



Classification

L3{*MASS, MA, Info_SI, Info_Reseaux*} - UFR S.A.T

Pr. Ousmane THIARE

othiare@ugb.edu.sn
[www.ousmanethiare.com]

16 avril 2020

Introduction

Classification
hiérarchique

Classification

Chapitre IV : Classification

Introduction

Classification
hiérarchique

1 Introduction

2 Classification hiérarchique



Classification

Introduction

Introduction

Classification hiérarchique

La classification sans a priori est depuis longtemps une problématique importante issue surtout de l'étude des phénomènes naturels et de la biologie en particulier. Toutes les méthodes ainsi développées appartiennent à une science la taxonomie littéralement la science des lois de l'ordre. Les méthodes de classification font parties intégrante de l'analyse de données. Dans le domaine de la reconnaissance des formes elle porte le nom de classification non-supervisée. Le terme anglais pour classification est clustering, les classes étant des clusters. Le terme anglais classification désigne davantage classement i.e. le fait d'affecter des objets à des classes prédéfinies, voire analyse de données en général.



Objectifs

La classification a pour principal objectif de rassembler les éléments (individus ou variables) qui se ressemblent et/ou de séparer ceux qui diffèrent. C'est-à-dire qu'il s'agit de créer des classes homogènes les plus éloignées les unes des autres. Si cet objectif est facilement compréhensible, il n'en est pas moins compliqué à atteindre. Nous sous-entendons lorsque nous cherchons à classer des éléments, qu'il existe des regroupements, soit en nombre inconnu soit en nombre supposé.



Objectifs

Les objectifs de la classification sont donc de regrouper les individus décrits par un ensemble de variables, ou regrouper les variables observées sur des individus et d'interpréter ces regroupements par une synthèse des résultats. L'intérêt de regrouper les individus est ici de les classer en conservant leur caractère multidimensionnel, et non pas seulement à partir d'une seule variable. Si les variables sont nombreuses il peut être intéressant de les regrouper afin de réduire leur nombre pour une interprétation plus facile. Les méthodes de classification sont donc complémentaires des analyses factorielles.



Objectifs

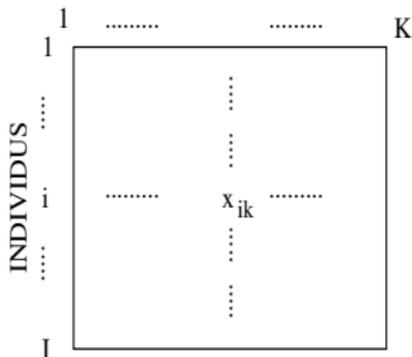


FIGURE: Représentation des données pour la classification



Données

Les données de départ sont souvent organisées comme une matrice X décrite par la figure 1, où x_{ik} est la valeur de la variable k pour l'individu i , I représente à la fois le nombre d'individus et l'ensemble $I = \{1, \dots, I\}$ et K représente à la fois le nombre de variables et l'ensemble $K = \{1, \dots, K\}$.

Les variables peuvent être quantitatives continues ou issues de tableaux de contingence ou encore qualitatives. Afin de traiter l'ensemble de ces types de variables, c'est la mesure de similarité ou dissimilarité qui doit être adaptée aux types de données.



Classification

Introduction

Introduction

Classification hiérarchique

Données

Une mesure de similarité ou de dissimilarité est une distance à l'exception que l'inégalité triangulaire n'est pas exigée. Ces mesures peuvent être des distances dans le cas de variables quantitatives. Ainsi, il est préférable d'employer une distance euclidienne, de Minkowsky pour les variables quantitatives continues et une distance du χ^2 pour des tableaux de contingences.



Méthodes

Il existe un grand nombre de méthodes et surtout beaucoup de variantes. Il est possible de les différencier grossièrement soit par leur structure de classification, soit par le type de représentation des classes. Ainsi, nous pouvons distinguer quatre types de représentation :

- Les partitions sont une notion la plus naturelle, chaque individu est affecté à une classe et une seule.
- Les hiérarchies sont un ensemble de partitions emboîtées. Ainsi une classe se divise en sous-classes.



