

Mastère Professionnel : Business Intelligence

Apprentissage automatique



Arbres de décision

Apprentissage supervisé



Hind Elouedi

Semestre 2 - 2021



- Définir les concepts de base relatifs aux arbres de décision.
- Comprendre l'algorithme de construction d'un arbre de décision.
- Utiliser les arbres de décision dans la classification.
- Comprendre la notion d'élagage.
- Définir les notions de Bagging et Boosting.

Points abordés



- Introduction
- Composants
- Construction d'un arbre
- Classification
- Elagage
- Attributs à valeurs continues
- Bagging et boosting
- Conclusion



Un peu d'histoire...

- **Statistiques:**

- Morgan et Sonquist (1963): Arbres de régression dans un processus de prédiction et d'explication (AID – Automatic Interaction Detection).
- Morgan et Messenger (1973): THAID (Theta AID), utilisation des arbres pour les classements et les discriminations.
- Kass (1980): CHAID (CHi-squared Automatic Interaction Detector), prédiction ou détection d'interactions entre variables.
- Breiman et al. (1984): CART (Classification And Regression Tree).

Introduction (2)



Un peu d'histoire...

- **Apprentissage automatique:**
 - Quinlan (1979): ID3 (Induction of Decision Tree), les travaux de Quinlan sont rattachés à ceux de Hunt (1962).
 - Quinlan (1993): Ensemble de travaux couronnés par beaucoup d'articles et la mise en place de la méthode C4.5, la référence incontournable sur les arbres de décision.

Introduction (3)



- **Arbre de décision:**
 - Technique de classification en apprentissage supervisé.
 - Technique utilisée en intelligence artificielle.
- **Avantages:**
 - Traitement des problèmes complexes.
 - Expression simple de la connaissance.
 - Facilité dans la compréhension et l'interprétation des résultats.
 - Participation des experts dans l'élaboration des règles.

Introduction (4)



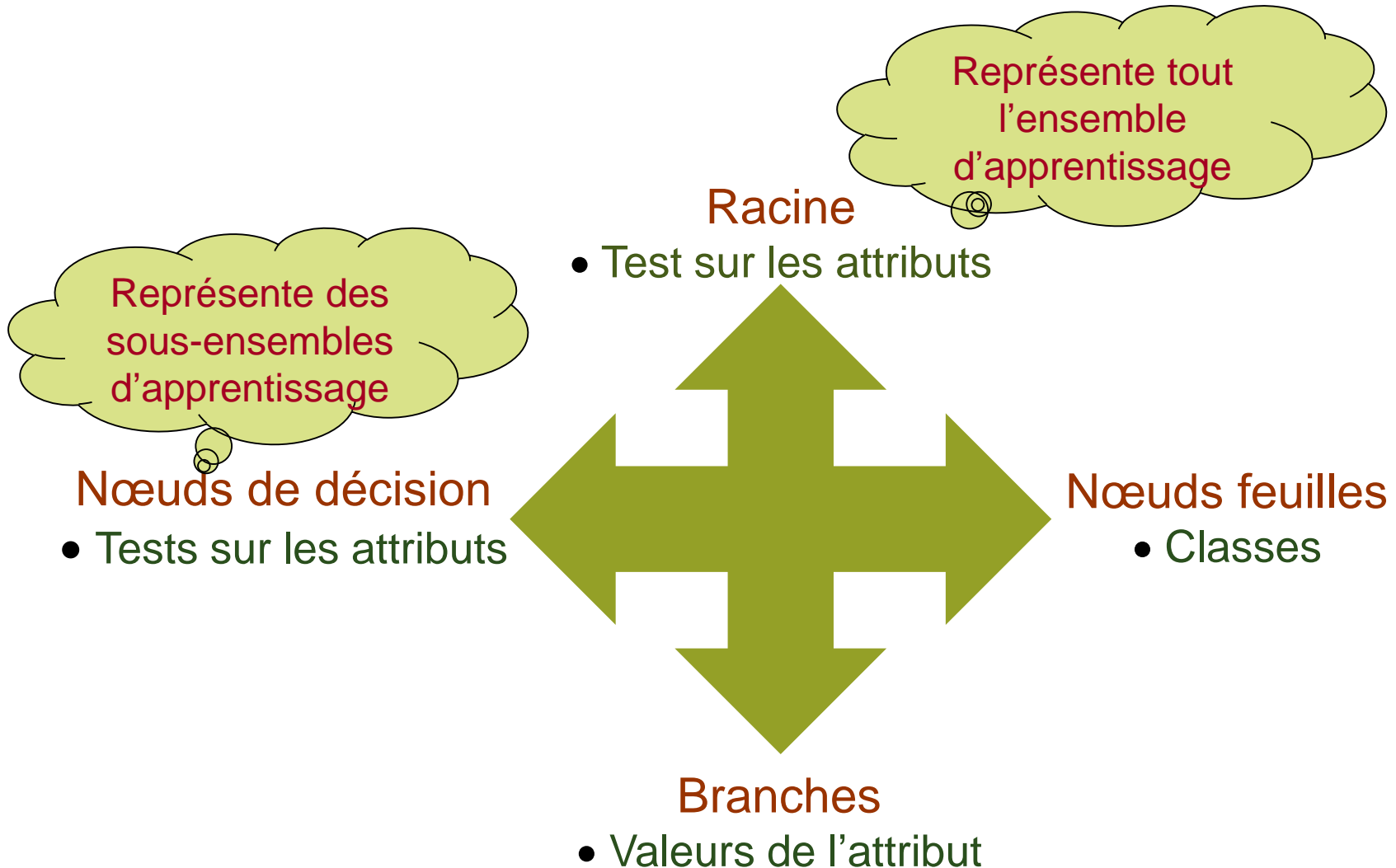
- Domaines d'application:
 - Gestion de crédits.
 - Diagnostic médical.
 - Analyse du marché.
 - Détection d'intrusion.
 - Contrôle de production.

