spécifique .
 - Applications de la probabilité de Bayes.
 - L'élément appartenu à la classe qui a une probabilité bayésienne la plus grande

Un exemple des variables discretes :

ID	X1:Age	X2:revenu	X3:Etudiant	X4:Credit Rating	Y:Achat d'un PC
1	Jeune	Elevé	Non	Bon	Non
2	Jeune	Elevé	Non	Mauvais	Non
3	Moyen	Elevé	Non	Bon	Oui
4	Agé	Moyen	Non	Bon	Oui
5	Agé	Faible	Oui	Bon	Oui
6	Agé	Faible	Oui	Mauvais	Non
7	Moyen	Faible	Oui	Mauvais	Oui
8	Jeune	Moyen	Non	Bon	Non
9	Jeune	Faible	Oui	Bon	Oui
10	Agé	Moyen	Oui	Bon	Oui
11	Jeune	Moyen	Oui	Mauvais	Oui
12	Moyen	Moyen	Non	Mauvais	Oui
13	Moyen	Elevé	Oui	Bon	Oui
14	Agé	Moyen	Non	Mauvais	Non
15	Moyen	Moyen	Oui	Bon	?

- L'apprentissage –algorithms

La classification naïve bayésienne

Le tableau contingence

Attribues	Modalité	Achat d'un ordinateur		Total
		oui	non	
Age	Jeune	2	3	5
	Moyen	4	0	4
	Agé	3	2	5
	Elevé	2	2	4
Revenu	Moyen	4	2	6
	Faible	3	1	4
Etudiant	OUI	6	1	7
	NON	3	4	7
	Bon	6	2	8
Crédit rating	Mauvais	3	3	6
Total		9	5	14

X=(Age=Moyen ,Revenu=Moyen,Etudiant=Oui,CR=Bon)

- L'apprentissage –algorithms

La classification naïve bayésienne

Application numérique :

$$P(C1): P(Y=Oui)=9:14=0,643$$

$$P(C2): P(Y=Non)=5:14=0,357$$

on calcule $P(X|Y=Ci)$:

$$P(\text{Age} = \text{Moyen} | y=Oui)=4/9=0,444$$

$$P(\text{Age} = \text{Moyen} | y=Non)=1/8=0,125$$

$$P(\text{revenu} = \text{Moyen} | y=Oui)=4/9=0,444$$

$$P(\text{Revenu}=\text{Moyen} | y=Non)=2/5=0,4$$

$$P(\text{Etudiant}=Oui | y=Oui)=6/9=0,667$$

$$P(\text{Etudiant}=Oui | y=Non)=1/5=0,2$$

$$P(CR=\text{Bon} | y=Oui)=6/9=0,667$$

$$P(CR=\text{Bon} | y=Non)=2/5=0,4$$

$$P(x/y=Oui)=0,444*0,444*0,667*0,667=0,087$$

$$P(x/y=Non)=0,125*0,4*0,2*0,4=0,004$$

$$P(Y=Oui | X=l15)= P(x/y=Oui) * P(Y=Oui)=0,056$$

$$P(Y=Non | X=l15)= P(x/y=Non) * P(Y=Non)=0,0014$$

- L'apprentissage –algorithms

La classification naïve bayésienne

Un exemple des variables continues

- On calcule l'espérance et la variance de chaque attribue.
- Déterminer la probabilité de chaque classe.
- Déterminer la probabilité de chaque attribue du cas spécifique en utilisant la loi Normale
- Applications de la probabilité de Bayes.
- L'élément appartenu à la classe qui a une probabilité bayésienne la plus grande

- Entraînement

Sexe	Taille (cm)	Poids (kg)	Pointure (cm)
masculin	182	81.6	30
masculin	180	86.2	28
masculin	170	77.1	30
masculin	180	74.8	25
féminin	152	45.4	15
féminin	168	68.0	20
féminin	165	59.0	18
féminin	175	68.0	23

- hypothèse de distribution Gaussienne pour les lois de probabilités des caractéristiques :

Sexe	Espérance (taille)	Variance (taille)	Espérance (poids)	Variance (poids)	Espérance (pointure)	Variance (pointure)
masculin	178	2.9333×10^1	79.92	2.5476×10^1	28.25	5.5833×10^0
féminin	165	9.2666×10^1	60.1	1.1404×10^2	19.00	1.1333×10^1

- L'apprentissage –algorithms

La classification naïve bayésienne

le cas spécifique :

- | Sexe | Taille (cm) | Poids (kg) | Pointure (cm) |
|---------|-------------|------------|---------------|
| inconnu | 183 | 59 | 20 |
- Quelle probabilité *a posteriori* est la plus grande ?
 $Pr[(183, 59, 20)|feminin]$ ou $Pr[(183, 59, 20)|masculin]$?

la loi Normale :

- On peut à présent déterminer le sexe de l'échantillon avec :

$$f_{j,k}(x) = \frac{1}{\sqrt{2\pi\sigma_{k,j}^2}} \exp\left(\frac{-1}{2\sigma_{k,j}^2}(x - \mu_{k,j})^2\right)$$

pour une variable j dans la classe k .