Méthodes

- Les arbres additifs sont une autre vision des hiérarchies ; une structure dont les noeuds terminaux sont les individus classés et les noeuds intérieurs les classes. Une extension des arbres additifs est la notion d'arbre au sens de la théorie de graphes.
- Les pyramides sont une généralisation des hiérarchies car elles permettent des empiétements entre les classes.



Méthodes

Les méthodes de classification cherchent à transformer le tableau de données en un autre tableau ayant de "bonnes propriétés". C'est donc un problème d'optimisation.



Classification

Classification hiérarchique

Introduction

Classification
hiérarchique

Il existe principalement deux familles d'algorithmes de classification hiérarchique :

- les algorithmes ascendants : la construction des classes se fait par des agglomérations successives des éléments deux à deux,
- les algorithmes descendants : la construction des classes se fait par dichotomies successives de l'ensemble des éléments.

Ces deux approches conduisent à une hiérarchie des partitions des éléments. La seconde approche est beaucoup moins employée que la première, nous présentons donc ici la première approche.



Classification

Classification hiérarchique

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Le principe repose donc sur la création à chaque étape d'une partition obtenue en agrégeant deux à deux les éléments (individus ou plus rarement variables) les plus proches. Les différentes façons de créer un nouveau couple constituent autant de différents algorithmes de classification hiérarchique ascendante



Classification

Classification hiérarchique

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Méthodes d'agrégation

Supposons que le nuage initial, par exemple N_I , à classer est muni d'une distance (ou d'une mesure de similarité ou dissimilarité) d . La façon de regrouper des individus ou des groupes d'individus repose sur des règles de calcul des distances entre ces classes (individus ou groupes d'individus) disjointes, appelées critère d'agrégation.



Classification

Classification hiérarchique

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Méthodes d'agrégation

Soit x , y et z trois classes. Si les classes x et y sont regroupées en une seule classe h , plusieurs critères d'agrégation sont possibles :

- distance du *saut minimal*

$$d(h, z) = \min\{d(x, z), d(y, z)\} \quad (1)$$

- distance du *saut maximal*

$$d(h, z) = \max\{d(x, z), d(y, z)\} \quad (2)$$

- distance *moyenne*

$$d(h, z) = \frac{d(x, z) + d(y, z)}{2} \quad (3)$$



Principe de la classification hiérarchique ascendante

Méthodes d'agrégation

- distance *moyenne généralisée*, en notant n_x et n_y le nombre d'individus de x et y :

$$d(h, z) = \frac{n_x d(x, z) + n_y d(y, z)}{n_x + n_y} \quad (4)$$

Ces méthodes d'agrégation ont l'avantage de conduire à des calculs simples et possèdent des propriétés mathématiques intéressantes. Cependant, les résultats ne sont pas toujours bons. En particulier, la distance du saut minimal peut entraîner des *effets de chaîne*, illustrés sur la figure 1. Sur le nuage de points représenté sur cette figure, les groupes A et B ne sont pas facilement discernables par la distance du saut minimal.



Principe de la classification hiérarchique ascendante Méthodes d'agrégation

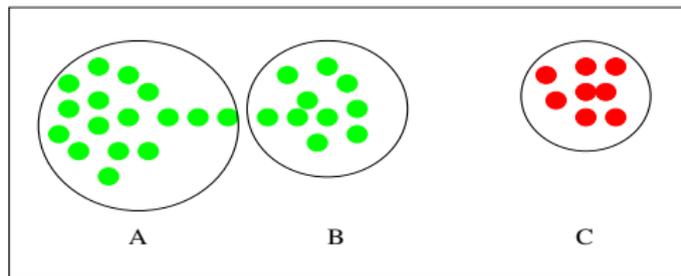


FIGURE: Illustration de l'effet de chaîne



Classification

Classification hiérarchique

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Méthodes d'agrégation

Il est difficile de déterminer au niveau de la chaîne quels points appartiennent à A et quels points appartiennent à B. Le critère de la distance moyenne donne de meilleurs résultats, mais comme nous le voyons sur la figure (les classes ont alors des formes de cercles), elle a tendance à considérer A et B comme deux classes, alors qu'il s'agit d'un seul sous-nuage.



