République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme **:**

MASTER en Mathématiques

Option : **Statistique**

Par

NITA Manel

Titre :

Méthodes de Classification Automatique

Membres du Comité d'Examen :

Dr.	**MERAGHNI Djamel**	UMKB	Encadreur
Dr.	**SAYAH Abdallah**	UMKB	Président
Dr.	**ROUBI Afaf**	UMKB	Examinateur

Juin 2019

Dédicace

≪Toutes les lettres ne sauraient trouver les mots qu'il faut... Tous les mots ne sauraient exprimer la gratitude,l'amour, Le respect, la reconnaissance.≫

Je dédie ce mémoire

A mes chères parents **Fateh** et **Djamila**

Pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études,

A mes chères sœurs **Souad** et **Nerdjes**

Pour leurs encouragements permanents, et leur soutien moral

A toute ma famille sans exception Pour leur soutien tout au long de mes études.

A toute mes amis **Rona**,**Noujoud**,**Raouia**, **Meriem**,**Rachida**,**Khaoula**,**Imen**

Merci d'être une partie merveilleuse de ma vie et vous le resterez inchallah.

REMERCIEMENTS

Avant toute chose, je remercie $\mathcal{ALLAH}$ le tout puissant de m'avoir donnée courage,patience et force durant toutes ces années d'étude

J'adresse mes sincères remerciements à tous les professeurs, intervenants et toutes les personnes qui par leurs paroles, leurs écrits, leurs conseils et leurs critiques ont guidé mes réflexions et ont accepté de me rencontrer et de répondre à mes questions durant mes recherches.

Je voudrais dans un premier temps remercier, mon encadreur de mémoire **M.MERAGHNI**, professeur à l'université de Mohamed KHIDER Biskra , pour sa patience, sa disponibilité et surtout ses conseils.

Mes sincères remerciements mon oncle **Ahmed CHIBAT**, docteur à l'université de Mentouri Constantine à pour sa patience et ses conseils dans mes études universitaires.

Je remercie également toute l'équipe pédagogique de l'université de Mohamed KHIDER Biskra et les intervenants professionnels responsables de ma encadrer.

Table des matières

Table des figures

Liste des tableaux

Introduction

La classification est depuis longtemps une problématique importante issue surtout de l'étude des phénomènes naturels et de la biologie en particulier. En mathématiques, on s'intéresse à la classification automatique ou non supervisée (clusternig ou cluster analysis en anglais). La classification automatique regroupe l'ensemble des méthodes statistiques visant à détecter des groupes, généralement appelés classes, dans un échantillion d'objets. L'essence de cette classification est que contrairement à l'analyse discrimante (classification supervisée), il n'est pas nécessaire de connaître à priori la structure d'un groupe.

Le but de la classification automatique est de regrouper les observations similaires et à séparer celles qui sont dissimilaires, comme le montre la figure 1 Puisque les mesure ou les notions de similarité peuvent être explicitées de multiples façons, alors de nombreuses méthodes de classification automatique ont été proposées depuis les années 1930.

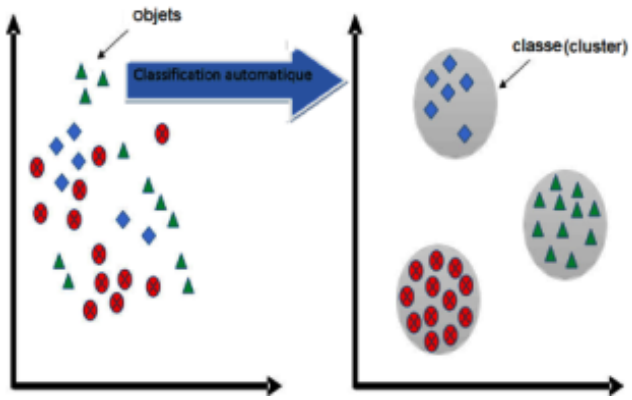

FIG. 1 – Classification automatique

En général, on peut parler de classification automatique si aucune information n'est disponible concernant l'appartenance de certaines données à certaines classes connues. Par

ailleurs, le nombre de groupes recherchés peut être connu a priori ou non. On peut résumer la classification automatique par les étapes suivantes :

1. Calcul des dissimilarités entre les individus.
2. Choix d'un algorithme de classification et exécution.
3. Interprétation des résultats : évaluation de la qualité de la classification et description des classes obtenues.

Ce travail, consacré aux deux méthodes de classification automatique (hiérarchique et non hiérarchique), se compose de deux chapitres.ces deux méthodes s'inscrivent dans une même perspective : l'analyse exploratoire d'un tableau rectangulaire $(n \times p)$.

Chapitre 1 : Classification hiérarchique

Dans ce chapitre, après quelques rappels statistiques, on présente les deux types de classification hiérarchique : classification ascendante hiérarchique (CAH) et classification descendante hiérarchique (CDH). On décrit en détail les algorithmes de la CAH et on traite un exemple d'application.

Chapitre 2 : Classification non hiérarchique

Dans ce chapitre, on présente d'abord quelques rappels de notions nécessaires, puis on propose les types les plus connus de la classification non hiérarchique : K-means, nuées dynamiques, centres mobiles. enfin, on applique les resultats théorique sur le même exemple du premier chapitre.

Les resultats numériques et les représentations graphiques de ce mémoire sont obtenus au moyen du logiciel de traitement et analyse statistiques R, introduit par R. Ihaka et R. Gentlman [12]

Chapitre 1

Classification hiérarchique

La classification hiérarchique (ou hierarchical clustering en anglais) constitue depuis longtemps une forme de classification très populaire. C'est l'une des approches les plus importantes pour l'exploration des données multivariées. L'objectif est d'identifier des groupes d'objets similaires dans un ensemble d'objets. Elle est utilisée dans différents domaines : la biologie [13], la taxinomie [1], la phytosociologie [21], les réseaux de télécommunications [11],... Elle a l'avantage d'être interprétable visuellement à l'aide des graphes ou arbres hiérarchiques. On distingue deux types de classification hiérarchique, schématisés dans la figure 1.1.

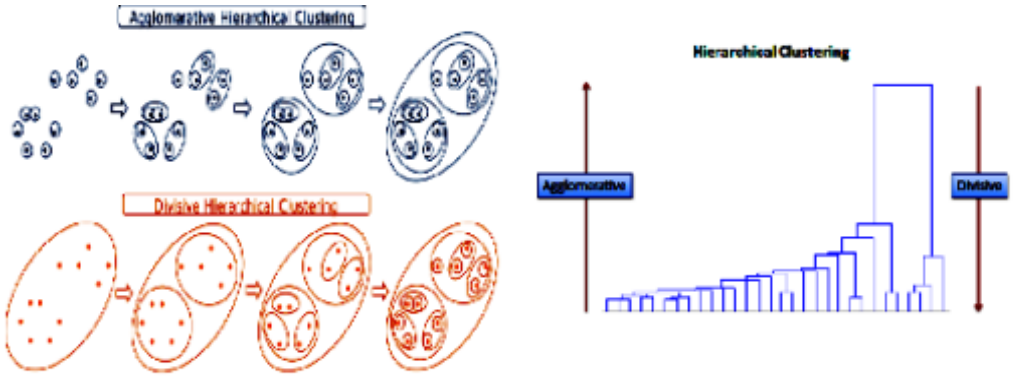

FIG. 1.1 – Classifications hiérarchiques agglomérative et divisive (src :[8])

⊛ **Classification ascendante hiérarchique, notée CAH, (agglomérative)** : à partir des éléments de départ (éléments terminaux), on forme de petites classes ne comportant

que les objets les plus semblables, puis à partir de celles-ci, on construit des classes de moins en moins homogènes jusqu'à obtenir la classe tout entière qui réunit tous les éléments.

⊛ **Classification descendante hiérarchique, notée CDH, (divisive) :** il s'agit d'une dichotomie de la classe entière jusqu'à obtenir tous les éléments terminaux. Elle favorise une évaluation décroissante opposée au premier type.

1.1 Données statistiques

La statistique est l'étude d'un phénomène par la collecte de données, leur traitement, leur analyse, l'interprétation des résultats et leur présentation afin de rendre les données compréhensibles par tous. Autant les arbres constituent la matière première du papier, autant, les données sont la matière brute d'où naît l'information. On pourrait définir les données comme des faits ou des chiffres se prêtant à des conclusions.

Définition 1.1 (Variable quantitative) *Ce sont les variables qui prennent des valeurs numériques(taille ,âge,...). S'expriment par des nombres réels sur lesquels les opérations arithmétiques de base (somme, moyenne,...) ont un sens. Peuvent être discrètes ou continues.*

Définition 1.2 (Variable qualitative) *Ces sont les variables qui prennent des valeurs non numériques , dans le sens où les opérations de base n'ont pas de sens. Les valeurs qu'elles prennent sont appelées des catégories, ou modalités. Ces variables sont exprimées sous forme littérale (par un mot, une phrase ou un code).Elle peut être* ***ordinale*** *ou* ***nominale****.*

On considère un ensemble $\Omega = \{e_1, \ldots, e_n\}$ de $n \geq 1$ individus décrits par $p \geq 1$ variables $x_1, ..., x_p$. La quantité $x_{ij} := x_j(e_i)$ représente la valeur prise par la variable x_j sur l'individu e_i.