- L'apprentissage –algorithms

La classification naïve bayésienne

- Application numérique :

$$\begin{aligned}P(\text{taille}|M) &= f_{t,m}(x) \\&= \frac{1}{\sqrt{2\pi \times 2,9333 \times 10^1}} \exp \left(\frac{-1}{2 \times 2.9333 \times 10^1} (183 - 178)^2 \right)\end{aligned}$$

● On réalise ce calcul pour chacune des variables et des groupes :

$P(M)$	$=$	0.5	$P(F)$	$=$	0.5
$P(\text{taille} M)$	$=$	4.8102×10^{-2}	$P(\text{taille} F)$	$=$	7.2146×10^{-3}
$P(\text{poids} M)$	$=$	1.4646×10^{-5}	$P(\text{poids} F)$	$=$	3.7160×10^{-2}
$P(\text{pointure} M)$	$=$	3.8052×10^{-4}	$P(\text{pointure} F)$	$=$	1.1338×10^{-1}
$P_p(M)$	$=$	1.3404×10^{-10}	$P_p(F)$	$=$	1.5200×10^{-5}

• Évaluation des modèles de classification

Evaluation de la qualité

Évaluation des modèles de classification

Dans le domaine de la recherche, il est courant de devoir comparer des modèles de classification afin de sélectionner le meilleur modèle pour nos données. Nous devons être capables de dire si un modèle est plus performant qu'un autre tout en le justifiant. Pour cela, il existe différents critères d'évaluation

- Compréhensibilité du modèle
- Complexité spatiale et temporelle du modèle
- Qualité du modèle (en utilisant les métriques de classification)

Evaluation de la qualité

La matrice de confusion : appelée également matrice d'erreur, est un tableau qui présente différentes prévisions et résultats de tests, en les comparant avec des valeurs réelles

- Évaluation des modèles de classification

Evaluation de la qualité

		Classe réelle	
		-	+
Classe prédite	-	True Negatives <i>(vrais négatifs)</i>	False Negatives <i>(faux négatifs)</i>
	+	False Positives <i>(faux positifs)</i>	True Positives <i>(vrais positifs)</i>

TP (Vrais Positifs): c'est un résultat où le modèle prédit correctement la classe positive.

TN (Vrais Négatifs): c'est un résultat où le modèle prédit correctement la classe négative.

FP (Faux Positifs): c'est un résultat où le modèle prédit incorrectement la classe positive

FN (Faux Négatifs): c'est un résultat où le modèle prédit incorrectement la classe négative

• Évaluation des modèles de classification

Evaluation de la qualité

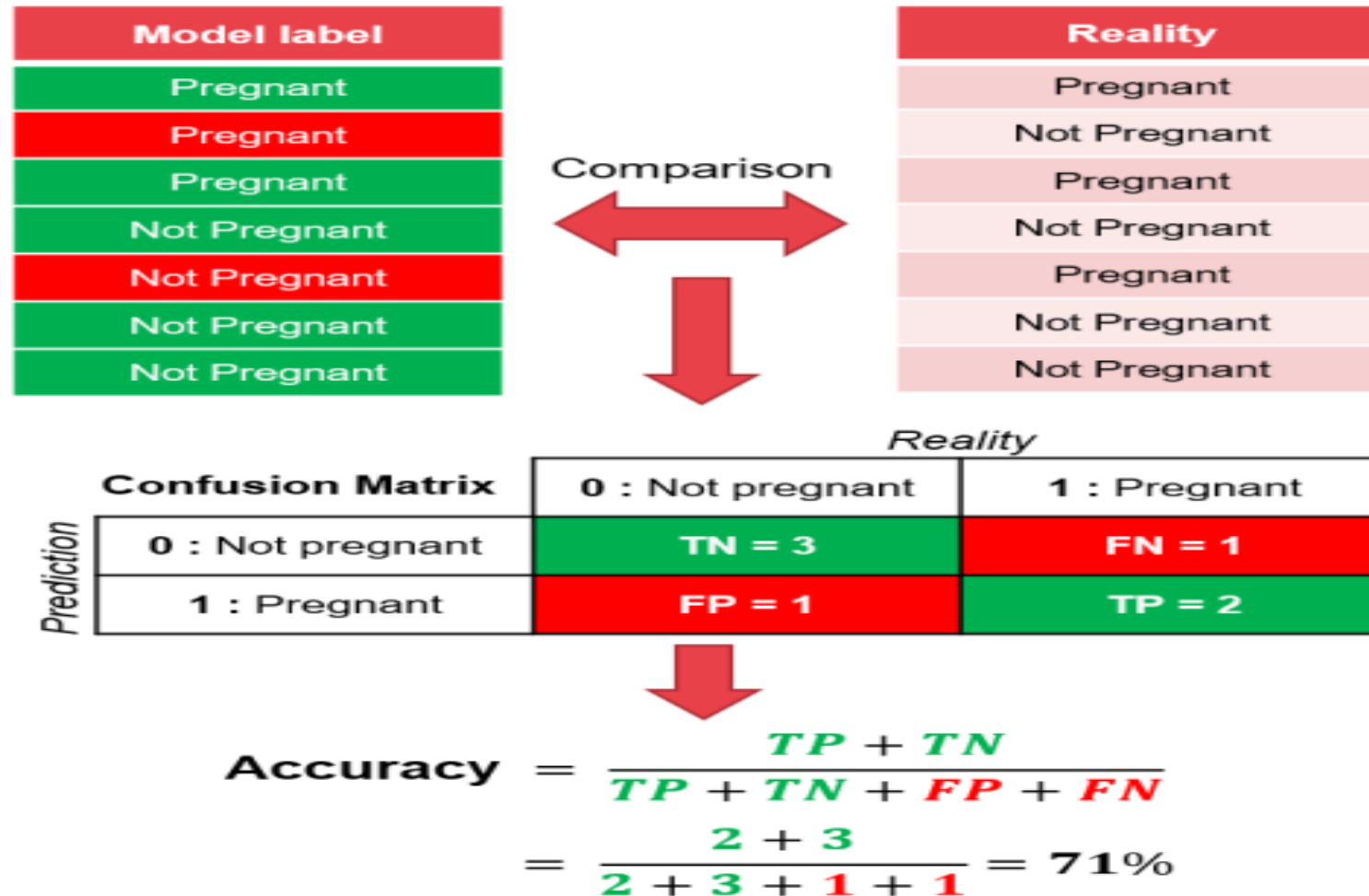
• Remarque :