■

■

■

Composants



Composants (2)



- Ensemble d'apprentissage

Attributs

Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C_1
Elevé	Supérieur	Oui	C_2
Elevé	Supérieur	Non	C_1
Elevé	Inférieur	Oui	C_2
Moyen	Supérieur	Non	C_1
Moyen	Supérieur	Oui	C_2
Moyen	Inférieur	Non	C_2
Moyen	Inférieur	Oui	C_2
Faible	Inférieur	Non	C_3
Faible	Inférieur	Oui	C_3

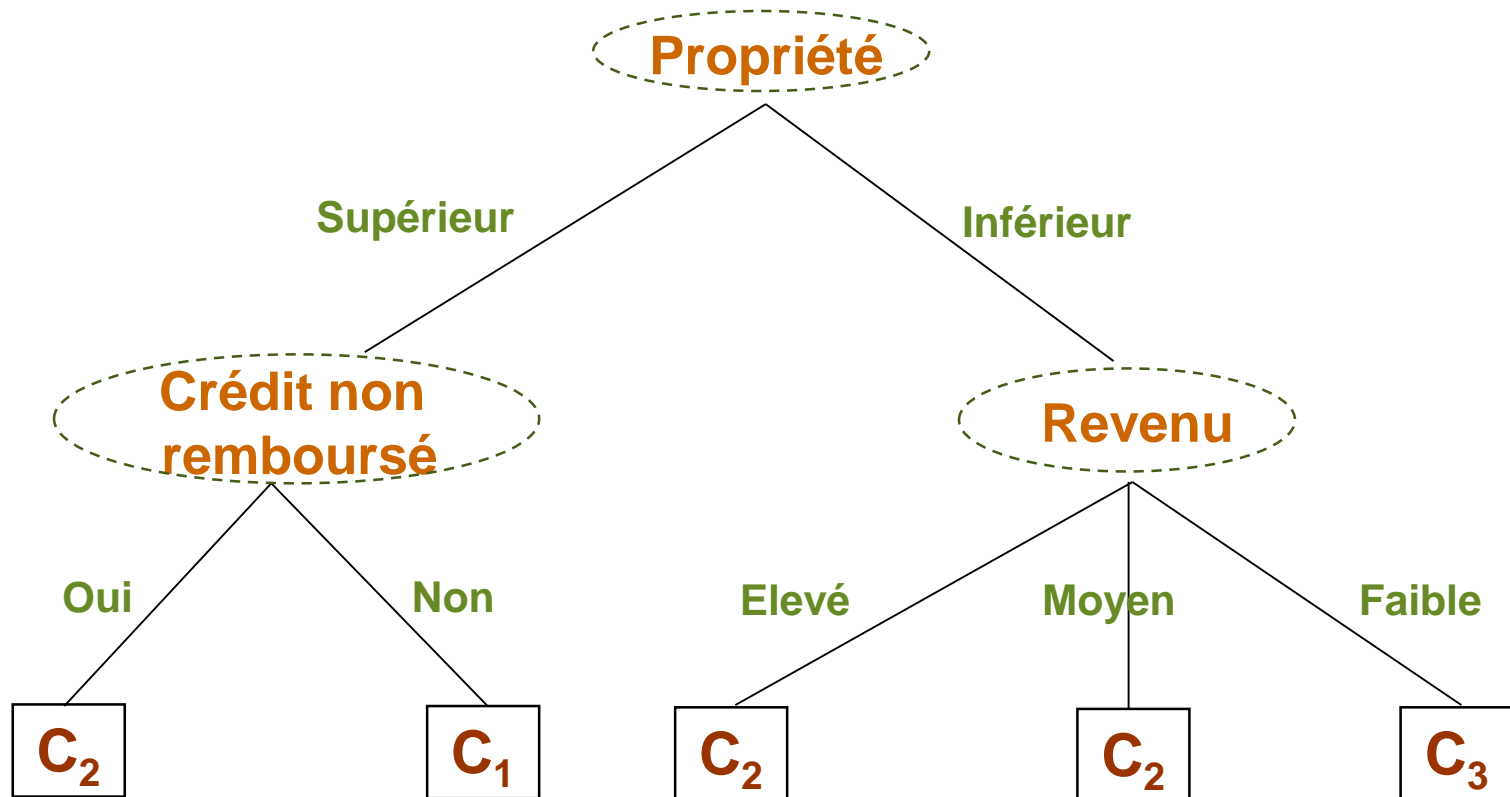
Valeurs des attributs

C_1 : Attribuer tout le crédit - C_2 : Attribuer une partie crédit - C_3 : Ne pas attribuer le crédit.