Définition 1.3 (Tableau des données) *L'ensemble des valeurs x_{ij} est présenté sous la forme d'une matrice X, de n lignes et p colonnes, appelée tableau des données :*

$$X := (x_{ij})_{n\times p} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{bmatrix} \in M_{n,p}(\mathbb{R})$$

Remarque 1.1

1. *Chaque variable peut être représentée par un vecteur de dimension n , appelé* ***vecteur variable*** *, correspondant aux valeurs prises par cette variable sur les n individus. On la note par x_j.*

$$x_j = (x_{1j}, \dots, x_{nj})^t \in \mathbb{R}^n.$$

2. *Chaque individu est décrit par p variables, formant un vecteur de dimension p, appelé* ***vecteur individu.*** *On la note par e_i ou ω_i.*

$$e_i = (x_{1j}, \dots, x_{nj})^t \in \mathbb{R}^p.$$

1.2 Mesure d'éloignement

Tout système ayant pour but d'analyser ou d'organiser automatiquement un ensemble de données ou de connaissances doit utiliser, sous une forme ou une autre, un opérateur capable d'évaluer précisément les ressemblances ou les dissemblances qui existent entre ces données. La notion de ressemblance (ou Proximité) a fait l'objet d'importantes recherches dans des domaines extrêmement divers. Pour qualifier cet opérateur, plusieurs notions comme la similarité, la dissimilarité ou la distance peuvent être utilisées.

1.2.1 Définitions

On appel similarité ou dissimilarité toute application à valeurs numériques qui permet de mesurer le lien entre les individus d'un même ensemble. Pour une similarité le lien entre deux individus sera d'autant plus fort que sa valeur est grande. Pour une dissimilarité le lien sera d'autant plus fort que sa valeur dissimilarité est petite.

Définition 1.4 (Indice de similarité) *Un indice de similarité sur* Ω *est une application* $S : \Omega \times \Omega \to \mathbb{R}^+$ *et tel que* $\forall e_1, e_2 \in \Omega$ *avec* $e_1 \neq e_2$ *vérifier les propriétés suivant :*

- $S(e_1, e_2) = S(e_2, e_1)$. *(Symétrie).*
- $S(e_1, e_1) = S(e_2, e_2) = S_{\max} > S(e_1, e_2)$.

Définition 1.5 (Indice de dissimilarité) *Un indice de dissimilarité sur* Ω *est une application* $Dis : \Omega \times \Omega \to \mathbb{R}^+$ *et tel que* $\forall e_1, e_2 \in \Omega$ *avec* $e_1 \neq e_2$*vérifier les propriétés suivant :*

- $Dis(e_1, e_2) = Dis(e_2, e_1)$
- $Dis(e_1, e_2) = Dis(e_2, e_1) \Leftrightarrow e_1 = e_2$.

1.2.2 Distances

Dans la vie courante, la distance entre deux points est un nombre positif qui mesure l'écart entre ces deux points. En mathématiques, on formalise ce concept de la manière suivante.

Définition 1.6 (Distance) *On appelle distance sur* Ω *une application* $d : \Omega \times \Omega \to \mathbb{R}^+$ *et tel que* $\forall e_1, e_2 \in \Omega$ *vérifier les propriétés suivant :*

- $d(e_1, e_2) = 0 \Leftrightarrow e_1 = e_2$
- $d(e_1, e_2) = d(e_2, e_1)$
- $d(e_1, e_2) \leq d(e_1, e_3) + d(e_1, e_3)$. $\forall e_3 \in \Omega$.

Remarque 1.2 *Une distance est une dissimilarité mais L'inverse n'est pas vrai.*

1.2.3 Mesure de distance

La Classification Hiérarchique utilise des mesures de dissemblance ou de distance entre les objets pour former des classes. Lorsque les données sont quantitatives les distances classiques sont :

Les distances définies par :

$$d_M^2(e_i, e_i) =^t (e_i - e_i) M (e_i - e_i)$$

où M est une matrice carrée, symétrique, d'ordre p définie positive, appelée **métrique.**

Remarque 1.3 *si $M = I$, d est la distance euclidienne simple, si $M = D_{1/S^2}$ (matrice diagonale des inverses des variances empiriques des p variables) On parle de distance euclidienne normalisée par l'inverse de la variance, si $M = V^{-1}$, d est la distance de Mahalanobis.*

Distances euclidienne

C'est probablement le type de distance le plus couramment utilisé. Il s'agit simplement d'une distance géométrique dans un espace multidimensionnel On appelle distance euclidienne entre e_i et $e_{i'}$ la distance :

$$d(e_i, e_{i'}) = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2}. \forall i, i' = 1, ..., n.$$

Remarque 1.4

1. *L'un des problèmes avec cette distance est qu'elle sensible à l'échelle des variables. Si l'une des variables est plus dispersée que les autres alors elle va les dominer dans les calculs.*
2. *Pour cela il est recommandé de transformer les variables ou de les mettre à l'échelle avant les calculs.*

Distances de Manhattan

Nommée ainsi car elle sert à « mesurer » la distance parcourue par une voiture dans la ville de Manhattan c'est pour ça cette méthode nommée aussi **taxi-distance**, ça formule est :

$$d\left(e_i, e_{i'}\right) = \sum_{j=1}^{p} \left|x_{ij} - x_{i'j}\right| .\forall i, i' = 1, ..., n.$$

Distance de Minkowski

La distance de Minkowski est une mesure de distance entre deux points de l'espace vectoriel normé (espace réel à N dimensions) et est une généralisation de la distance euclidienne et de la distance de Manhattan.

$$d\left(e_i, e_{i'}\right) = \left(\sum_{j=1}^{p} \left|x_{ij} - x_{i'j}\right|^q\right)^{\frac{1}{q}} .q \geq 1.\forall i, i' = 1, ..., n$$

Remarque 1.5

▷ *Lorsque le paramètre $q = 1$, nous avons la métrique **city block (Manhattan)**.*

▷ *Quand $q = 2$, nous avons la distance euclidienne.*

Distance de Canberra

La distance de Canberra est une mesure numérique de la distance entre des paires de points dans un espace vectoriel, introduite en 1966 et affinée en 1967 par Lance.G.N. et Williams.W. T.

$$d\left(e_i, e_{i'}\right) = \sum_{j=1}^{p} \frac{x_{ij} - x_{i'j}}{x_{ij} + x_{i'j}} .\forall i, i' = 1, ..., n.$$

Distance Tchebyshev

La distance de Tchebyshev est la mesure de distance qui représente la distance maximale absolue dans une dimension à deux points de n dimension. Il a des applications réelles dans les échecs et dans de nombreux autres domaines.

$$d\left(e_i, e_{i'}\right) = \max_{j=1,...,p} \left|x_{ij} - x_{i'j}\right| .\forall i, i' = 1, ..., n.$$

Distance de Mahalanobis

Cette distance tient compte de la covariance. Elle est donnée par la formule suivante :

$$d\left(e_i, e_{i'}\right) = \left(x_i - x_j\right)^t V^{-1} \left(x_i - x_j\right).$$

Où V^{-1} est l'inverse de matrice de covariance. Elle doit être estimée en utilisant les données.

Remarque 1.6 *Dans le cas des données binaires, nous avons d'autres types de mesures de dissimilarité : Jaccard, Dice ou Czekanowski, Ochiai, Russel et Rao, Rogers et Tanimoto. Pour des détails sur ces mesures, voir [2, page* $29 - 30$*].*

1.3 Espaces de classification

Définition 1.7 (Partition) *On appelle une partition* $P = \{C_i; i \in I\}$ *,où* I *est un ensemble d'indices, de* Ω *un ensemble de parties* $C_i \subset \Omega$ *possédant les propriétés :*

- $\forall i \in I,\ C_i \neq \emptyset$.
- $\forall i, j \in I,\ i \neq j \Rightarrow C_i \cap C_j = \emptyset$.
- $\cup_{i \in I} C_i = P$.

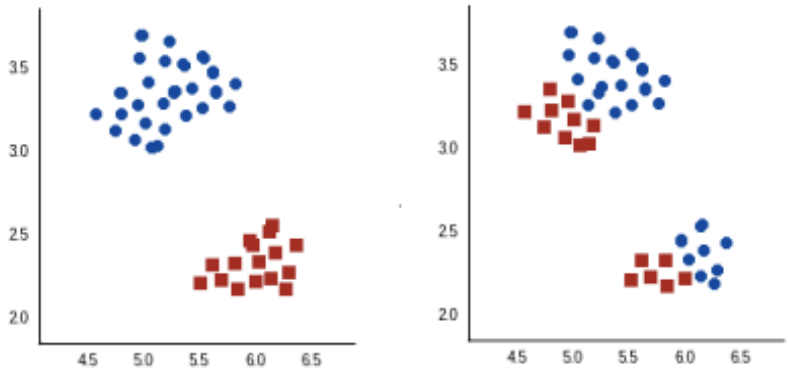

FIG. 1.2 – Deux-partition-en-classe-rouge-et-blue(src :[20])

Exemple 1.1 *Pour* $\Omega = \{1, \cdots, 7\}$, *l'ensemble* $P = \{C_1, C_2, C_3\}$, *avec* $C_1 = \{7\}$, $C_2 = \{4, 5, 6\}$ *et* $C_3 = \{1, 2, 3\}$ *est une partition en trois classes.*

Définition 1.8 (Hiérarchie) *On appelle hiérarchie* H *de* Ω *tout ensemble de parties de* Ω *vérifiant*

▷ $\emptyset, \Omega \notin H$.

▷ $\forall e \in \Omega, \{e\} \in H$ *: (la hiérarchie contient tous les singletons).*

▷ $\forall C_1, C_2 \in H$, $C_1 \cap C_2 = \emptyset$ *ou* $C_1 \subset C_2$ *(ou* $C_2 \subset C_1$*) : (deux classes de la hiérarchie sont soit disjointes soit contenues l'une dans l'autre.)*

Remarque 1.7 *Une hiérarchie* H *est dite indicée s'il existe une application* i *positive définie sur* H *telle que*

▷ $\forall e \in \Omega$, $i(\{e\}) = 0$.