- TP et TN nous indiquent quand le classificateur obtient les choses correctes
- FP et FN nous indiquent quand le classificateur se trompe

Les mesures	Les formules
La justesse, score (ou "accuracy" en anglais)	$\frac{TP + TN}{P + N}$
Précision	$\frac{TP}{TP + FP}$
rappel ("recall" en anglais), ou sensibilité ("sensitivity" en anglais)	$\frac{TP}{P}$
F-mesure ("F-score" en anglais)	$\frac{2 \times \text{Précision} \times \text{Rappel}}{\text{précision} + \text{rappel}}$
spécificité ("specificity" en anglais)	$\frac{TN}{N}$

- Évaluation des modèles de classification

Evaluation de la qualité



- **Partitionnement – Clustering**
dendrogramme

Objectif :

Effectuer un regroupement en k ($k \ll n$) groupes de manière à rassembler dans chaque groupe les individus “les plus semblables” selon un critère à définir (en général assimilé à une distance).

Méthode générale :

Classification des n individus en k classes telles que : l'homogénéité soit maximale à l'intérieur de chaque classe, l'hétérogénéité soit maximale d'une classe à l'autre.

Remarque :

Les classes et le nombre k de classes sont inconnus

- **La partie hiérarchique (plan)**
 1. Les 4 étapes de la méthode CAH
 2. L'algorithme
 3. Application de l'algorithme

- **Partitionnement – Clustering**
dendrogramme

- **Notions de bases de la méthode CAH**

1. Choix d'un indice de dissimilarité
2. Choix d'un indice d'agrégation
3. Quelle partie du dendrogramme faut-il conserver

Choix 1 : d'un indice de dissimilarité

- Distance Euclidienne. $d(I_i, I_j) = \sqrt{\sum_k (x_{ik} - x_{jk})^2}$
le type de distance le plus couramment utilisé
- Distance Euclidienne au carré. $d(I_i, I_j) = \sum_k (x_{ik} - x_{jk})^2$
- Distance du City-block (Manhattan) : $d(I_i, I_j) = \sum_k |x_{ik} - x_{jk}|$
- Distance de Tchebychev : $d(I_i, I_j) = \text{Max}_k |x_{ik} - x_{jk}|$
- Distance à la puissance. $d(I_i, I_j) = \left(\sum_k |x_{ik} - x_{jk}|^p \right)^{1/p}$
- Percent disagreement. $d(I_i, I_j) = \frac{\text{Nombre de } x_{ik} \neq x_{jk}}{K}$
utile si les données de nature catégorielle.
- 1 - r de Pearson : $d(I_i, I_j) = \sqrt{1 - r_{ij}}$

- Partitionnement – Clustering
dendrogramme

- Notions de bases de la méthode CAH

1. Choix d'un indice de dissimilarité
2. Choix d'un indice d'agrégation
3. Quelle partie du dendrogramme faut-il conserver

Choix 2 : d'un indice d'agrégation

Objectif: calculer la distance entre deux classes quelconques
sans avoir à recalculer celles qui existent entre les individus

- Saut minimum ou "single linkage" (distance minimum).

$$D(A,B) = \min_{I \in A} \min_{J \in B} d(I,J)$$

- Diamètre ou "complete linkage" (distance maximum).

$$D(A,B) = \max_{I \in A} \max_{J \in B} d(I,J)$$

- Moyenne non pondérée des groupes associés.

$$D(A,B) = \frac{1}{n_A n_B} \sum_{I \in A, J \in B} d(I,J)$$

- Moyenne pondérée des groupes associés.

$$D(A,B) = \frac{1}{(n_A + n_B)(n_A + n_B - 1)} \sum_{I, J \in A \cup B} d(I,J)$$

- Partitionnement – Clustering
dendrogramme

On peut aussi utiliser la méthode de Ward comme méthode d'agrégation

Algorithme d'agrégation par critère de Ward

- ▶ **Principe général** : Au départ de l'algorithme, l'inertie inter classes est maximale (n classes). A la fin, celle-ci est nulle (1 classe).
⇒ On cherche à minimiser à chaque étape la perte d'inertie inter-classes (ou à minimiser le gain d'inertie intra-classe).
- ▶ **A chaque étape** : On regroupe les 2 classes pour lesquelles la perte d'inertie inter-classes est minimale. Cela revient à regrouper les classes j et j' pour lesquelles la perte