Classification

Classification hiérarchique

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Méthodes d'agrégation

Pour remédier à ce problème, des critères d'agrégation selon la variance sont liés à des calculs d'inertie. Cette méthode est particulièrement facile à mettre en oeuvre après une analyse factorielle, les éléments étant donnés par leurs coordonnées sur les premiers axes factoriels.



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante Méthodes d'agrégation

Cette méthode porte également le nom de la méthode de Ward. La solution au problème évoqué ci-dessus est donc de considérer les éléments - prenons les individus - comme un nuage de points N_I dans \mathbb{R}^K . L'idée est ensuite d'agréger les individus en minimisant l'inertie (ou la variance) intraclasse et en maximisant l'inertie interclasse. Ainsi l'inertie totale du nuage N_I est égale à la somme de l'inertie interclasse et de l'inertie intraclasse :

$$I = I_{intra} + I_{inter} \quad (5)$$

Reprenons la figure 2 illustrant cette proposition. Le même nuage est représenté deux fois en reliant les points pour le calcul de l'inertie totale à gauche et de la somme des inerties interclasse et intraclasse à droite.



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante Méthodes d'agrégation

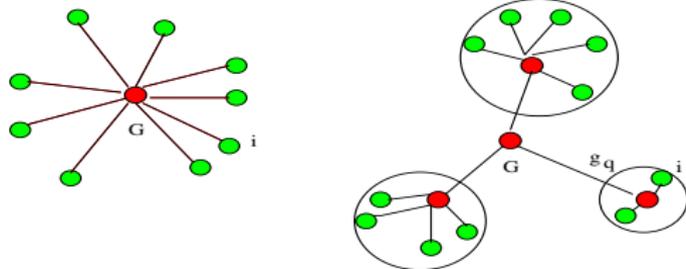


FIGURE: Illustration de la formule de Huygens



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante Méthodes d'agrégation

Considérons que chaque individu i est muni d'une masse m_i et chaque classe q est munie d'une masse m_q . Avec les notations de la figure 2, la formule de décomposition de Huygens s'écrit :

$$I = \sum_{q \in Q} m_q d^2(g_q, G) + \sum_{q \in Q} \sum_{i \in I_q} m_i d^2(x_i, g_q) \quad (6)$$

où d représente la distance choisie initialement, g_q est le centre de gravité du sous-nuage N_{I_q} et G le centre de gravité du nuage des individus N_I .



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Méthodes d'agrégation

Ainsi la qualité globale d'une partition est liée à l'homogénéité interne des sous-nuages et donc également à l'éloignement des sous-nuages. Par exemple, la figure 3 illustre deux partitions en deux sous-nuages, celui de gauche avec une inertie intraclasse faible, celui de droite avec une inertie intraclasse élevée.

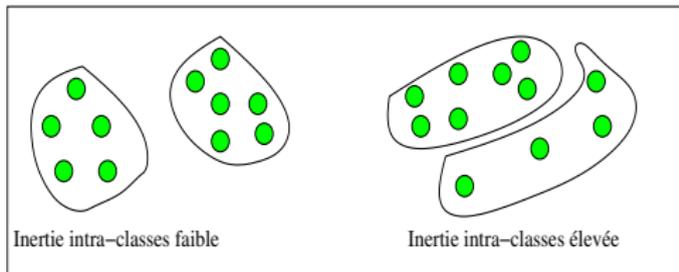


FIGURE: Illustration de la formule de Huygens



Classification

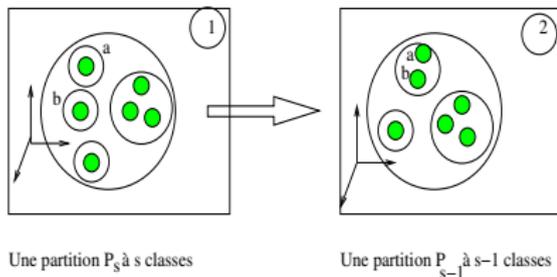
Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante Méthodes d'agrégation