▷ $A \subseteq B \Rightarrow i(A) \leq i(B)$.

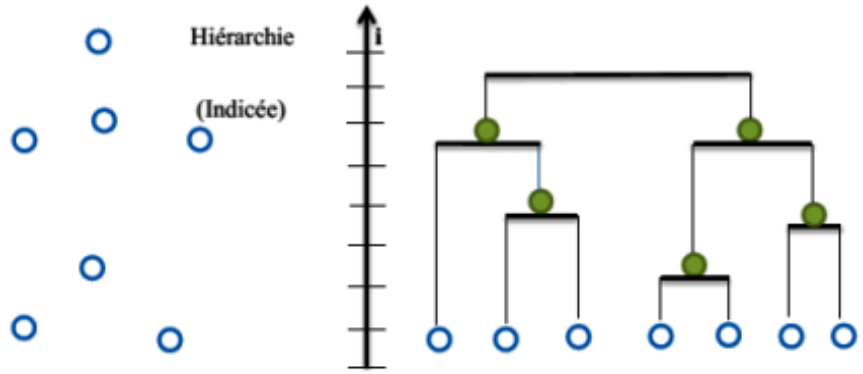

FIG. 1.3 – Hiérarchie indicée (src :[23])

1.4 Classification ascendante hiérarchique(CAH)

La classification ascendante hiérarchique **(CAH)** n'est pas la technique d'analyse de données la plus ancienne mais la problématique de la classification date de quelques milliers d'années, c'est essentiellement un outil au service du marketing.

1.4.1 Principe de CAH

Le principe de la CAH est de rassembler des individus selon un critère de ressemblance défini au préalable qui s'exprimera sous la forme d'une matrice de distances, exprimant la distance existant entre chaque individu pris deux à deux. Deux observations identiques auront une distance nulle. Plus les deux observations seront dissemblables, plus la distance sera importante. La CAH va ensuite rassembler les individus de manière itérative afin de produire un dendrogramme ou arbre de classification. La classification est ascendante car elle part des observations individuelles ; elle est hiérarchique car elle produit des classes ou groupes de plus en plus vastes, incluant des sous-groupes en leur sein. En découpant cet arbre à une certaine hauteur choisie, on produira la partition désirée.

1.4.2 Matrice de distance

Pour mesurer la proximité entre les paires d'objets, on peut utiliser aussi bien les distances que les dissimilarités qu'on regroupe dans une matrice appelée **matrice des distances** ou **des dissimilarités.** On la note par D.Cette matrice est carrée, symétrique d'ordre n ,définie positive, nulle sur diagonale pour un nuage d'effectif n il y a $\dfrac{n(n-1)}{2}$ distances à calculer.

$$D = \begin{pmatrix} 0 & d(e_1,e_2) & d(e_1,e_3) & \cdots & d(e_1,e_{n-1}) & d(e_1,e_n) \\ d(e_2,e_1) & 0 & d(e_2,e_3) & \cdots & d(e_2,e_{n-1}) & d(e_2,e_n) \\ d(e_3,e_1) & d(e_3,e_2) & \ddots & \cdots & d(e_3,e_{n-1}) & d(e_3,e_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ d(e_{n-1},e_1) & d(e_{n-1},e_2) & d(e_{n-1},e_3) & \cdots & 0 & d(e_{n-1},e_n) \\ d(e_n,e_1) & d(e_n,e_2) & d(e_n,e_3) & \cdots & d(e_n,e_{n-1}) & 0 \end{pmatrix} \in M_{(n,n)}(\mathbb{R})$$

Remarque 1.8 *Généralement pour cette matrice on l'utilise la distance euclidienne.*

1.4.3 Critères d'agrégation

Lorsque les proximités entre les objets sont calculées, nous pouvons déterminer comment ils seront réunis dans des groupes. Pour cela nous utilisons les méthodes de liaisons pour mettre les paires d'objets qui sont proches dans le même groupe binaires (les fusions des objets se font toujours deux à deux). Les groupes nouvellement formés se réunissent encore dans des nouveaux groupes jusqu'à ce que l'arbre hiérarchique soit complété.

Notation 1

- *Soient un groupe r et un groupe s le nombre des objets dans chacun des deux groupes est noté par n_r et n_s.La distance entre les groupes r et s est notée $d_c(r,s)$.*

- **Méthode de liaison simple (ou méthode du plus proche voisin), critère du saut minimal**

La distance entre deux groupes est donnée par la **plus petite** distance entre les objets, où l'un des deux éléments est pris dans le premier groupe et l'autre élément est pris dans le deuxième groupe.

$$d_c(r,s) = \min\left\{d\left(x_{ri}, x_{sj}\right)\right\}; i = 1,\ldots,n_r, j = 1,\ldots,n_s.$$

Où $d\left(x_{ri}, x_{sj}\right)$ est la distance entre l'observation i du groupe r et l'observation j du groupe s.

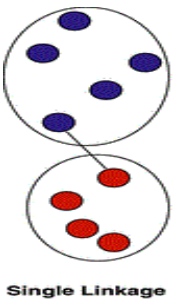

FIG. 1.4 – Saut minumum(src :[19])

Remarque 1.9 *Le groupement à liaison simple peut provoquer un problème qui s'appelle le chaînage. Ceci se produit quand les groupes ne sont pas bien séparés.*

- **Méthode de liaison complète (ou méthode du plus lointain voisin) critère du saut maximal**

La distance entre deux groupes est donnée par la **plus grande** distance entre les objets, où l'un des deux éléments est pris dans le premier groupe et l'autre élément est pris dans le deuxième groupe.

$$d_c(r,s) = \max\left\{d\left(x_{ri}, x_{sj}\right)\right\}; i = 1, \ldots, n_r, j = 1, \ldots, n_s.$$

Remarque 1.10 *La méthode de la liaison complète ne provoque pas de chaînage.*

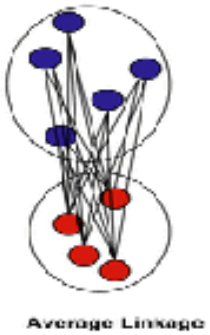

FIG. 1.5 – Saut maximale(src :[19]))

- **Méthode de liaison moyenne**

Cette méthode définie la distance entre les groupes comme :La distance moyenne de tous les éléments d'un groupe à tous les points de l'autre groupe.

$$d_c(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} d\left(x_{ri}, x_{sj}\right).$$

Remarque 1.11 *Cette méthode tend à combiner les groupes qui ont des faibles variances. Elle tend aussi à produire des groupes qui ont approximativement une variance égale.*

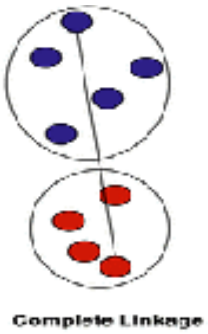

FIG. 1.6 – Moyenne distance(src[19])

• **Méthode de liaison barycentrique (ou à centroïdes)**

Cette méthode requiert les données brutes tout aussi bien que les distances. La distance entre les groupes est la distance entre les centroïdes. Ils sont habituellement les moyennes. Ces moyennes changent à chaque nouvelle fusion.

$$d_c(r,s) = d\left(\overline{x}_r, \overline{x}_s\right).$$

Où $\overline{x}_r$ est la moyenne des observations du groupe r, et $\overline{x}_s$ est la moyenne des observations du groupe s.

Remarque 1.12 *La distance entre les centroïdes est habituellement la distance Euclidienne.*

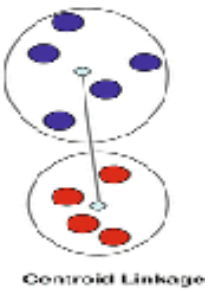

FIG. 1.7 – Centroid distance(src :[19])

- **Méthode de centre de gravité**

Définition 1.9 (Centre de gravité) *C'est le vecteur dont les composantes sont les moyennes arithmétiques des variables* $x_1; x_2, \cdots, x_p$. *On le note par* g. *On l'appelle aussi (**Individu moyenne** ou **poids moyenne**)*

$$g = (\overline{x_1}, \overline{x_2}, \cdots, \overline{x_p})^t \in \mathbb{R}^p.$$

Où :

$$\overline{x_j} = \sum_{i=1}^{n} p_i x_{ij}; \qquad j = \overline{1, p}.$$

Si g_{C_1},g_{C_2} est le centre de gravité de la classe C_1 et C_2 respectivement alors la distance entre les classes C_1 et C_2 est la distance entre les centres des gravités :

$$d(C_1, C_2) = d(g_{C_1}, g_{C_2}).$$

- **Méthode de ward**

Considérons l'évaluation de l'inertie intra-classe au fur et à mesure de la classification. Au début, toutes les classes sont composées d'une unique observation. Chaque classe est donc parfaitement homogène, et l'inertie intra-classe est nulle À la dernière étape de l'algorithme, toutes les observations sont regroupées pour former une unique classe. L'inertie inter-classe est alors nulle, et l'inertie intra-classe est maximum (puisque $I_W + I_B$) .Il n'est par ailleurs pas difficile de constater qu'à chaque étape de la classification, l'inertie intra-classe augmente

alors que l'inertie inter-classe diminue, car chaque étape fusion fait perdre de l'homogénétie aux deux classes fusionnées. Les détails sur ces inerties se trouvent dans la section 2.1.

L'objectif étant d'aboutir à une partition en K classes d'inertie intra-classe minimum, la stratégie va consiste à regrouper à chaque étape les deux classes dont la fusion entraîne le plus faible gain d'inertie intra-classe (ou de manière équivalente la plus faible perte d'inertie inter-classe).