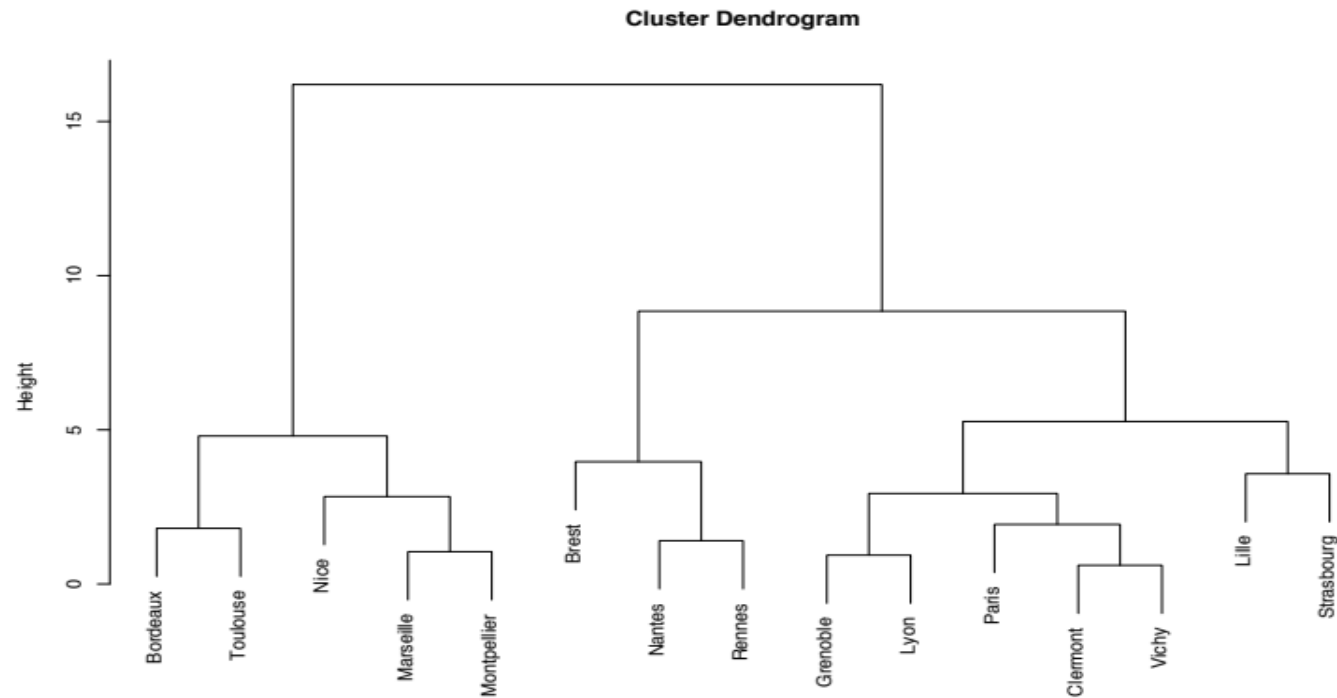
$$\Delta_{jj'} = \frac{n_j n_{j'}}{n_j + n_{j'}} \|C_j - C_{j'}\|^2$$

est minimale.

- Partitionnement – Clustering
dendrogramme

3-ème étape : Quelle partie du dendrogramme faut-il conserver

Dendrogramme obtenu par la CAH



- Partitionnement – Clustering
dendrogramme

Choix du nombre de classes

Choix du nombre de classes

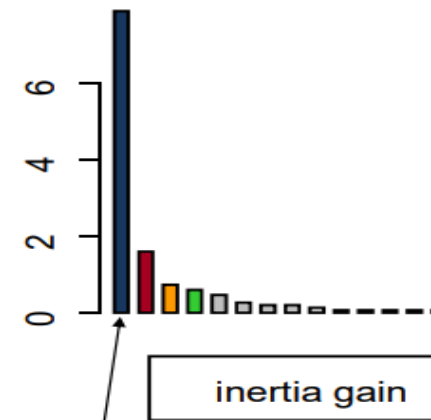
- ▶ *Coupure* de l'arbre à un niveau donné de l'indice \implies *partition*.
- ▶ La coupure doit se faire :
 - ▶ **après** les agrégations correspondant à des valeurs **peu élevées** de l'indice,
 - ▶ **avant** les agrégations correspondant à des niveaux **élevés** de l'indice, qui dissocient les groupes bien distincts dans la population.
- ▶ *Règle empirique* : sélection d'une coupure lors d'un saut important de l'indice par *inspection visuelle* de l'arbre.
- ▶ Ce saut traduit le passage brutal entre des classes d'une certaine homogénéité de l'ensemble à des classes beaucoup moins homogènes.
- ▶ Dans la plupart des cas, il y a *plusieurs paliers* et donc *plusieurs choix de partitions* possibles.

- Partitionnement – Clustering
dendrogramme

Les données température

Pertes d'inertie inter lors du passage de

15 classes en 14 classes : 0.01
14 classes en 13 classes : 0.02
13 classes en 12 classes : 0.03
12 classes en 11 classes : 0.05
11 classes en 10 classes : 0.06
10 classes en 9 classes : 0.09
9 classes en 8 classes : 0.17
8 classes en 7 classes : 0.19
7 classes en 6 classes : 0.26
6 classes en 5 classes : 0.42
5 classes en 4 classes : 0.56
4 classes en 3 classes : 0.69
3 classes en 2 classes : 1.56
2 classes en 1 classe : 7.88



Grosse perte si on passe de 2 classes à 1 seule donc on préfère garder 2 classes

Somme des pertes d'inertie = 12

- Partitionnement – Clustering
dendrogramme

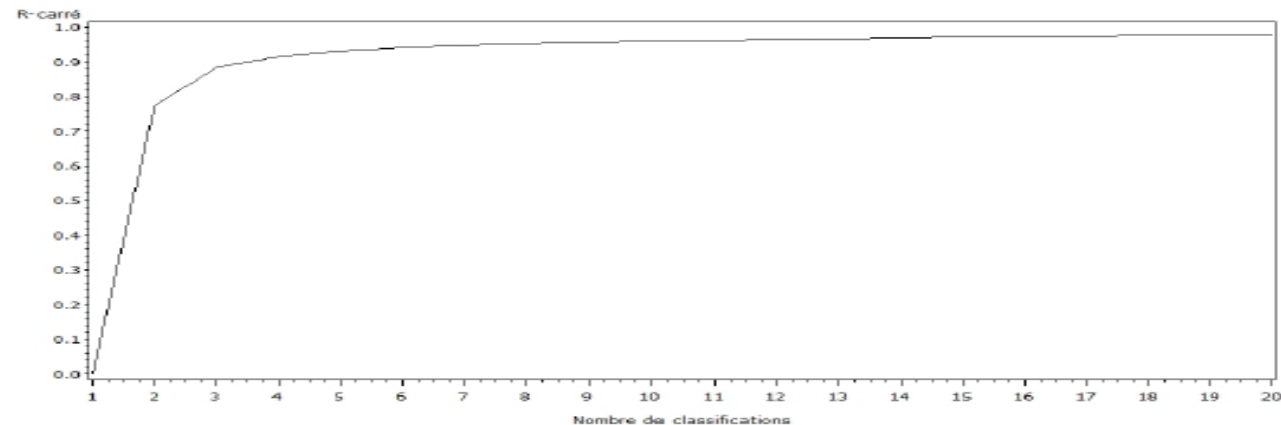
Choix du nombre de classes : R^2

Souvent, plusieurs choix de partitions possibles

⇒ utilisation d'un critère numérique :

$$R^2(l_1, \dots, l_k) = \frac{\mathcal{I}_{inter}(l_1, \dots, l_k)}{\mathcal{I}_G}$$