Ainsi pour une agrégation, nous cherchons à faire varier le moins possible l'inertie intraclasse, ce qui est équivalent à rendre minimale la perte d'inertie interclasse résultant de cette agrégation. Considérons une partition P_s à s classes (ou sous-nuages), en associant deux classes a et b à P_s , nous obtenons une partition P_{s-1} à $s-1$ classes (cf. figure 4).



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante Méthodes d'agrégation

L'élément c obtenu par l'agrégation de a et b a pour masse $m_c = m_a + m_b$, et il peut être décrit par son centre de gravité donné par :

$$c = \frac{m_a a + m_b b}{m_a + m_b} \quad (7)$$

L'inertie interclasse de a et de b peut se décomposer par la formule de Huygens par :

$$I_{inter(ab)} = m_a d^2(a, G) + m_b d^2(b, G) \quad (8)$$

$$= m_a d^2(a, c) + m_b d^2(b, c) + m_c d^2(c, G) \quad (9)$$



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Méthodes d'agrégation

Or l'inertie de la partition P_s est donnée par :

$$I_s = I_{inter(ab)} + I_{intra(a)} + I_{intra(b)} \quad (10)$$

et celle de la partition P_{s-1} par :

$$I_{s-1} = I_{inter(c)} + I_{intra(a)} + I_{intra(b)} \quad (11)$$

$$= m_c d^2(c, G) + I_{intra(a)} + I_{intra(b)} \quad (12)$$



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante Méthodes d'agrégation

Ainsi la perte d'inertie $\Delta I_{inter(ab)}$ due au passage de la partition P_s à la partition P_{s-1} est donnée par :

$$\Delta I_{inter(ab)} = I_{inter(P_s)} + I_{inter(P_{s-1})} = m_a d^2(a, c) + m_b d^2(b, c) \quad (1)$$

En remplaçant c par sa valeur en fonction de a et b , nous obtenons :

$$\Delta I_{inter(ab)} = \frac{m_a m_b}{m_a + m_b} d^2(a, b) \quad (14)$$



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Méthodes d'agrégation

Cette variation représente un indice de dissimilarité (appelé aussi indice de niveau) qui est l'inertie de l'haltère (ou variance du dipôle) (a ; b). Il est aisé de vérifier que la somme des indices de dissimilarité entre toutes les partitions est l'inertie totale du nuage N_I . Le principe de la méthode de Ward est donc de déterminer les éléments a et b d'une partition P_s qui ont un indice de dissimilarité minimal.



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante Algorithme

L'algorithme de classification hiérarchique ascendante est simple et facile à programmer. Son déroulement suit les étapes suivantes :

- Étape 1 : Nous considérons le nuage N_I comme une partition P_I de I éléments.
- Étape 2 : Une transformation des données s'effectue par la construction à partir de la matrice X décrite par la figure 1. d'une matrice de distances entre les I individus, à partir de la distance retenue initialement. Nous recherchons ensuite les deux éléments à agréger (i.e. les deux éléments les plus "proches" en terme de distance ou d'indice de dissimilarité). L'agrégation des deux éléments fournit une partition P_{I-1} à $I-1$ individus.



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Algorithme

- Étape 3 : Nous construisons la nouvelle matrice $((I - 1) \times (I - 1))$ des distances, puis nous recherchons les deux nouveaux éléments à agréger. L'agrégation des deux éléments fournit une partition à $I-2$ individus.
- Étape m : Nous calculons la matrice $((I - (m - 1)) \times (I - (m - 1)))$ des distances, puis nous cherchons à agréger deux éléments jusqu'à ce qu'il n'en reste plus qu'un qui constitue la dernière partition P_1 .