Composants (3)



Arbre de décision





- Problématique
- Procédure de construction
- Paramètres
- Application

Construction (2)



- **Problématique**

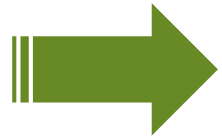
- Construire un arbre de décision à partir d'un ensemble d'apprentissage.



Généralisation: Être capable de classer correctement un nouvel objet.



Construction (3)



Processus très coûteux



Proposer un meilleur algorithme

- Choisir le **meilleur** attribut.
- **Partitionner** l'ensemble d'apprentissage.
- **Répéter** jusqu'à ce que chaque élément de l'ensemble d'apprentissage soit correctement classé.

Mais comment faire?

Construction (4)



- **Algorithmes:** Top Down Induction of Decision Trees (TDIDT)
 - ➡ Diviser pour régner (Induction descendante)
 - ID3 (Quinlan, 1979)
 - CART (Breiman et al., 1984)
 - ASSISTANT (Bratko, 1984)
 - C4.5 (Quinlan, 1993)

▪
▪
▪



Procédure de construction

- Processus récursif:
 - L'arbre commence à un nœud qu'on appelle **racine** et qui représente toutes les données.
 - Si les objets sont de la même classe, alors le nœud devient une **feuille** libellée par le nom de la classe.
 - Sinon, **sélectionner** les nœuds qui séparent **le mieux** les objets en classes **homogènes**.
 - Le traitement récursif s'arrête quand au moins l'un des critères d'arrêt est vérifié.

Construction (6)



Procédure de construction (2)

- Recherche à chaque niveau de l'attribut le plus **discriminant**.
- Partition des données (T):
 - Si tous les éléments de T sont dans la même classe alors retour;
 - Pour chaque attribut A, évaluer **la qualité du partitionnement** sur A;
 - Utiliser **le meilleur partitionnement** pour **diviser** T en T1, T2, ...Tk;
 - Pour i = 1 à k faire **Partition(T_i)**;

Construction (7)



Paramètres

- Mesure de sélection des attributs.
- Stratégie de partitionnement.
- Critères d'arrêt.

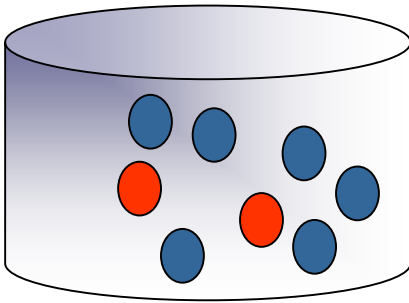


- **Choix de l'attribut:** Plusieurs mesures:
 - Gain d'information.
 - Indice de Gini.
 - Ratio de gain.
- **Mesure de l'information:**
 - L'**entropie** de **Shannon** exprime la quantité d'information: Le nombre de bits nécessaires pour coder l'information.

Paramètres (3)



- **Exemple**



La probabilité de tirer une boule bleue est:

$$\frac{6}{6+2} = \frac{3}{4}$$

La probabilité de tirer une boule rouge est:

$$\frac{2}{6+2} = \frac{1}{4}$$

- **Apport d'information**

- Nombre de bits nécessaires pour distinguer chaque boule parmi N:
 - P bits permettent de coder 2^P informations.
 - $\log_2(N)$ bits permettent de coder N informations.

Paramètres (4)



- Si je tire une boule (parmi N boules) et que je ne connais que sa couleur (par exemple elle est rouge), l'information acquise sera:

$$\log_2(N) \text{ bits} - \log_2(Nr) \text{ bits}$$



- Si je tire une boule au hasard et qu'on me donne sa couleur, l'information acquise sera:

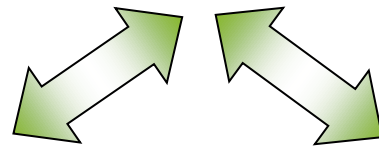
$$\text{Prob(Bleue)} (\log_2(N) - \log_2(Nb)) + \text{Prob(Rouge)} (\log_2(N) - \log_2(Nr))$$

$$\frac{3}{4} (\log_2 8 - \log_2 6) + \frac{1}{4} (\log_2 8 - \log_2 2)$$

Paramètres (5)



$$\text{Prob(Bleue)} (\log_2(N) - \log_2(N_b)) + \text{Prob(Rouge)}(\log_2(N) - \log_2(N_r))$$



$$\frac{N_b}{N} (\log_2 \frac{N}{N_b}) + \frac{N_r}{N} (\log_2 \frac{N}{N_r}) \longleftrightarrow - \frac{N_b}{N} (\log_2 \frac{N_b}{N}) - \frac{N_r}{N} (\log_2 \frac{N_r}{N})$$

$$\longleftrightarrow - \text{Prob(Bleue)} \log_2(\text{Prob(Bleue)}) - \text{Prob(Rouge)} \log_2(\text{Prob(Rouge)})$$

$$- \frac{3}{4} (\log_2 \frac{3}{4}) - \frac{1}{4} (\log_2 \frac{1}{4})$$

C'est la quantité d'information apportée par la couleur.