Repérage du point k où il y a rupture de pente dans le R^2 :



- Partitionnement – Clustering
dendrogramme

L'Algorithme pour dresser le dendrogramme :

Etape 0 : dresser la matrice de distance

Étape 1 : sélectionner la distance minimale entre les individus i et j

Etape2 : actualiser la matrice de distance

Etape 3 : si la matrice de distance comporte un seul élément on s'arrête sinon on revient à l'étape N 1

Application :

Classification Ascendante Hiérarchique: CAH

On suppose de réaliser une
classification des 7 points: M1,,,M7

	X1	X2
1 (M1)	0	4
2 (M2)	1	1
3 (M3)	1	2
4 (M4)	1	5
5 (M5)	3	4
6 (M6)	4	3
7 (M7)	6	2

- Partitionnement – Clustering
k-means

Calcul des distances euclidiennes

	M1	M2	M3	M4	M5	M6	M7
M1	0	3,1	2,2	1,4	3	4,1	6,3
M2		0	1	4	3,6	3,6	5,1
M3			0	3	2,8	3,1	5
M4				0	2,2	3,6	5,8
M5					0	1,4	3,6
M6						0	2,2
M7							0

$$\text{dist}(M1, M1) = 0$$

$$M1(0,4); M2(1,1)$$

$$\text{dist}(M1, M2) = ((0-1)^2 + (4-1)^2)^{1/2}$$

$$\text{dist}(M1, M2) = 3,1$$

ainsi de suite ...

La plus courte distance est 1 entre M2 et M3, on crée donc la classe:

C1={M2, M3}

Remplir la matrice suivant

	C1	M1	M4	M5	M6	M7
C1	0	2,2	3	2,8	3,1	5
M1		0	1,4	3	4,1	6,2
M4			0	2,2	3,6	5,8
M5				0	1,4	3,6
M6					0	2,2
M7						0

Calculons d'abord les distances au plus proche voisin de la classe C1 avec les 5 points restants

Par exemple:

$$\text{Dist}(C1, M1) = \min_{i=2,3} \text{dist}(Mi, M1)$$

Dist(M2, M1)=3,1 et
dist(M3, M1)=2,2

Donc **dist(C1, M1)=2,2** puisque c'est le min ... Ainsi de suite

La nouvelle classe C2={M1, M4}

- Partitionnement – Clustering
k-means

Même démarche

	C1	C2	M5	M6	M7
C1	0	2,2	2,8	3,1	5
C2		0	2,2	3,6	5,8
M5			0	1,41	3,6
M6				0	2,2
M7					0

La nouvelle classe •
C3={M5,M6} •

CAH

	C1	C2	C3	M7
C1	0	2,2	2,8	5
C2		0	2,2	5,8
C3			0	2,2
M7				0

La nouvelle classe •
C4={C1,C2} •

Fin de l'algorithme : CAH

	C5	M7
C5	0	2,2
M7	2,2	0

Le dernier classe est: •
C6={C5,M7} •

- # Partitionnement – Clustering

partitionnement non hiérarchique 'K-Means '

PARTIE 2 : partitionnement non hiérarchique 'K-Means '

Objectif :

Le clustering K-Means est une méthode de quantification vectorielle, utilisé pour partitionner des données observées en k clusters ; Vous donnez au modèle un ensemble de données avec des caractéristiques définies et vous lui dites combien de clusters vous voulez qu'il produise. Le modèle classera le jeu de données dans le nombre de clusters que vous lui avez attribué.

Comment fonctionne le K-Means ?

Si k est donné, l'algorithme K-Means peut être exécuté dans les étapes suivantes:

Etape 1 : Partition d'objets en k sous-ensembles non vides

Etape 2 : Identification des centroïdes de cluster (point moyen) de la partition courante.

Etape 3 : Affectation de chaque point à un cluster spécifique

Etape 4 : Calculez les distances de chaque point et attribuez des points au cluster où la distance du centroïde est minimale.

Etape 5 : Après avoir réattribué les points, trouvez le centroïde du nouveau groupe formé.