Calculons cette perte d'inertie Soit $g_{C_1C_2}$ le centre de gravité de réunion $(C_{1\cup}C_2)$
On a :

$$g_{C_1C_2} = \frac{p_{C_1}g_{C_1} + p_{C_2}g_{C_2}}{p_{C_1} + p_{C_2}}.$$

Où p_{C_1} et p_{C_2} sont les poids des deux classes.

Soit I_1l'inertie avant le regroupement de C_1 et C_2 et I_2 l'inertie après leur regroupement. La perte d'inertie est égale à :$I_1 - I_2$,telle que : $I_1 = p_{C_1}d^2\left(g_{C_1}, g\right) + p_{C_2}d^2\left(g_{C_2}, g\right)$ et $I_2 = p_{C_1\cup C_2}d^2\left(g_{C_1C_2}, g\right)$, avec $p_{C_1\cup C_2} = p_{C_1} + p_{C_2}$. et g est le centre de gravité de E alors :

$$I_1 - I_2 = p_{C_1}d^2\left(g_{C_1}, g\right) + p_{C_2}d^2\left(g_{C_2}, g\right) - p_{C_1\cup C_2}d^2\left(g_{C_1C_2}, g\right).$$

Définition 1.10 *Un calcul élémentaire montre que cette variation vaut :*

$$\delta\left(C_1, C_2\right) = \frac{p_{C_1}p_{C_2}}{p_{C_1} + p_{C_2}}d^2\left(g_{C_1}, g_{C_2}\right).$$

Proposition 1.1 *Le niveau d'agrégation est égal à la perte d'inertie, c'est à dire*

$$\delta\left(C_1, C_2\right) = I_1 - I_2.$$

Proof. La différence entre I_1 et I_2 est :

$$\begin{aligned}
I_1 - I_2 &= p_{C_1} d^2\left(g_{C_1}, g\right) + p_{C_2} d^2\left(g_{C_2}, g\right) - p_{C_1 \cup C_2} d^2\left(g_{C_1 C_2}, g\right) \\
&= p_{C_1}\left(g_{C_1} - g\right)^2 + p_{C_2}\left(g_{C_2} - g\right)^2 - \left(p_{C_1} + p_{C_2}\right)\left(g_{C_1 C_2} - g\right)^2 \\
&= p_{C_1}\left(g_{C_1}^2 - 2 g g_{C_1} + g^2\right) + p_{C_2}\left(g_{C_2}^2 - 2 g g_{C2} + g^2\right) - \left(p_{C_1} + p_{C_2}\right)\left(\frac{p_{C_1} g_{C_1} + p_{C_2} g_{C_2}}{p_{C_1} + p_{C_2}} - g\right)^2 \\
&= p_{C_1} g_{C_1}^2 - 2 p_{C_1} g g_{C_1} + p_{C_1} g^2 + p_{C_2} g_{C_2}^2 - 2 p_{C2} g g_{C2} + p_{C_2} g^2 \\
&- \left(p_{C_1} + p_{C_2}\right)\left(\frac{p_{C_1} g_{C_1} + p_{C_2} g_{C_2}}{p_{C_1} + p_{C_2}}\right)^2 + 2\left(p_{C_1} + p_{C_2}\right)\frac{p_{C_1} g_{C_1} + p_{C_2} g_{C_2}}{p_{C_1} + p_{C_2}} g - \left(p_{C_1} + p_{C_2}\right) g^2 \\
&= p_{C_1} g_{C_1}^2 - 2 p_{C_1} g g_{C_1} + p_{C_1} g^2 + p_{C_2} g_{C_2}^2 - 2 p_{C2} g g_{C2} + p_{C_2} g^2 \\
&- \frac{\left(p_{C_1} g_{C_1} + p_{C_2} g_{C_2}\right)^2}{p_{C_1} + p_{C_2}} + 2 p_{C_1} g g_{C_1} + 2 p_{C2} g g_{C2} - p_{C_1} g^2 - p_{C_2} g^2 \\
&= p_{C_1} g^2 + p_{C_2} g_{C_2}^2 - \frac{\left(p_{C_1} g_{C_1} + p_{C_2} g_{C_2}\right)^2}{p_{C_1} + p_{C_2}} \\
&= p_{C_1} g^2 + p_{C_2} g_{C_2}^2 - \frac{p_{C_1}^2 g_{C_1}^2 + p_{C_2}^2 g_{C_2}^2 + 2 p_{C_1} p_{C_2} g_{C_1} g_{C_2}}{p_{C_1} + p_{C_2}} \\
&= \frac{\left(p_{C_1} + p_{C_2}\right)\left(p_{C_1} g^2 + p_{C_2} g_{C_2}^2\right) - p_{C_1}^2 g_{C_1}^2 - p_{C_2}^2 g_{C_2}^2 - 2 p_{C_1} p_{C_2} g_{C_1} g_{C_2}}{p_{C_1} + p_{C_2}} \\
&= \frac{p_{C_1} p_{C_2} g_{C_1}^2 + p_{C_1} p_{C_2} g_{C_2}^2 - 2 p_{C_1} p_{C_2} g_{C_1} g_{C_2}}{p_{C_1} + p_{C_2}} \\
&= \frac{p_{C_1} p_{C_2}\left(g_{C_1} - g_{C_2}\right)^2}{p_{C_1} + p_{C_2}} = \frac{p_{C_1} p_{C_2}}{p_{C_1} + p_{C_2}} d^2\left(g_{C_1}, g_{C_2}\right)
\end{aligned}$$

C'est le résultat voulu. ■

Remarque 1.13

1. Si $C_1 = \{a\}$ et $C_2 = \{b\}$,alors :

$$\delta(a, b) = \frac{p_a p_b}{p_a + p_b} d^2(a, b).$$

2. Si $C_1 = \{a, b\}$ et $C_2 = \{c\}$,alors :

$$\delta(ab, c) = \frac{(p_a + p_c)\delta(a, c) + (p_b + p_c)\delta(b, c) - p_c \delta(a, b)}{p_a + p_b + p_c}.$$

C'est la formule de Lance de williams. La méthode de Ward s'applique dans le cas de distance euclidienne.

1.4.4 Algorithme de CAH

Un algorithme de **CAH** est une procédure qui se compose des étapes suivantes :

1. Nous considérons le nuage N_I comme une partition P de I éléments.
2. Grouper les objets dans un arbre hiérarchique binaire : on fusionne les objets qui sont proches en utilisant un critère d'agrégation choisi. Les groupes formés sont fusionnés dans des groupes de plus en plus grands jusqu'à ce que l'arbre hiérarchique soit formé.
3. Déterminer l'endroit où l'arbre hiérarchique est découpé en groupes **:** les éléments de chaque branche en dessous du niveau de découpage donnent un groupe.

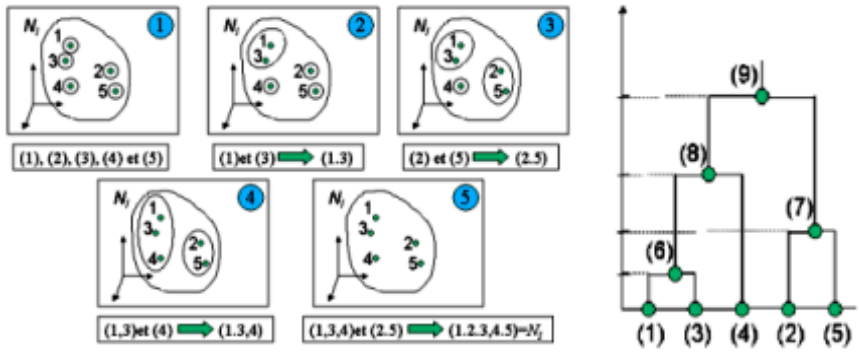

FIG. 1.8 – Algorithme de CAH avec un nuage de I=5 individus(src :[17])

1.4.5 Dendrogramme

L'arbre hiérarchique crée par les méthodes de liaison est plus facilement compris quand il est représenté graphiquement.

Définition 1.11 (Dendrogramme) *Un dendrogramme est un diagramme en arbre qui montre la structure des partitions et comment les groupes sont fusionnés à chaque étape. Le dendrogramme peut être présenté horizontalement ou verticalement. Il y a une valeur numérique*

associée à chaque étape, où les branches se rejoignent (les groupes). Cette valeur représente la distance entre les groupes. L'axe vertical est utilisé pour la représentation de ces valeurs.

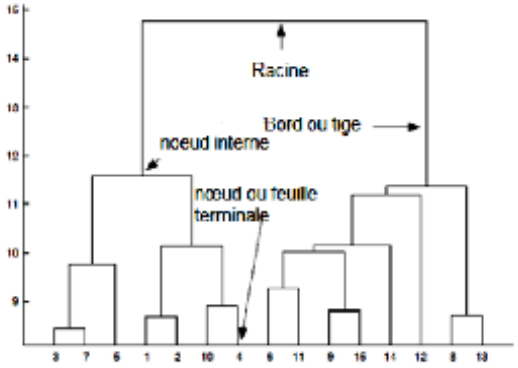

FIG. 1.9 – Dendrogramme

En haut de l'arbre, il y n'y a qu'un seul groupe. C'est toute la population qui est réunie en un seul groupe. En bas de l'arbre, il y a n groupes car chaque individu constitue un groupe. Entre ces deux extrémités, il y a des groupements intermédiaires. La population est partitionnée en un ensemble de groupes. Le nombre de groupes est fixé par le niveau de la hiérarchie en remontant l'axe vertical.

1.4.6 Coupure de dendrogramme

Après avoir créé l'arbre hiérarchique, nous pouvons découper cet arbre pour partitionner les données en groupes. Nous pouvons créer les groupes de deux manières :

– Trouver les divisions naturelles dans les données.