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante Algorithme

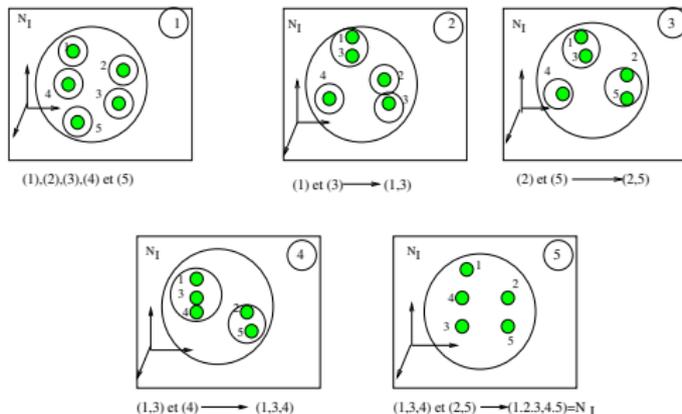


FIGURE: Illustration de l'algorithme de classification avec un nuage $I=5$ individus



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante Algorithme

Noeud	(6)	(7)	(8)	(9)
benjamin	(1)	(2)	(6)	(8)
ainé	(3)	(5)	(4)	(7)
effectif	2	2	3	5

FIGURE: Relations entre les noeuds de l'arbre



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Algorithme

Afin d'illustrer cet algorithme, nous donnons un exemple d'un nuage N_i de cinq individus sur la figure 6.

Les étapes successives de cet algorithme peuvent être représentées par un arbre hiérarchique également appelé dendrogramme où sont représentées en ordonnées les indices de dissimilarité (cf. figure 6.).



Principe de la classification hiérarchique ascendante Algorithme

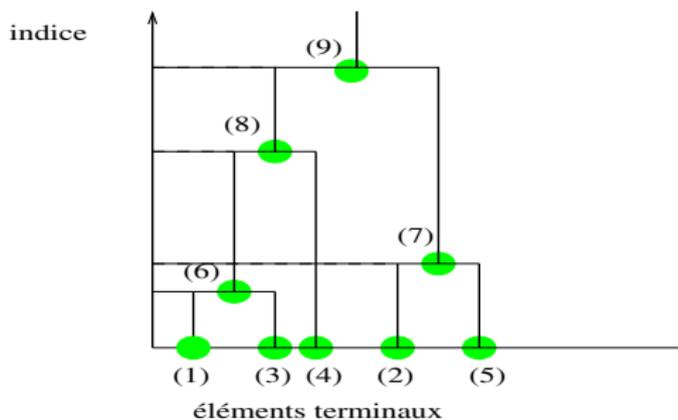


FIGURE: Exemple de dendrogramme



Principe de la classification hiérarchique ascendante

Algorithme

Vocabulaire lié au dendrogramme

- Les *éléments terminaux* de l'arbre (ou de la hiérarchie) sont les individus (ou variables selon ce qui est classé).
- Les noeuds de l'arbre correspondent aux regroupements de deux éléments appelés aîné et benjamin. L'arbre de la figure 7 peut ainsi être décrit par la figure 6.
- L'agrégation repose sur les inégalités des distances entre elles. Nous pouvons obtenir le même classement en des couples d'éléments en classant ces couples par ordre croissant des distances. Un tel classement est appelé *ordonnance*.



Principe de la classification hiérarchique ascendante

Algorithme

Vocabulaire lié au dendrogramme

- La hiérarchie peut être décrite par une famille H d'éléments de I telle que :
 - $I \in H, \{i\} \in H \forall i \in I,$
 - $\forall A, B \in H, A \cap B \in \{A, B, \emptyset\}$ i.e. deux classes sont soit disjointes, soit l'une est incluse dans l'autre.Ainsi toute classe est la réunion des classes qui sont incluses en elle. La famille des sous-ensembles construits par la classification ascendante hiérarchique forme une hiérarchie. C'est en fait une hiérarchie binaire, il en existe d'autres.