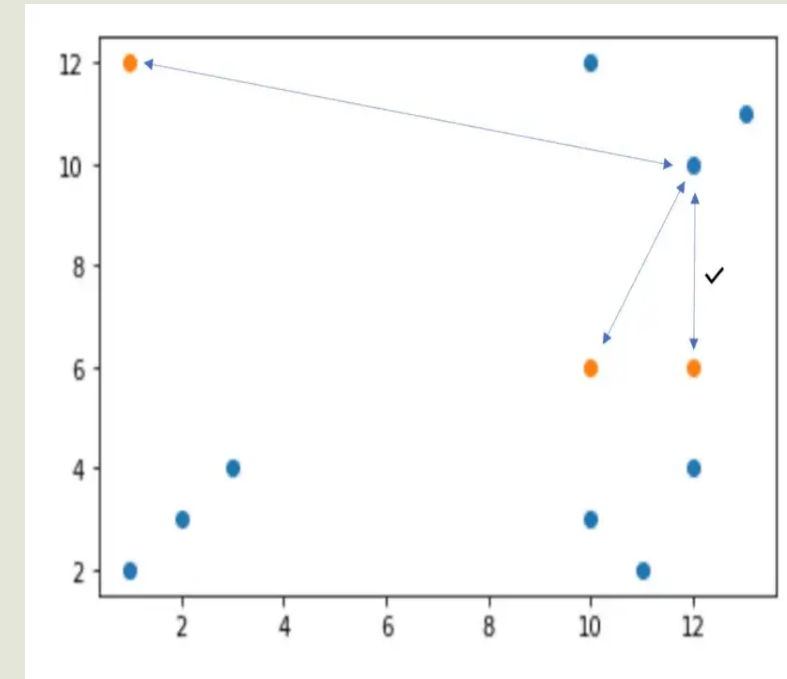
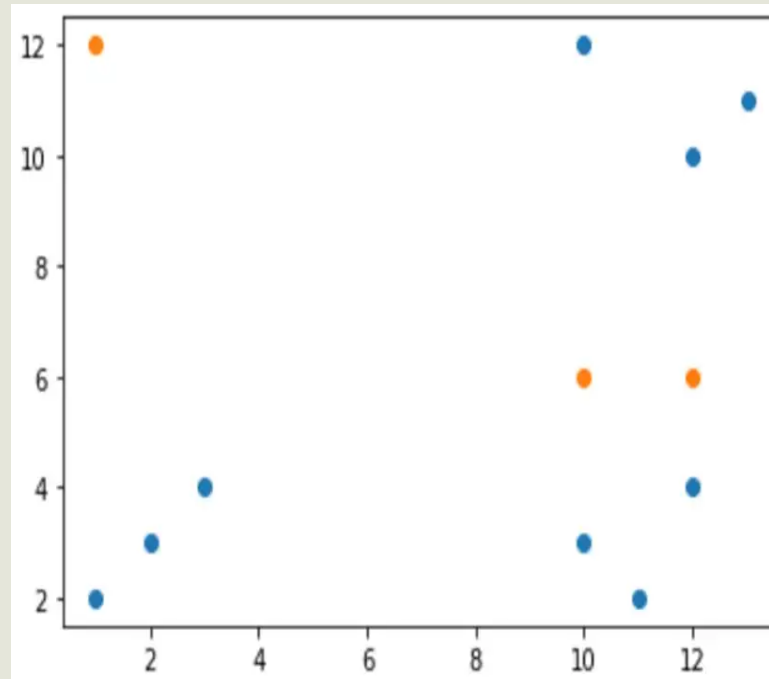
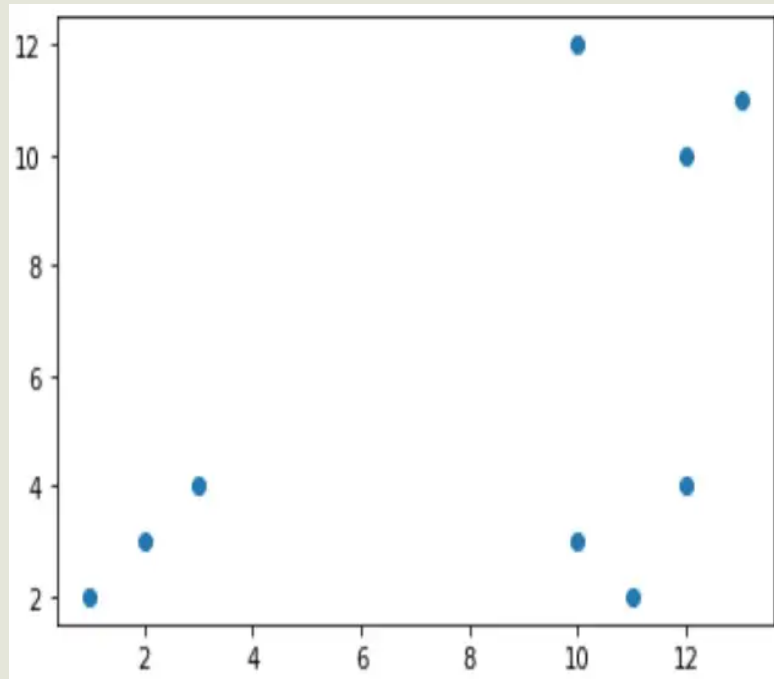
• Partitionnement – Clustering

partitionnement non hiérarchique 'K-Means '

Suivant le processus en utilisant un cas réel, étape par étape. Supposons que j'ai 9 points de données :

points = [[2,3],[3,4],[1,2],[10,12],[12,10],[13,11],[10,3],[11,2],[12,4]]

Disons que je veux classer les 9 points dans k= 3 clusters.