– Spécifier des groupes arbitraires.

Trouver des divisions naturelles dans les données :

L'arbre hiérarchique peut être divisé en groupes bien distincts et bien séparés d'une manière naturelle. Ceci est évident à trouver lorsque dans le dendrogramme les groupes sont compacts dans certaines zones et pas dans d'autres. Le coefficient d'inconsistance peut identifier ces divisions où les similarités entre les objets changent brutalement.

Spécifier des groupes arbitraires :

Au lieu de chercher les divisions naturelles, nous pouvons décider quel est le nombre de groupes que nous voulons obtenir. Ceci se fait avec le niveau dans la hiérarchie.

Plus le niveau est élevé moins il y a de groupes. Si par exemple nous voulons deux groupes, nous plaçons la barre juste en dessous du lien le plus haut, et si on veut obtenir trois groupes, nous plaçons la barre en dessous du deuxième lien.comme indiqué dans la fig.1.10.

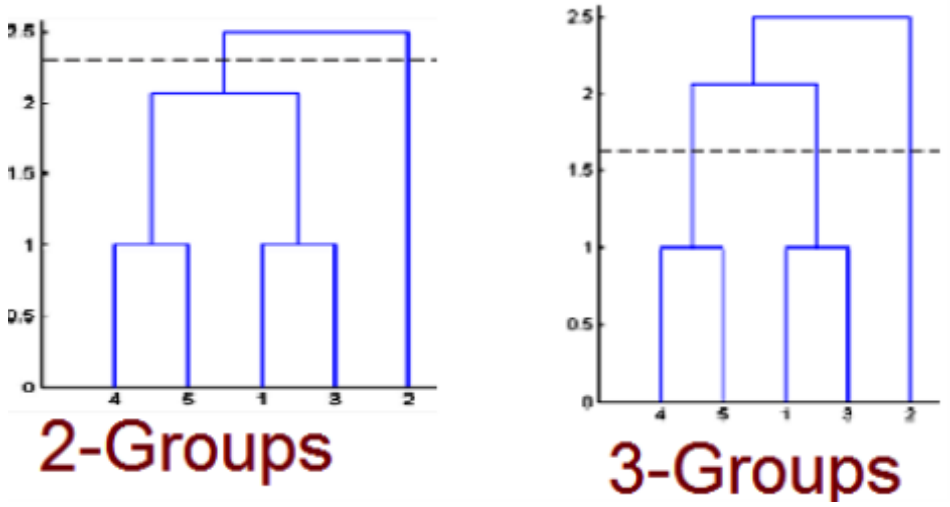

FIG. 1.10 – Coupure en deux et trois groupes.

1.4.7 Avantages et inconvénients de CAH

Avantages

* Pas de dépendance au choix de centres initiaux.
* Pas de fixation à priori du nombre de classes.
* Détecte des classes de forme diverse ou des centre de classes.

Inconvénients

* Complexité algorithmique(non linéaire).
* A chaque étape, le critère de partionnement n'est pas globale mais dépend des classes déjà obtenues.

1.5 Méthode de classification descendante hiérarchique (CDH)

Une classification hiérarchique descendante (divisive) réalisé un dendrogramme non pas par agrégation ,mais par division depuis, tous les éléments formant une classe jusqu'à la partition formée de tous les éléments simple.

Les méthodes de **CDH** déterminent ,à chaque étape, le groupe courant le moins homogène et le splittent en deux sous groupes

Leur schéma générale est le suivant :

⋆ Ressembler tous les objets dans un même cluster. Définir une valeur seuil de distance (ou de dissimilarité).

⋆ Comparer tous les objets deux à deux dans chaque cluster et marquer la paire d'objets ayant la plus grand distance.

⋆ Si cette distances est supérieure à la valeur seuil, splitter le cluster correspondant en deux et retourner au étape 2.Sinon fin de la procédure.

1.5.1 Avantages et inconvénients de la CDH

Avantages : Par rapport à la plus part des algorithmes en classification automatique, l'algorithme de **CDH** ne nécessite pas l'utilisation d'une seuil arbitraire pour la formation des classes qui peut éventuellement mener la recherche d'une partition dans une direction non réaliste.

Inconvénients

Les resulats sont en générale grossiers, les niveaux des nœuds de la hiérarchie ne sont plus définis que par l'ordre dans lequel ils apparaissent.

1.6 Exemple d'application

La classification hiérarchique donne des partitions emboîtées. A la base, chaque observation constitue un groupe. Puis les groupes vont s'agglomérer deux à deux pour donner des groupes plus grands. A la fin, nous obtenons un seul groupe qui est constitué de toutes les observations ensemble. Lorsque nous exécutons la méthode de la classification hiérarchique nous obtenons toutes les partitions possibles. Mais parmi ces partitions quelle est la meilleure? C'est celle qui fait que les éléments qui sont dans un groupe sont très proches les uns des autres et en même temps sont loin des autres groupes.

Ou l'objectif d'ou cet exemple est d'appliqué les résultats théoriques ; les données prises à partir du site web [[25]] : on utilise différentes fonctions sur logiciel R à l'aide des package (fpc,stats...) pour réaliser les résultats numériques et graphiques.

1.6.1 Présentation des données

On prend quinze pays arabes et on va étudier leur coût de vie en , de L'objectif est d'identifier des groupes homogènes, partageant des caractéristiques similaires.Les données dans le tableau **??** qui croise les quinze pays :{Cout de vie (C.V), Culture (Cult), Economie (Eco), Libertés (Liber), Santé, Infrastructure (I.S), Sécurité (Séc), Climat (Clim)} nommes individus et leurs coût de vie nomées variables.

Pays	(C.V)	(Cult)	(Eco)	(Liber)	(Santé)	(I.S)	(Séc)	(Clim)
Tunis	63	61	45	17	73	36	86	85
Jordan	56	60	45	33	80	28	71	68
Koweit	46	60	74	50	70	56	71	18
Liban	68	56	50	42	85	36	21	61
Maroc	43	35	44	42	63	36	93	78
Bahrein	59	67	61	33	43	22	86	30
Syrie	69	49	44	8	68	44	71	60
Qatar	54	63	92	25	47	24	79	13
Egypt	55	49	47	25	76	36	79	32
UEA	49	53	76	25	76	58	29	10
Algeria	51	51	44	25	51	36	36	89
Oman	50	44	53	25	56	44	64	2
KSA	60	68	53	8	71	28	36	22
Iraq	100	49	45	17	55	36	0	38
Soudan	53	19	44	2	28	40	21	7

TAB. 1.1 – Les pays arabes et leurs cout de vie

Croisement des objets

Lorsque nous faisons des représentations graphiques des points de cet ensemble par croisement deux à deux variables, nous obtenons les résultats dans le figure 1.11.

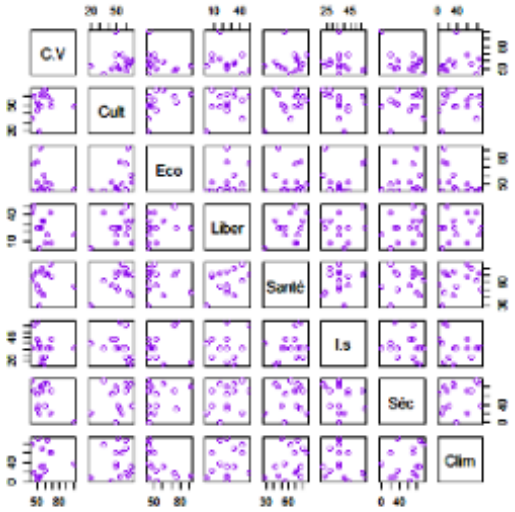

FIG. 1.11 – Croissement deux à deux(src : [12])

CAH

La classification ascendante hiérarchique applique plusieurs stratégies d'agrégation pour obtenir une arbre de classification. on va utiliser sur notre données les stratégies : de Saut maximal (Complete) comme la montre le figure 1.12.

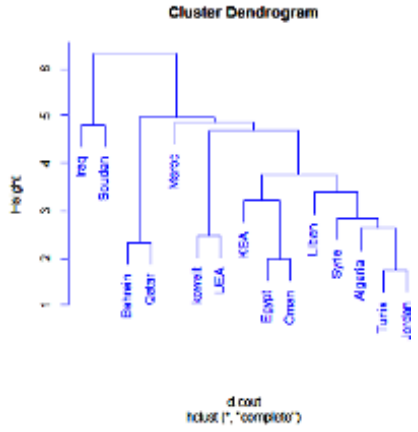

FIG. 1.12 – Dendrogramme de pays(src :[12])

Coupure de dendrogramme

Après d'avoir la création de l'arbre hiérarchique des pays, nous pouvons découper cet arbre pour partitionner les données en groupes que la montre la figure 1.13.

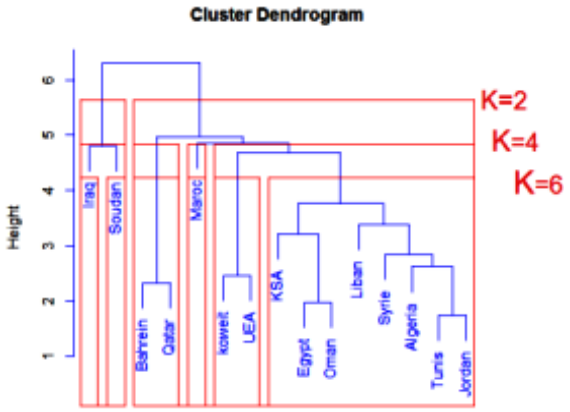

FIG. 1.13 – Coupure en 2,4 et 6 classes(src :[12])

Le dendrogramme propose de se diviser en quatre groupes. Ou plus, peut également être coupé en deux groupes seulement.