● Mesure de l'information

- Si on a n classes (C_1, C_2, \dots, C_n) de probabilités respectives p_1, p_2, \dots, p_n , la quantité d'information relative à la connaissance de la classe est définie par **l'entropie d'information:**

$$I = \sum_{i=1..n} -p_i \log_2 p_i$$

- $I = 0$ quand $\exists i / p_i = 1$ (une seule classe).
- I est maximale quand $\forall i / p_i = 1/n$ (classes équiprobables).



● Gain d'information (ID3)

- $\text{freq}(T, C_j)$: Nombre d'objets de T appartenant à la classe C_j .
- L'information relative à T est définie:

Quantité moyenne
d'information nécessaire
pour identifier la classe
d'un objet de T

$$\text{Info}(T) = - \sum_{j=1}^n \frac{\text{freq}(T, C_j)}{|T|} \log_2 \frac{\text{freq}(T, C_j)}{|T|}$$

- Une mesure similaire de T après partition selon l'attribut A (contenant n valeurs) est:

$$\text{Info}_A(T) = \sum_{i \in D_A} \frac{|T_i|}{|T|} \text{Info}(T_i)$$

D_A = Domaine de valeurs de l'attribut A .

- Le gain d'information mesure le gain obtenu suite au partitionnement selon l'attribut A : $\text{Gain}(T, A) = \text{Info}(T) - \text{Info}_A(T)$



On sélectionne l'attribut offrant le plus de gain.



- **Attributs multivalués**

- Le Critère de gain d'information présente une limite: **il favorise les attributs ayant plusieurs valeurs.**
- Lorsqu'un attribut a plusieurs valeurs possibles, son gain peut être très élevé, car il classifie parfaitement les objets.
- Par contre, cela peut générer un arbre de décision d'une profondeur de 1 (ou faible) qui ne sera pas très bon pour les instances futures.



- **Ratio de gain (C4.5)**

- Une mesure de l'information contenue dans l'attribut A (mesure de dispersion) est définie:

$$\text{Split Info}(T, A) = - \sum_{i \in D_A} \frac{|T_i|}{|T|} \log_2 \frac{|T_i|}{|T|}$$

- Le ratio de gain mesure le gain calibré par Split Info.

$$\text{Gain Ratio}(T, A) = \frac{\text{Gain}(T, A)}{\text{Split Info}(T, A)}$$

Quantité d'information générée par T et utile pour la classification

➡ On sélectionne l'attribut offrant le ratio de gain le plus élevé.



- **Stratégie de partitionnement**

- Pour chaque valeur de l'attribut, on va associer une branche dans l'arbre.
- Problème avec les attributs continus.



Découper en sous-ensembles ordonnés



- **Critères d'arrêt**
 - Si tous les objets appartiennent à la même classe.
 - S'il n'y a plus d'attributs à tester.
 - S'il n'y a pas d'objets avec la valeur d'attribut.
 - Absence d'apport informationnel des attributs.
(tous les ratios de gain ≤ 0)

Construction (8): Application



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C ₁
Elevé	Supérieur	Oui	C ₂
Elevé	Supérieur	Non	C ₁
Elevé	Inférieur	Oui	C ₂
Moyen	Supérieur	Non	C ₁
Moyen	Supérieur	Oui	C ₂
Moyen	Inférieur	Non	C ₂
Moyen	Inférieur	Oui	C ₂
Faible	Inférieur	Non	C ₃
Faible	Inférieur	Oui	C ₃

$$\text{Info}(T) = - \sum_{j=1}^3 \frac{\text{freq}(T, C_j)}{|T|} \log_2 \frac{\text{freq}(T, C_j)}{|T|}$$

$$\text{Info}(T) = - 3/10 \log_2 3/10 - 5/10 \log_2 5/10 - 2/10 \log_2 2/10 = 1.485$$

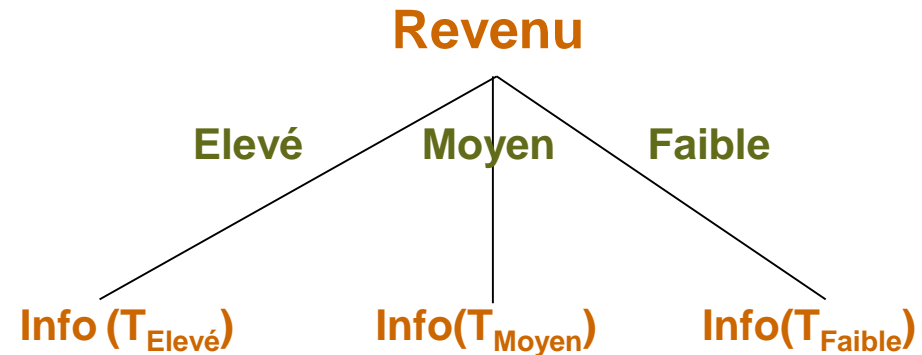
Application (2): $\text{Info}_{\text{Revenu}}(T)$