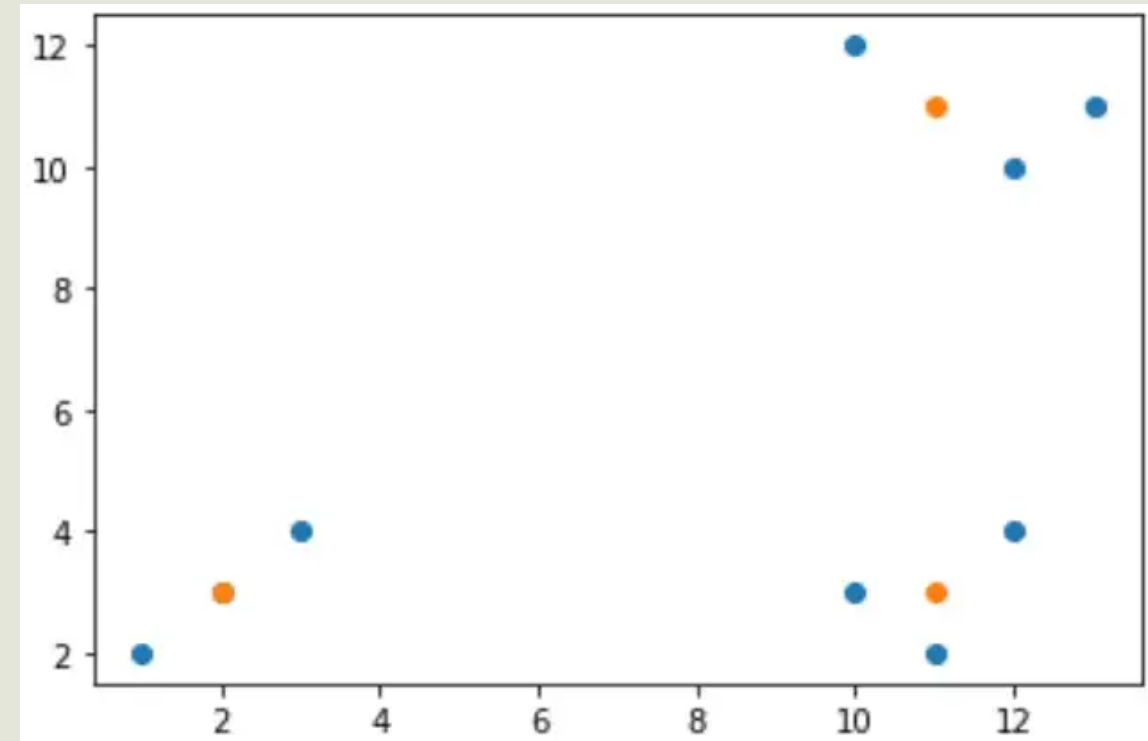
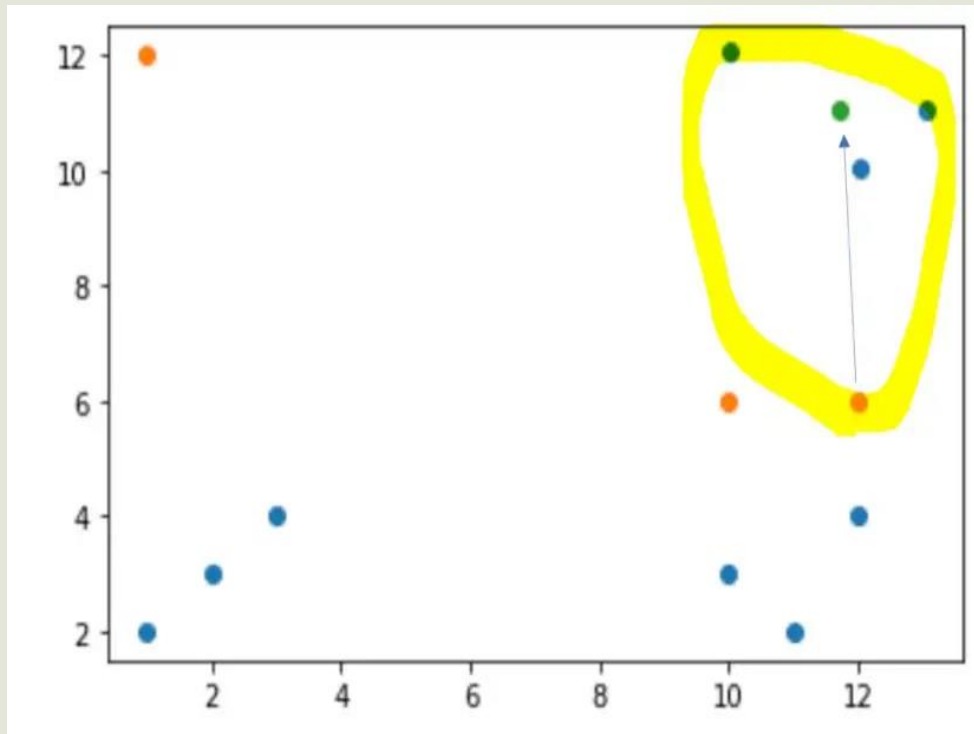
• Partitionnement – Clustering

partitionnement non hiérarchique 'K-Means '

Suivant le processus en utilisant un cas réel, étape par étape. Supposons que j'ai 9 points de données :

```
points = [[2,3],[3,4],[1,2],[10,12],[12,10],[13,11],[10,3],[11,2],[12,4]]
```

Disons que je veux classer les 9 points dans $k=3$ clusters.



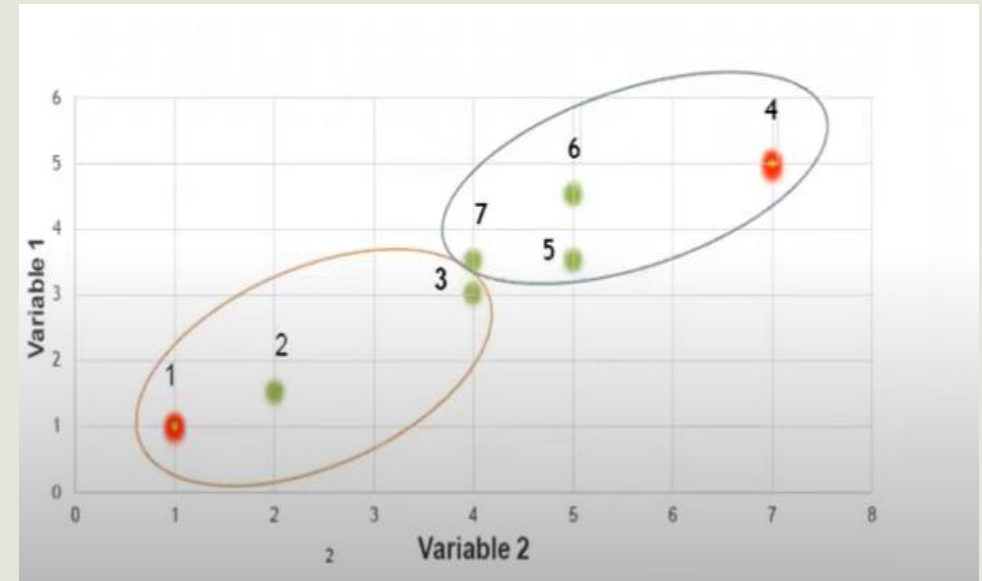
- Partitionnement – Clustering

partitionnement non hiérarchique 'K-Means '