Chapitre 2

Classification non Hiérarchique

La méthodes non hiérarchique font partie des méthodes exploratoires, génère une classification en partionenent des données, ce qui donne généralement un ensemble des groupes ne sont chevauchant pas ,sans relation hiérarchique entre eux. Elle connue aussi comme **méthode de partitionnement.** Son avantages sont principalement ;sa simplicité et sa rapidité d'exécution même sur un grand nombre d'observations ; ou contraire, nous pourrons lui reprocher une classification finale un peu trop approximative et le besoin de spécifier à priori un nombre de classes à obtenir en fin de procédure.

Une méthode de classification non hiérarchique permettent de subdiviser l'ensemble des individus en un certain nombre de classes en employant une stratégie d'optimisation itérative dont le principe général est de générer une partition initiale, puis de chercher à l'améliorer en réattribuant les données d'une classe à l'autre. Il n'est bien entendu pas souhaitable d'énumérer toutes les partitions possibles. Ces algorithmes recherchent donc des maxima locaux en optimisant une fonction objectif traduisant le fait que les individus doivent être similaires au sein d'une même classe, et dissimilaires d'une classe à une autre. Les classes de partition final, prises deux à deux, sont d'intersection vide est représentée par noyau.

Les algorithmes de partitionnement sont divisés en trois grandes sous-familles :**Centre mobile(Forgy 1965)** ;**nuées dynamiques(Diday et al 1980)** ;**K-means(MacQueen,1967)**.

2.1 Critères de homogénéité

Le critère mesurant l'homogénéité des classes c'est à dire la qualité de la partition est souvent l'inertie intra classe de la partition et l'algorithme itératif utilisé pour minimiser localement.

2.1.1 Inertie inter-classe et intra -classe

On considère un ensemble $\Omega = \{e_1, \ldots, e_n\}$ de $n \geq 1$ individus et on désigne par :

g le centre de gravité du ces individus :

$$g = \sum_{i=1}^{n} p_i e_i,$$

tel que p_i est le poids de l'individu e_i.où $p_i > 0$ et $\sum_{i=1}^{n} p_i = 1$

μ_k : le poids de classe C_k :

$$\mu_k = \sum_{i \epsilon C_K} p_i; k = 1, \cdots, K.$$

g_k : le centre de gravité de C_k :

$$g_k = \frac{1}{\mu_k} \sum_{i \epsilon C_k} p_i e_i.$$

Et

$$d_M^2 (e_i, e_i) =^t (e_i - e_i) M (e_i - e_i),$$

où M est une métrique.

Définition 2.1 (Inertie Totale) *L'inertie totale I_T du nuage des n individus est :*

$$I_T = \sum_{i=1}^{n} p_i d_M^2 (e_i, g).$$

Où g est le centre de gravité du nuage.

Remarque 2.1 *L'inertie totale est indépendante de la partition.*

Définition 2.2 *L'inertie I_a du nuage des n individus par rapport à un point $a \in \mathbb{R}^p$ est :*

$$I_a = \sum_{i=1}^{n} p_i d_M^2 (e_i, a) .$$

Définition 2.3 (Inter-classe) *L'inertie inter-classe I_{Inter} de la partition P est :*

$$I_{Inter} = \sum_{k=1}^{K} \mu_k d_M^2 (g_k, g) .$$

Définition 2.4 (Intra-classe) *L'inertie intra-classe I_{Intra} de la partition P est :*

$$I_{Intra} = \sum_{k=1}^{K} I (C_k)$$

Où

$$I (C_k) = \sum_{i \epsilon C_k} p_i d_M^2 (e_i, g_k) .$$

Proposition 2.1 (Théoréme de Huygens) *Soit I_a l'inertie d'un point a est I_g l'inertie totale alors :*

$$\forall a \in \mathbb{R}^p, I_a = I_g + \|g - a\|_M^2 = I_g + \sum_{i=1}^{n} p_i d_M^2 (g, a)$$

Proof. On a l'inertie I_a d'un point a par définition 2.2 est :

$$\begin{aligned} I_a &= \sum_{i=1}^{n} p_i d_M^2 (e_i, a) = \sum_{i=1}^{n} p_i \parallel e_i - a \parallel_M^2 \\ &= \sum_{i=1}^{n} p_i \parallel e_i - g + g - a \parallel_M^2 \end{aligned}$$

$$= \sum_{i=1}^{n} p_i \parallel e_i - g \parallel_M^2 + 2\sum_{i=1}^{n} p_i \, {}^t(g-a)M \, (e_i - g)$$

$$+ \quad \sum_{i=1}^{n} p_i \parallel g - a \parallel_M^2 \text{ (car } M \text{ est symétrique)}$$

$$= I_g + \sum_{i=1}^{n} p_i d_M^2 (a, g) + 2^t(g-a)M \sum_{i=1}^{n} p_i \, (e_i - g)$$

$$\text{Or} \quad \sum_{i=1}^{n} p_i (e_i - g) = 0 \text{ alors}$$

$$I_a = I_g + \sum_{i=1}^{n} p_i d_M^2 (a, g)$$

d'où le resultat ■

Proposition 2.2 *On a la relation fondamentale suivante* :

$$I_T = I_{Intra} + I_{Inter}.$$

Donc on en déduit que minimiser l'inertie intra-classe c'est à dire l'homogénéité des classes est équivalent à maximiser l'inertie inter-classe, c'est à dire la séparation entre les classes. Cette relation se déduit du ***théorème de Huygens.***

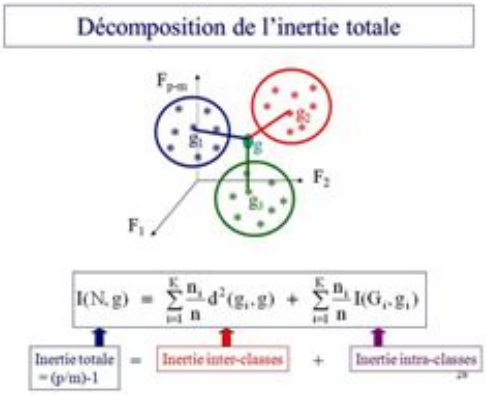

FIG. 2.1 – Décomposition-de-l'inertie-totale(src :[8])

Proof. On considére g_k le centre de gravité de la classe C_k.D'aprés le théoréme de Huygens,l'inertie du nuage des points de C_k par rapport au centre de gravité g s'écrit :

$$\begin{aligned} I_T &= \sum_{i=1}^{n} p_i d_M^2(e_i, g) = \sum_{i=1}^{n} p_i \parallel e_i - g \parallel_M^2 \\ &= \sum_{k=1}^{K} \sum_{i \epsilon C_k} p_i \parallel e_i - g_k + g_k - g \parallel_M^2 \\ &= \sum_{k=1}^{K} \sum_{i \epsilon C_k} p_i \left[\|e_i - g_k\|_M^2 + \|g_k - g\|_M^2 \right] \text{ (d'aprés théoréme de Huygens)} \\ &= \sum_{k=1}^{K} \sum_{i \epsilon C_k} p_i \|e_i - g_k\|_M^2 + \sum_{k=1}^{K} \sum_{i \epsilon C_k} p_i \|g_k - g\|_M^2 \\ &= \sum_{k=1}^{K} \sum_{i \epsilon C_k} p_i d_M^2(e_i, g_k) + \sum_{k=1}^{K} \sum_{i \epsilon C_k} p_i d_M^2(g_k, g) \end{aligned}$$

Alors $I_T = I_{Intra} + I_{Inter}$

d'où le resultat. ■

Proposition 2.3 *L'inertie d'une classe C_k s'écrit également indépendamment du centre de gravité ,en ne faisant intervenir que les distances des individus deux à deux :*

$$I(C_k) = \sum_{i \epsilon C_k} p_i d_M^2(e_i, g_k) = \sum_{i \epsilon C_k} \sum_{\grave{\imath} \epsilon C_k} \frac{p_i p_{\grave{\imath}}}{2\mu_k} d_M^2(e_i, e_{\grave{\imath}}).$$

Cette égalité se déduit du théoréme de Huygens et de la définition de l'inertie par rapport à un point.

Proof. On considère g_k le centre de gravité de la classe C_k.D'aprés le théorème de Huygens, l'inertie du nuage des points de C_k par rapport à un point e_i s' écrit :

$$\begin{aligned} I_{e_i} &= I_{g_k} + \parallel g_k - e_i \parallel_M^2 \\ &= I_{g_k} + \overbrace{\sum_{i \epsilon C_k} {}_i p_i}^{\mu_k} d_M^2(e_i, g_k) \end{aligned}$$

En multipliant cette égalité par p_i et en sommant pour i dans C_k ont trouve :

$$\sum_{i \epsilon C_k} p_i I_{e_i} = \overbrace{\sum_{i \epsilon C_k} p_i}^{\mu_k} I_{g_k} + \mu_k \overbrace{\sum_{i \epsilon C_k} p_i d_M^2(\omega_i, g_k)}^{I_{g_k}}$$

En remplaçant I_{e_i}par sa valeur dans cette égalité on trouve :

$$\sum_{i \epsilon C_k} p_i \left(\sum_{\grave{\imath} \epsilon C_k} p_{\grave{\imath}} d_M^2(e_i, e_{\grave{\imath}}) \right) = 2\mu_k I_{g_k}$$

et comme $I_{g_k} = I(C_k)$ on retrouve l'égalité :

$$I_{g_k} = I(C_k) = \frac{\sum_{i \epsilon C_k} \sum_{\grave{\imath} \epsilon C_k} p_i p_{\grave{\imath}}}{2\mu_k} d_M^2 \left(e_i, e_{\grave{\imath}}\right).$$

d'où le resultat. ■

2.2 Différentes algorithmes.

2.2.1 Méthode de centre mobile (Forgy 1965)

La méthode des centres mobiles due à Forgy [Forgy, 1965][9] est la plus classique et celle qui reste très utilisée. Cette méthode s'applique lorsque l'on sait à l'avance combien de classes on veut obtenir. Appelons K ce nombre.