Principe de la classification hiérarchique ascendante

Algorithme

Vocabulaire lié au dendrogramme

- - Une *hiérarchie indicée* est une hiérarchie pour laquelle il existe une fonction v de H dans \mathbb{R}^+ i.e. telle que :

$$A \subset B \Leftrightarrow v(A) \leq v(B), \forall A, B \in H \quad (15)$$

La hiérarchie est généralement indiquée par les valeurs des distances (ou indices de dissimilarité) correspondant à chaque étape d'agrégation.

- En coupant l'arbre par une droite horizontale, nous obtenons une *partition*. Une hiérarchie donne ainsi une chaîne de l partitions de 1 à l classes.



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Les hiérarchies indiquées ont une propriété particulièrement intéressante, car elle peuvent être vues comme un ensemble muni d'une *ultramétrie*. Une ultramétrie est une distance d particulière. En tant que distance d associée au nuage N_I , elle est une application qui vérifie :

- $x = y \Leftrightarrow d(x, y) = 0, \quad \forall x, y \in N_I,$

Cette distance d est une ultramétrie si elle vérifie une condition plus forte que l'inégalité triangulaire donnée par $d(x, y) \leq d(x, z) + d(y, z), \quad \forall x, y, z \in N_I$. La distance du saut minimal est la plus grande ultramétrie inférieure à la métrique d initiale.



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Les hiérarchies indiquées ont une propriété particulièrement intéressante, car elle peuvent être vues comme un ensemble muni d'une *ultramétrie*. Une ultramétrie est une distance d particulière. En tant que distance d associée au nuage N_I , elle est une application qui vérifie :

- $x = y \Leftrightarrow d(x, y) = 0, \forall x, y \in N_I,$
- $d(x, y) = d(y, x), \forall x, y \in N_I$ (relation de symétrie),

Cette distance d est une ultramétrie si elle vérifie une condition plus forte que l'inégalité triangulaire donnée par $d(x, y) \leq d(x, z) + d(y, z), \forall x, y, z \in N_I$. La distance du saut minimal est la plus grande ultramétrie inférieure à la métrique d initiale.



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Principe de la classification hiérarchique ascendante

Les hiérarchies indiquées ont une propriété particulièrement intéressante, car elle peuvent être vues comme un ensemble muni d'une *ultramétrie*. Une ultramétrie est une distance d particulière. En tant que distance d associée au nuage N_I , elle est une application qui vérifie :

- $x = y \Leftrightarrow d(x, y) = 0, \forall x, y \in N_I$,
- $d(x, y) = d(y, x), \forall x, y \in N_I$ (relation de symétrie),
- $d(x, y) \leq d(x, z) + d(y, z), \forall x, y, z \in N_I$ (inégalité triangulaire).

Cette distance d est une ultramétrie si elle vérifie une condition plus forte que l'inégalité triangulaire donnée par $d(x, y) \leq d(x, z) + d(y, z), \forall x, y, z \in N_I$. La distance du saut minimal est la plus grande ultramétrie inférieure à la métrique d initiale.



Classification

Agrégation selon l'inertie

Introduction

Classification
hiérarchique

Interprétation

L'interprétation repose essentiellement sur la lecture du dendrogramme. Elle devient problématique lorsque le nombre d'individus est très important. Elle doit se faire de haut en bas afin d'examiner d'abord les partitions qui possèdent peu de classes, pour ensuite entrer dans des considérations plus détaillées. Nous cherchons, essentiellement la partition qui présente le plus d'intérêt. Pour cela, il faut chercher à construire des classes homogènes. Une bonne partition, i.e. une bonne coupure de l'arbre, doit comporter peu de classes avec une inertie intraclasse faible et une inertie interclasse élevée. Pour le choix de la coupure, nous pouvons également nous aider de la courbe des indices. Ainsi nous devons rechercher le noeud après lequel il y a une perte d'indice importante. Ceci peut également se lire sur le dendrogramme.