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C_1
Elevé	Supérieur	Oui	C_2
Elevé	Supérieur	Non	C_1
Elevé	Inférieur	Oui	C_2
Moyen	Supérieur	Non	C_1
Moyen	Supérieur	Oui	C_2
Moyen	Inférieur	Non	C_2
Moyen	Inférieur	Oui	C_2
Faible	Inférieur	Non	C_3
Faible	Inférieur	Oui	C_3

$$\text{Info}_{\text{Revenu}}(T) = \sum_{i \in D_{\text{Revenu}}} \frac{|T_i|}{|T|} \text{Info}(T_i)$$

$$D_{\text{Revenu}} = \{\text{Elevé}, \text{Moyen}, \text{Faible}\}$$



$$\text{Info}(T_{\text{Elevé}}) = - 2/4 \log_2 2/4 - 2/4 \log_2 2/4 = 1$$

$$\text{Info}(T_{\text{Moyen}}) = - 1/4 \log_2 1/4 - 3/4 \log_2 3/4 = 0.812$$

$$\text{Info}(T_{\text{Faible}}) = - 2/2 \log_2 2/2 = 0$$

$$\text{Info}_{\text{Revenu}}(T) = 4/10 \text{Info}(T_{\text{Elevé}}) + 4/10 \text{Info}(T_{\text{Moyen}}) + 2/10 \text{Info}(T_{\text{Faible}}) = 0.725$$

Application (3): Gain ration (T, Revenu)



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C_1
Elevé	Supérieur	Oui	C_2
Elevé	Supérieur	Non	C_1
Elevé	Inférieur	Oui	C_2
Moyen	Supérieur	Non	C_1
Moyen	Supérieur	Oui	C_2
Moyen	Inférieur	Non	C_2
Moyen	Inférieur	Oui	C_2
Faible	Inférieur	Non	C_3
Faible	Inférieur	Oui	C_3

$$\text{Gain}(T, \text{Revenu}) = \text{Info}(T) - \text{Info}_{\text{Revenu}}(T) = 0.761$$

$$\text{Split Info}(T, \text{Revenu}) = - \sum_{i \in D_{\text{Revenu}}} \frac{|T_i|}{|T|} \log_2 \frac{|T_i|}{|T|}$$

$$\text{Split Info}(T, \text{Revenu}) = - 4/10 \log_2 4/10 - 4/10 \log_2 4/10 - 2/10 \log_2 2/10 = 1.522$$

$$\text{Gain Ratio}(T, \text{Revenu}) = \frac{0.761}{1.522} = 0.5$$

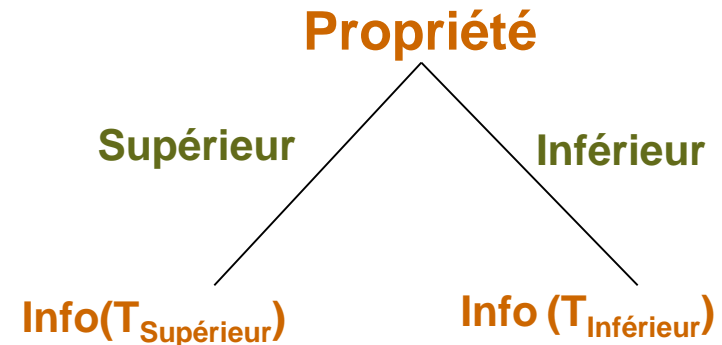
Application (4): Info_{Propriété}(T)



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C ₁
Elevé	Supérieur	Oui	C ₂
Elevé	Supérieur	Non	C ₁
Elevé	Inférieur	Oui	C ₂
Moyen	Supérieur	Non	C ₁
Moyen	Supérieur	Oui	C ₂
Moyen	Inférieur	Non	C ₂
Moyen	Inférieur	Oui	C ₂
Faible	Inférieur	Non	C ₃
Faible	Inférieur	Oui	C ₃

$$\text{Info}_{\text{Propriété}}(T) = \sum_{i \in D_{\text{Propriété}}} \frac{|T_i|}{|T|} \text{Info}(T_i)$$

$$D_{\text{Propriété}} = \{\text{Supérieur}, \text{Inférieur}\}$$



$$\text{Info}(T_{\text{Supérieur}}) = - 3/5 \log_2 3/5 - 2/5 \log_2 2/5 = 0.971$$

$$\text{Info}(T_{\text{Inférieur}}) = - 3/5 \log_2 3/5 - 2/5 \log_2 2/5 = 0.971$$

$$\text{Info}_{\text{Propriété}}(T) = 5/10 \text{Info}(T_{\text{Supérieur}}) + 5/10 \text{Info}(T_{\text{Inférieur}}) = 0.971$$