Principe

On représente les individus comme des points de l'espace ayant pour coordonnées des mesures ($\geq$ dim 2.) On cherche à regrouper autant que possible les individus les plus semblables tout en séparant au mieux les classes . On choisit de procéder de façon automatique, uniquement à partir des mesures, des ressemblances et des différences.

Algorithme de cette méthode

L'algorithme de méthode des centres mobiles procède comme suit :

Etape0 :Pour initialiser l'algorithme, on tire au hasard k individus appartenant à la population, $C_1(0)$, $C_2(0)$,..., $C_k(0)$:ce sont les k centres initiaux.

Etape1 :**(Constitution de classes)** On regroupe les individus autours de ces k centres de sorte à former k classes $\Gamma_1(0), \Gamma_2(0), \cdots, \Gamma_k(0)$ de la maniére suivante : chaque classe $\Gamma_l(0)$ est constituée des points plus proches du centre $C_l(0)$ que des autres centres $\Gamma_m(0)$ pour $m \neq l$..

Etape2 :(Calcul des nouveaux centres) On calcule alors les centres de gravité G_1,G_2 ,$\cdots$, G_k des k classes obtenues et on désigne ces points comme nouveaux centres $C_1(0) = G_1$, $C_2(0) = G_2$,$\cdots$, $C_k(0) = G_k$.

Etape3 :(Répétition des étapes 1et 2) On répète les étapes 1et 2 jusqu'à ce que le découpage en classes obtenu ne soit presque plus modifié par une itération suplémentaire. On peut montrer que la variance intra-classe ne peut que décroître lorsque l'on passe d'un découpage en classes au suivant comme le montre la figure 2.2.

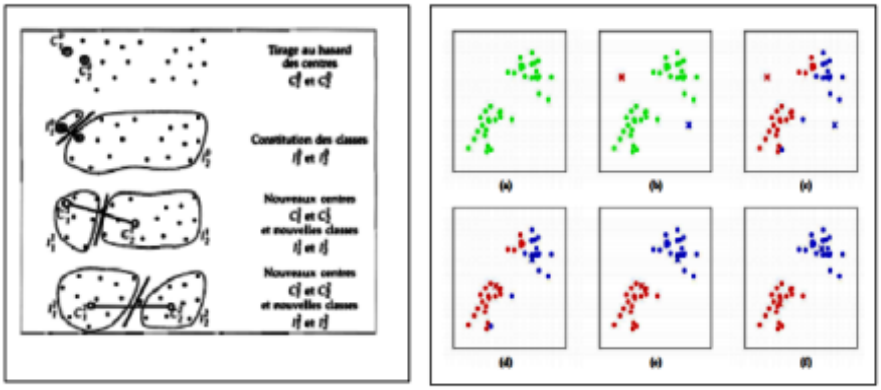

FIG. 2.2 – Algorithme des centres mobiles(src :[5])

Avantages et inconvénients

Avantages :

⋆ Temps d'exécution proportionnel au nombre d'individus ce qui la rend applicable à de grands volumes de données.

⋆ Nombre d'itérations nécessaires est faible.

Inconvénients

⋆ Ne s'applique qu'à des données continues ce qui nécessite des transformations.

⋆ Absence de solutions optimales mais des meilleurs solutions possibles par rapport aux hypothèse d'origine.

⋆ Le nombre de segments est fixé au départ. Il ya donc un risque qu'on s'éloigne du véritable nuage des individus.

2.2.2 Méthode de K-means (MacQueen,1967)

La méthode de K-means a été utilisé par MacQueen.James en (1967) [**?**],l'algorithme K-means est l'un des plus connus et simples algorithmes de classification automatique des données. Dans cet algorithme, les classes sont représentées par leurs centres, qui correspond à la moyenne de l'ensemble des objets contenus dans la classe. La procédure suit un moyen simple et facile de classer un ensemble de données donné à travers un certain nombre de classes K fixées.

Principe

Le principe de la méthode des "K-means" c'est que la classification se fait sur la base du critère des plus proches voisins. Celui-ci signifie que chaque individu est affecté à une classe s'il est très proche de son centre de gravité.

Algorithme

Dans la méthode des "K-means", le choix des centres initiaux s'effectue sur la base d'un tirage aléatoire sans remise de K. individus à partir de la population à classifier. La partition des classes est modifiée avec chaque affectation d'un individu i.
L'algorithme de la méthode des "K-means" se déroule comme suit :

Etape1 : Choisir K individus au hasard (comme centre des classes initiales)

Etape2 : Affecter chaque individu au centre le plus proche.

Etape3 : Recalculer le centre de chacune de ces classes.

Etape4 : Répéter l'étape 2 et 3 jusqu'à stabilité des centres.

Etape5 : Editer la partition obtenue.

Avantages et inconvénients

Avantages

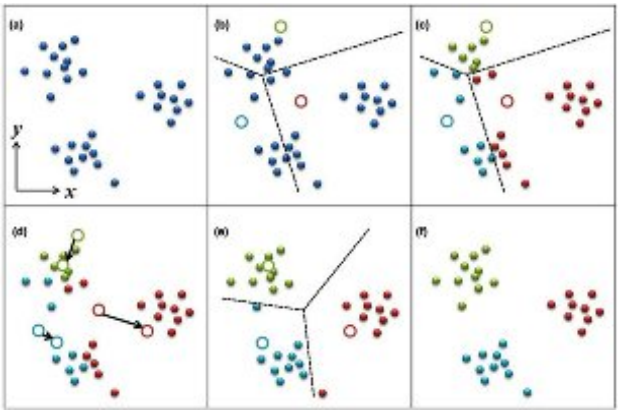

FIG. 2.3 – K-means algorithme pour trois dimension ($K = 3$)(src :[24])

⋆ **Simple :** Il est facile d'implémenter K-means et d'identifier des groupes de données inconnus à partir d'ensembles de données complexes. Les résultats sont présentés de manière rapide.

⋆ **Flexible :** L'algorithme K-means s'adapte aux divers changements de vos données. En cas de souci, l'ajustement du segment de cluster permettra d'apporter rapidement des modifications nécessaires à l'algorithme.

⋆ Convient aux gros data sets : K-means convient à un grand nombre d'ensembles de données et est calculé beaucoup plus rapidement que le plus petit. Il peut également produire des clusters plus élevées.

⋆ **Efficace :** L'algorithme utilisé permet de partitionner les gros de datasets. Son efficacité est fonction de la forme des clusters. Les K-Means fonctionnent bien dans les clusters hyper-sphériques.

⋆ **Facile à interpréter** : Les résultats sont très faciles à interpréter. K-Means génère des descriptions de cluster sous une forme minimisée pour maximiser la compréhension des données.

Inconvénients

⋆ **Spécifiez les valeurs** K : Pour que la classification par K-moyennes (K-means) soit efficace, vous devez spécifier le nombre de clusters (K) au début de l'algorithme.

⋆ **Problèmes de prédiction :** Il est difficile de prévoir les valeurs K ou le nombre de

clusters . Il est également difficile de comparer la qualité des clusters produites.

⋆ **Traiter les données numériques** : l'algorithme K-moyennes ne peut être exécuté que dans des données numériques.

2.2.3 Méthode de Nuées dynamiques(Diday 1980)

La méthode de classification des nuées dynamiques (Diday 1980)[**?**, **?**, **?**, **?**] repose essentiellement sur la répartition d'une population en catégories (classes) tout en utilisant la notion de noyau associé a chaque classe.

Principe

Considérer un ensemble d'individus qui appartient a un ensemble E, et chercher la meilleure partition à K classes fixées de cet ensemble selon le critère d'inertie. Le processus est itératif et à chaque étape la qualité de la partition s'améliore. Le nombre de classe souhaité est déterminé à priori ainsi que le nombre d'éléments centraux désirés, c'est-à-dire le nombre d'éléments au centre du noyau qui seront énumérés. Au départ, un ensemble de points ou noyaux d'une classe peut être tiré au hasard. Autour de ces points se regroupent les éléments les plus proches pour former une partition. La distance calculée par rapport au centre de classe est la distance euclidienne. A partir de cette partition créée, une autre famille de noyaux est définie, elle regroupe les points les plus proches formant une nouvelle classe et ainsi de suite jusqu'à obtention d'un nombre fini de classes. Si, après un certain nombre d'itérations, les classes formées sont stables, les données sont dites "classifiables" et constituent des "formes fortes". Les individus qui changent de classes selon les tirages sont les "individus charnières".

Algorithme

Etape1 : On part de K centres.

Etape2 : Ces centres déterminent une partition.

Etape3 : On remplace les centres par les centres de gravité de chaque sous ensemble.

Etape4 : On recommence en l'étape 2.