Exercice d'application :

A Simple example k-means (using K=2)

Individual	Variable 1	Variable 2
1	1	1
2	1.5	2
3	3	4
4	5	7
5	3.5	5
6	4.5	5
7	3.5	4.5



• Partitionnement – Clustering

partitionnement non hiérarchique 'K-Means '

Etape 1 : initialisation ,on choisie au hasard 2 centroïdes pour 2 clusters . Dans notre cas ,m1=(1,1) et m2= (5,7)

.	INDIVIDU	VECTEUR MOYEN
G1	1	(1,1)
G2	4	(5,7)

Etape 2:

CENTROIDE 1	CENTROIDE 2
$\sqrt{(1-1)^2 + (1-1)^2} = 0$	$\sqrt{(5-1)^2 + (7-1)^2} = 7.21$
$\sqrt{(1-1.5)^2 + (1-2)^2} = 1.12$	$\sqrt{(5-1.5)^2 + (7-2)^2} = 6.10$
$\sqrt{(1-3)^2 + (1-4)^2} = 3.61$	$\sqrt{(5-3)^2 + (7-4)^2} = 3.61$
$\sqrt{(1-5)^2 + (1-7)^2} = 7.21$	$\sqrt{(5-5)^2 + (7-7)^2} = 0.00$
$\sqrt{(1-3.5)^2 + (1-5)^2} = 4.72$	$\sqrt{(5-3.5)^2 + (7-5)^2} = 2.50$
$\sqrt{(1-4.5)^2 + (1-5)^2} = 5.31$	$\sqrt{(1-4.5)^2 + (1-5)^2} = 2.06$
$\sqrt{(1-3.5)^2 + (1-4.5)^2} = 4.30$	$\sqrt{(1-3.5)^2 + (1-4.5)^2} = 2.92$

- Partitionnement – Clustering

partitionnement non hiérarchique 'K-Means '

Les nouveau centroïdes : $G1 = \left(\frac{1+1.5+3}{3} ; \frac{1+2+4}{3} \right) = (1.83 ; 2,33)$

$$G2 = \left(\frac{(5+3.5+4.5+3.5)}{4} ; \frac{7+5+5+4.5}{4} \right) = (4.12 ; 5,38)$$

On refait mm démarche pour G1 et G2, on obtient :

$$G1 = \left(\frac{1+1.5}{2} ; \frac{1+2}{2} \right) = (1.25 ; 1,5)$$

$$G2 = \left(\frac{(3+5+3.5+4.5+3.5)}{5} ; \frac{4+7+5+5+4.5}{5} \right) = (3.9 ; 5,1)$$

- Partitionnement – Clustering

partitionnement non hiérarchique 'K-Means '

	Centroid 1	Centroid 2
1	$J(1.83 - 1)^2 + (2.33 - 1)^2 = 1.57$	$J(4.12 - 1)^2 + (5.38 - 1)^2 = 5.38$
2	$J(1.83 - 1.5)^2 + (2.33 - 2)^2 = 0.47$	$J(4.12 - 1.5)^2 + (5.38 - 2)^2 = 4.29$
3	$J(1.83 - 3)^2 + (2.33 - 4)^2 = 2.04$	$J(4.12 - 3)^2 + (5.38 - 4)^2 = 1.78$
4	$J(1.83 - 5)^2 + (2.33 - 7)^2 = 5.64$	$J(4.12 - 5)^2 + (5.38 - 7)^2 = 1.84$
5	$J(1.83 - 3.5)^2 + (2.33 - 5)^2 = 3.15$	$J(4.12 - 3.5)^2 + (5.38 - 5)^2 = 0.73$
6	$J(1.83 - 4.5)^2 + (2.33 - 5)^2 = 3.78$	$J(4.12 - 4.5)^2 + (5.38 - 5)^2 = 0.54$
7	$J(1.83 - 3.5)^2 + (2.33 - 4.5)^2 = 2.74$	$J(4.12 - 3.5)^2 + (5.38 - 4.5)^2 = 1.08$

On remarque qu' il n y avait pas de variations au niveau de cluster

Par suite , l algorithme est arrivée au bout . donc le résultats finals comporte 2 clusters {1 ;2} et {3,4,5,6,7}

- Partitionnement – Clustering

partitionnement non hiérarchique 'K-Means '

Conclusion :

On peut retenir que le clustering K-Means est une technique très utilisée pour l'analyse des groupes de données. elle fournit rapidement des résultats de formation. Cependant, ses performances ne sont généralement pas aussi compétitives que celles des autres techniques de clustering sophistiquées, car de légères variations dans les données peuvent entraîner une variance élevée.

- SYNTHÈSE

La méthodologie de la classification et du clustering est différente, et les résultats attendus de leurs algorithmes le sont également. En bref, la classification et le clustering sont tous deux utilisés pour résoudre des problèmes différents.

Le clustering et la classification sont deux méthodes d'apprentissage automatique efficaces qui contribuent à l'amélioration des processus métier. Bien que ces processus soient identiques, vous pouvez les utiliser de manière unique pour comprendre les objectifs de vos clients et améliorer leur expérience d'achat. En examinant, en profilant et en ciblant vos consommateurs à l'aide de la classification et du clustering dans l'apprentissage automatique, vous préparerez une base de clients fiables et aurez un meilleur retour sur investissement.

The background features a dark, textured collage of white line-art icons representing various educational fields: a globe, a microscope, a book, a percentage sign, a ruler, and a compass.

**MERCI DE VOTRE
Attention**