Application (5): Gain ration (T, Propriété)



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C_1
Elevé	Supérieur	Oui	C_2
Elevé	Supérieur	Non	C_1
Elevé	Inférieur	Oui	C_2
Moyen	Supérieur	Non	C_1
Moyen	Supérieur	Oui	C_2
Moyen	Inférieur	Non	C_2
Moyen	Inférieur	Oui	C_2
Faible	Inférieur	Non	C_3
Faible	Inférieur	Oui	C_3

$$\text{Gain}(T, \text{Propriété}) = \text{Info}(T) - \text{Info}_{\text{Propriété}}(T) = 0.514$$

$$\text{Split Info}(T, \text{Propriété}) = - \sum_{i \in D_{\text{Propriété}}} \frac{|T_i|}{|T|} \log_2 \frac{|T_i|}{|T|}$$

$$\text{Split Info}(T, \text{Propriété}) = - 5/10 \log_2 5/10 - 5/10 \log_2 5/10 = 1$$

$$\text{Gain Ratio}(T, \text{Propriété}) = \frac{0.514}{1} = 0.514$$

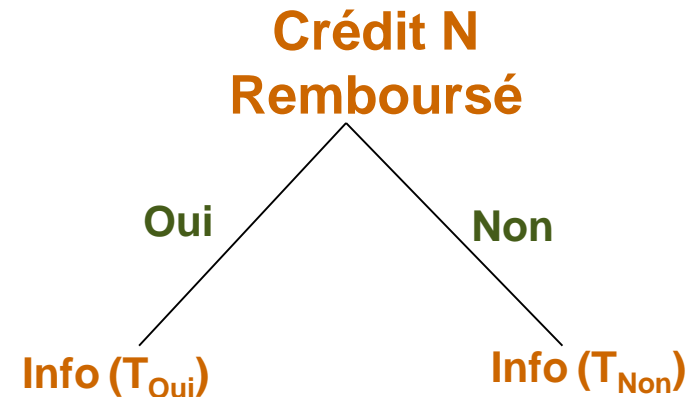
Application (6): Info_{Crédit N remboursé}(T)



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C ₁
Elevé	Supérieur	Oui	C ₂
Elevé	Supérieur	Non	C ₁
Elevé	Inférieur	Oui	C ₂
Moyen	Supérieur	Non	C ₁
Moyen	Supérieur	Oui	C ₂
Moyen	Inférieur	Non	C ₂
Moyen	Inférieur	Oui	C ₂
Faible	Inférieur	Non	C ₃
Faible	Inférieur	Oui	C ₃

$$\text{Info}_{\text{Crédit N remboursé}}(T) = \sum_{i \in D_{\text{Crédit N remboursé}}} \frac{|T_i|}{|T|} \text{Info}(T_i)$$

$$D_{\text{Crédit non remboursé}} = \{\text{Oui}, \text{Non}\}$$



$$\text{Info}(T_{\text{Oui}}) = -4/5 \log_2 4/5 - 1/5 \log_2 1/5 = 0.722$$

$$\text{Info}(T_{\text{Non}}) = -3/5 \log_2 3/5 - 1/5 \log_2 1/5 - 1/5 \log_2 1/5 = 1.371$$

$$\text{Info}_{\text{Crédit N remboursé}}(T) = 5/10 \text{Info}(T_{\text{Oui}}) + 5/10 \text{Info}(T_{\text{Non}}) = 1.046$$

Application (6): Gain ration (T, Crédit N remboursé)



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C_1
Elevé	Supérieur	Oui	C_2
Elevé	Supérieur	Non	C_1
Elevé	Inférieur	Oui	C_2
Moyen	Supérieur	Non	C_1
Moyen	Supérieur	Oui	C_2
Moyen	Inférieur	Non	C_2
Moyen	Inférieur	Oui	C_2
Faible	Inférieur	Non	C_3
Faible	Inférieur	Oui	C_3

$$\text{Gain}(T, \text{Crédit N remboursé}) = \text{Info}(T) - \text{Info}_{\text{Crédit non remboursé}}(T) = 0.439$$

$$\text{Split Info}(T, \text{Crédit non remboursé}) = - \sum_{i \in D_{\text{Crédit N remboursé}}} \frac{|T_i|}{|T|} \log_2 \frac{|T_i|}{|T|}$$

$$\text{Split Info}(T, \text{Crédit non remboursé}) = - 5/10 \log_2 5/10 - 5/10 \log_2 5/10 = 1$$

$$\text{Gain Ratio}(T, \text{Crédit non remboursé}) = \frac{0.439}{1} = 0.439$$

Application (7): Arbre de décision (Niveau 1)