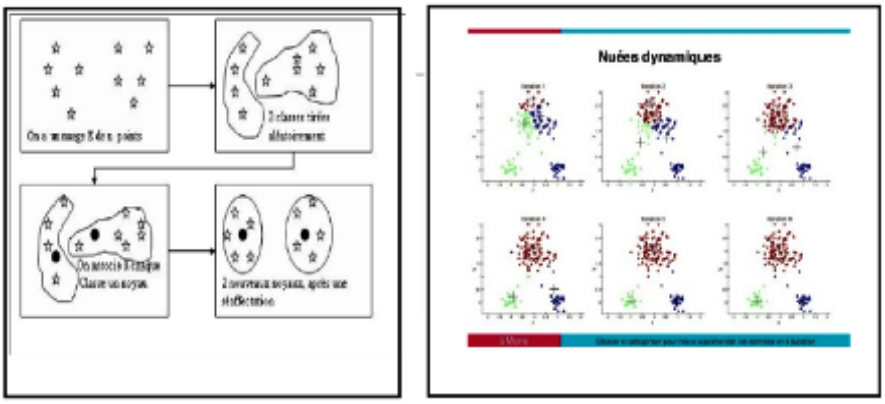

FIG. 2.4 – Nuées-dynamiques-par-k=2et 3(src :[3])

Avantages et inconvénients

Avantages

⋆ Complexité linéaire.

⋆ Applicable à de grands volumes de données.

⋆ Permet de détecter les individus isolés ou hors norme.

⋆ Amélioration continue de la qualité des classes.

Inconvénients

⋆ Partition finale dépend des choix initiaux : pas d'optimum global.

⋆ Nombre de classes K fixé.

⋆ Ne détectent bien que les formes sphériques.

2.3 Exemple d'application

K-means, à la différence de la CAH, ne fournit pas d'outil d'aide à la détection du nombre de classes. Nous devons les programmer sous R ou utiliser des procédures proposées par des packages dédiés. Le schéma est souvent le même : on fait varier le nombre de groupes et on surveille l'évolution d'un indicateur de qualité de la solution c.-à-d. l'aptitude des individus à être plus proches de ses congénères du même groupe que des individus des autres groupes.

2.3.1 Présentation des données

Pour la présentation des résultats, pour commencer, les résultats trouvés peuvent être affichés directement, la présentation des résultats pourra se faire par des graphiques en affichant par exemple les données dans l'espace, chaque ensemble de données appartenant à une même classe est colorée par une couleur différente, etc. Comme l'exemple de la figure 2.5.

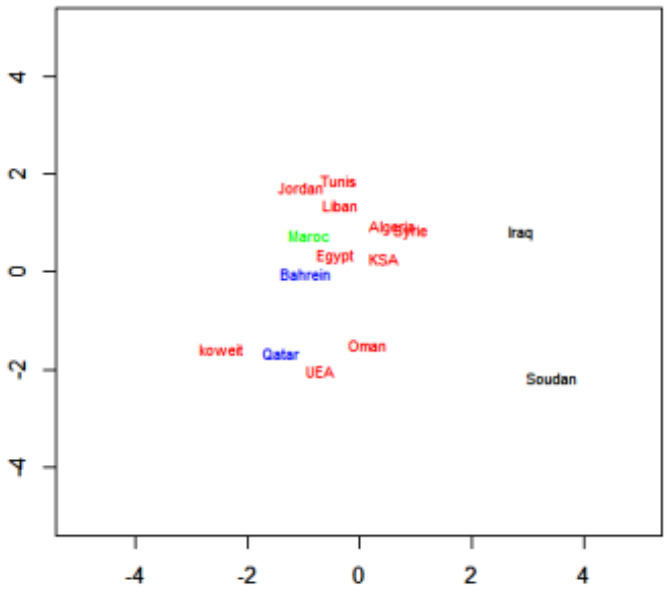

FIG. 2.5 – Classification de pays(src :[12])

2.3.2 Interprétation des classes

Dans la méthode des $K-$means nous devons choisir le nombre de groupes au départ. Nous allons par la suite vérifier si ce nombre de groupes est bon ou non. Le critère qui nous permet de voir cela est la valeur de **silhouette** sur chaque partition choisie. Ce critère nous permet de voir si le nombre de groupes est bien choisie et si les éléments de chaque groupe sont proches les uns des autres et sont éloignés des éléments des autres groupes. Dans la figure 2.6 on utilisé le même exemple de la première chapitre.

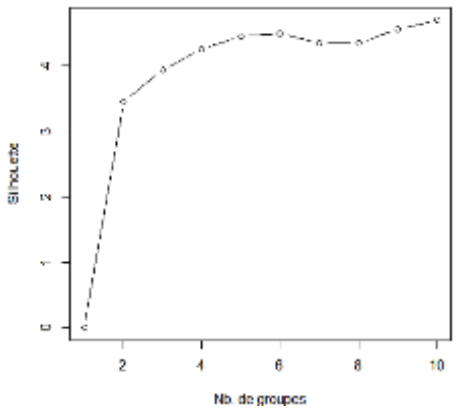

FIG. 2.6 – Nombre de groupes(src :[12])

Dans ce graphe on utiliser l'indice de Calinski Harabasz, on recherche alors à maximiser critère on peut aussi choisir l'indice silhouette moyenne.

Conclusion

Dans ce travail ; on a essayé de présenté deux méthodes d'analyse statistique exploratoires mutidimentionnel : classification hiérarchique et non hiérarchique ; la classification hiérarchique est l'une des grandes types de classification automatique avec la classification non-hiérarchique La différence entre les deux que le première type permet de voir toute une gamme pour le nombre de groupes et de choisir celui qui satisfait le mieux. mais le deuxième type partent déjà d'un nombre fixé de groupes et focalisent sur la répartition optimale., l'objectif de la classification est d'obtenir une représentation schématique simple (dendrogramme) de tableau initiale à partir d'une partition des individus dans des objets homogènes.

Bibliographie

[1] Benzecri, J.P. (1973). L'analyse des données. T1 : La taxinomie 619.Dunod, Paris. .

[2] Boubou.M.(2007).Contribution aux méthodes de classification non supervisées via des approches prétopologiques et d'agrégation d'opinion.Doctoral dissertation, Université Claude Bernard-Lyon I.

[3] CHERAD.A et SID.N. (2009). Optimisation du réseau du gaz lift dans la partie nord du champ de Hassi Messaoud. .Univesité des sciences et de la technologie Houari Boumedienne.

[4] Cloud, AI.(2017) .machine learning-Solution or Problem.

[5] Coquillard.P. (2018) Méthodes de classification..Master SV. Cours VIII.

[6] Crucian.M.(2018) .Classification automatique.Conservatoire National des Arts et Métiers,Paris,France.

[7] Diday, E, Lemaire.J, Pouget.J., and Testu.F.(1982). Elements d'analyse de données. Dunod, Paris.

[8] Folley.J. (2015) Agglomerative Hierarchical Clustering.

[9] Forgy.E.W.,(1965).Cluster analysis of multivariate data : efficiency versus interpretability of classifications,.Biometrics. (768, 769) .

[10] Hartigan J. A., Wong.M.A,(1979) K-Means Clustering Algorithm , Journal of the Royal Statistical Society, Series C, vol. 28, no 1, $100 - 108$.

[11] Jumbu.M.(1989).Exploration informatique et statistique des données.528.

[12] Ihaka, R., Gentleman, R. (1996) *R : A Language for Data Analysis and Graphics*. Journal of Computational and Graphical Statistics **5** : $299 - 314$.

[13] Lerman I.C, (1981), Classification et analyse ordinale des données, Dunod, Paris.

[14] Lloyd.S. P.(1982) .Least square quantization in PCM ,Bell Telephone Laboratories. Journal much later.129 − 137.

[15] Mac-Queen.J.B.(1967). Some Methods for classification and Analysis of Multivariate Observations.Proceedings 281 − 297.

[16] Malik.U. (2018) Hierarchical Clustering

[17] Martin.A.(2004) .l'analyse des données Polycopié de cours ENSIETA.

[18] Monbel V (2006).Clustering.Université de Rennes 1.

[19] Peng.L and Yaqing S.(2014). Cluster Analysis of RNA-Sequencing Data.

[20] Rangeon.N.(2019) Réalisez une analyse exploratoire de données.

[21] Roux G. (1967). Propos de quelques methodes de Classification en phytosociologie.Rev. de stat.. Vol. XV,59 − 72.

[22] Steinhaus.H. (1957). Sur la division des corps matériels en parties. Bull. Acad. Polon. Sci., vol. 4, no 12, 801 − 804.

[23] Tabet A, Houcine.W, et Ziani S. (2012) Clustering Hiérarchique de données à base de Ward.Université Abou Bakr Belkaid. Tlemcen.

[24] Yu-Zhong Chen (2016) Universal structural estimator and dynamics approximator for complex networks.Chicago, United States.

[25] https ://alphabeta.argaam.com/article/detail/9277.

Annexe A : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous.

src	Source.
Ω	Ensemble des individus.
n, p	Le nombre des individus et variables respectivement.
$\mathbb{X}$	Tableau des données.
$\mathbb{R}^n$	Espace des nombres réels de dimension n.
$\mathbb{R}^p$	Espace des nombres réels de dimension p.
S	Indice de similarité.
Dis	Indice de dissimilarité.
d	Distance.
M	Métrique.
I	Métrique usuelle.
D_{1/S^2}	Matrice diagonale des inverses des variances des p variables.
V^{-1}	L'inverse de matrice de covariance.
P	Partition.
C_i	Classe
H	Hiérarchie.
$C.A.H$	Classification Ascendante Hiérarchique
D	Matrice des distances.
n_r, n_s	Le nombre des objets de groupe r, s respectivement.

$d_c(r,s)$	La distances entre r et s.
$\overline{x}_r, \overline{x}_s$	Moyenne de groupes r et s.
g	Centre de gravité.
p_i	Le poid d'individu e_i.
g_{C_1}, g_{C_2}	Le centre de gravité de classe C_1et C_2.
N_I	Nuage des individus.
$C.D.H$	Classification Descendante Hiérarchique.
μ_k	Le poids de classe C_k.
g_k	Le centre de gravité de classe C_k.
I_{intra}	Inertie intra-classe.
I_{inter}	Inertie inter-classe.
K	Le nombre des groupes.