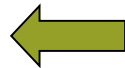


Gain Ratio(T, Revenu) = 0.5

Gain Ratio(T, Propriété) = 0.514

Gain Ratio(T, Crédit non remboursé) = 0.439

Racine



Propriété

Supérieur

Inférieur

Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C ₁
Elevé	Supérieur	Oui	C ₂
Elevé	Supérieur	Non	C ₁
Elevé	Inférieur	Oui	C ₂
Moyen	Supérieur	Non	C ₁
Moyen	Supérieur	Oui	C ₂
Moyen	Inférieur	Non	C ₂
Moyen	Inférieur	Oui	C ₂
Faible	Inférieur	Non	C ₃
Faible	Inférieur	Oui	C ₃

Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C ₁
Elevé	Supérieur	Oui	C ₂
Elevé	Supérieur	Non	C ₁
Moyen	Supérieur	Non	C ₁
Moyen	Supérieur	Oui	C ₂

Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Inférieur	Oui	C ₂
Moyen	Inférieur	Non	C ₂
Moyen	Inférieur	Oui	C ₂
Faible	Inférieur	Non	C ₃
Faible	Inférieur	Oui	C ₃

Application (8): Propriété = Supérieur (1)



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C ₁
Elevé	Supérieur	Oui	C ₂
Elevé	Supérieur	Non	C ₁
Moyen	Supérieur	Non	C ₁
Moyen	Supérieur	Oui	C ₂

$$\text{Info}(T_{\text{Supérieur}}) = \text{Info}(S) = - 3/5 \log_2 3/5 - 2/5 \log_2 2/5 = 0.971$$

Application (9): Propriété = Supérieur (2)



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C ₁
Elevé	Supérieur	Oui	C ₂
Elevé	Supérieur	Non	C ₁
Moyen	Supérieur	Non	C ₁
Moyen	Supérieur	Oui	C ₂

$$\text{Info}_{\text{Revenu}}(S_{\text{Elevé}}) = - \frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} = 0.918$$

$$\text{Info}_{\text{Revenu}}(S_{\text{Moyen}}) = - \frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$\text{Info}_{\text{Revenu}}(S_{\text{Faible}}) = 0$$

$$\text{Info}_{\text{Revenu}}(S) = \left(\frac{3}{5}\right) * 0.918 + \left(\frac{2}{5}\right) * 1 + (0 * 0) = 0.951$$

$$\text{Gain}(S, \text{Revenu}) = 0.02$$

$$\text{Split Info}(S, \text{Revenu}) = - \frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} - 0 = 0.971$$

$$\text{Gain Ratio}(S, \text{Revenu}) = 0.02$$

Application (10): Propriété = Supérieur (3)



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C_1
Elevé	Supérieur	Oui	C_2
Elevé	Supérieur	Non	C_1
Moyen	Supérieur	Non	C_1
Moyen	Supérieur	Oui	C_2

$$\text{Info}_{\text{Crédit non remboursé}}(S_{\text{Oui}}) = - 2/2 \log_2 2/2 = 0$$

$$\text{Info}_{\text{Crédit non remboursé}}(S_{\text{Non}}) = - 3/3 \log_2 3/3 = 0$$

$$\text{Info}_{\text{Crédit non remboursé}}(S) = ((3/5) * 0) + ((2/5) * 0) = 0$$

$$\text{Gain}(S, \text{Crédit non remboursé}) = 0.971$$

$$\text{Split Info}(S, \text{Crédit non remboursé}) = - 2/5 \log_2 2/5 - 3/5 \log_2 3/5 = 0.971$$

$$\text{Gain Ratio}(S, \text{Crédit non remboursé}) = 1$$

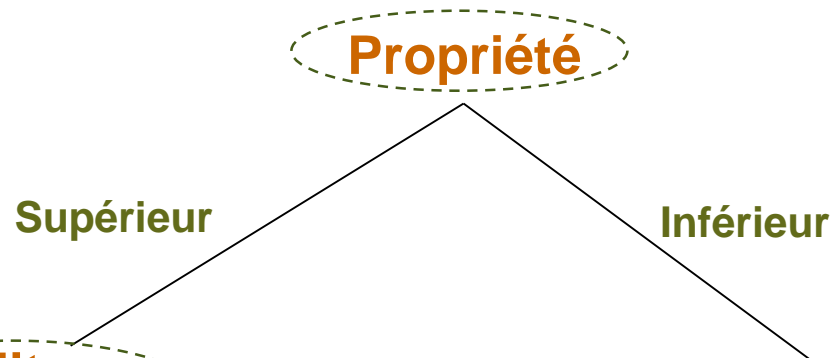
Application (11): Arbre de décision (Niveau 2)



Gain Ratio(S, Revenu) = 0.02

Gain Ratio(S, Crédit non remboursé) = 1

Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C ₁
Elevé	Supérieur	Oui	C ₂
Elevé	Supérieur	Non	C ₁
Moyen	Supérieur	Non	C ₁
Moyen	Supérieur	Oui	C ₂



Crédit non remboursé

Oui

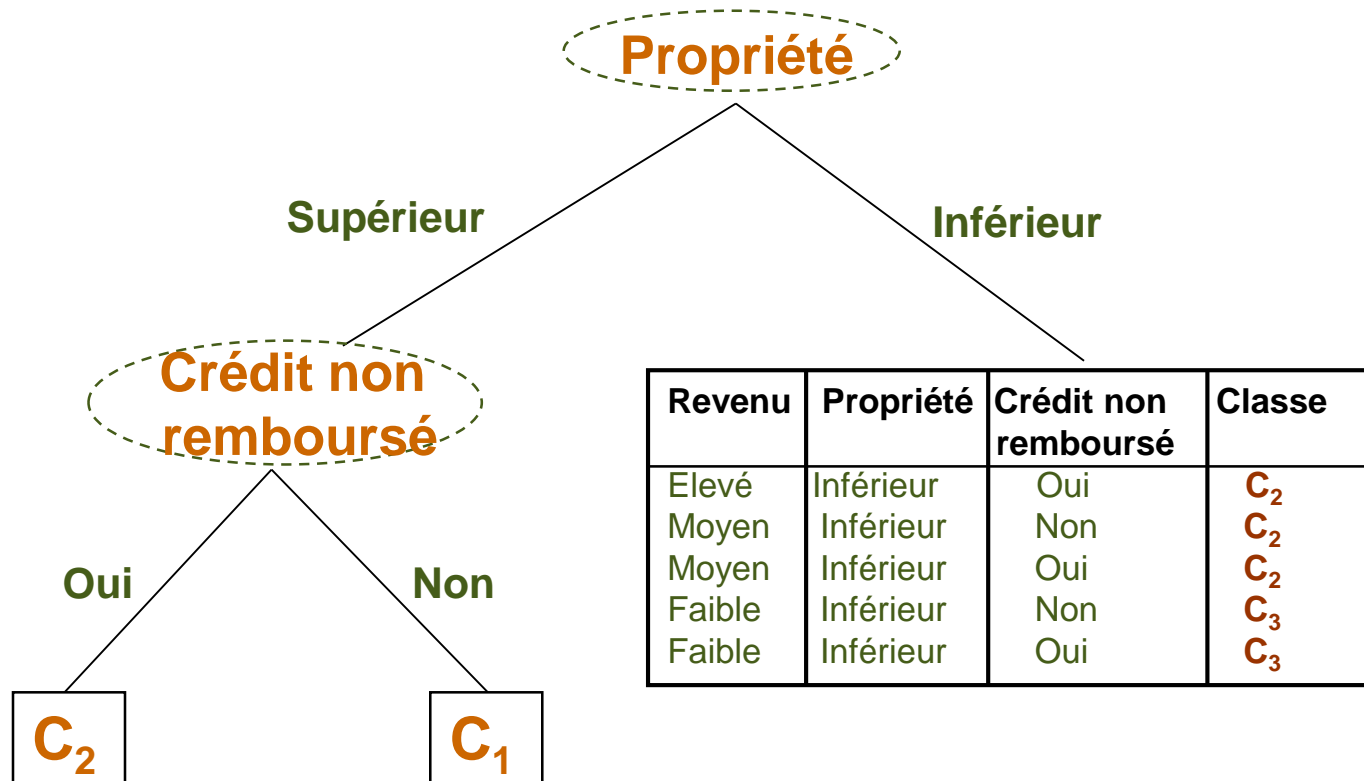
Non

Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Inférieur	Oui	C ₂
Moyen	Inférieur	Non	C ₂
Moyen	Inférieur	Oui	C ₂
Faible	Inférieur	Non	C ₃
Faible	Inférieur	Oui	C ₃

Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Oui	C ₂
Moyen	Supérieur	Oui	C ₂

Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Supérieur	Non	C ₁
Elevé	Supérieur	Non	C ₁
Moyen	Supérieur	Non	C ₁

Application (12): Arbre de décision (Niveau 2)



Application (13): Propriété = Inférieur (1)



Revenu	Propriété	Crédit non remboursé	Classe
Elevé	Inférieur	Oui	C_2
Moyen	Inférieur	Non	C_2
Moyen	Inférieur	Oui	C_2
Faible	Inférieur	Non	C_3
Faible	Inférieur	Oui	C_3

$$\text{Info}(T_{\text{Inférieur}}) = \text{Info}(I) = - 3/5 \log_2 3/5 - 2/5 \log_2 2/5 = 0.971$$

Application...



Continuez les calculs pour
l'application...

Exposés - Élagage



- **Méthodes d'élagage**
 - **MCCP**: Minimal Cost Complexity Pruning (Breiman, 1984).
 - **MEP**: Minimum Error Pruning (Niblett et Bratko, 1986).
 - **CVP**: Critical Value Pruning (Mingers, 1987).
 - **PEP**: Pessimistic Error Pruning (Quinlan, 1987).
 - **REP**: Reduced Error Pruning (Quinlan, 1987, 1993).
 - **EBP**: Error Based Pruning (Quinlan, 1993).
- **Date d'exposé : Pour le 9 mars 2021**