La classification hiérarchique :

1. Introduction sur les méthodes hiérarchiques

Il existe de nombreuses techniques visant à partitionner une population en différentes classes ou sous-groupes. Le clustering hiérarchique est l'une d'entre elles. Une méthode hiérarchique crée une décomposition hiérarchique d'un ensemble d'objets. Elle peut être classée comme étant soit agglomérative, soit divisionnaire, en fonction de la manière dont la décomposition hiérarchique est formée :

1.1. Méthode hiérarchique agglomérative

Une méthode de clustering hiérarchique agglomérative utilise une stratégie ascendante. Elle commence généralement en laissant chaque objet former son propre cluster et fusionne de manière itérative les clusters en clusters de plus en plus grands, jusqu'à ce que tous les objets soient dans un seul cluster ou que certaines conditions de terminaison soient satisfaites. Le cluster unique devient la racine de la hiérarchie. Pour l'étape de fusion, il trouve les deux clusters les plus proches l'un de l'autre selon une mesure de similarité et combine les deux pour former un cluster.

1.2. Méthode hiérarchique divisionnaire

L'approche par division, également appelée approche descendante, commence par placer tous les objets dans un cluster, qui est la racine de la hiérarchie. Il divise ensuite le cluster racine en plusieurs sous-clusters plus petits et partitionne récursivement ces clusters en plus petits. Le processus de partitionnement se poursuit jusqu'à ce que chaque cluster au niveau le plus bas soit suffisamment cohérent, soit qu'il ne contienne qu'un seul objet, soit que les objets d'un cluster soient suffisamment similaires les uns aux autres.

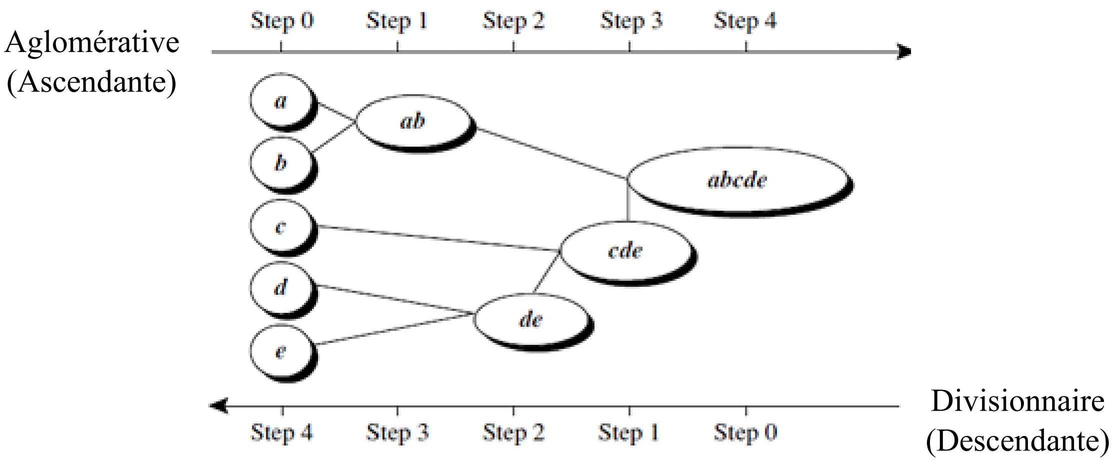


Figure 1. Regroupement hiérarchique agglomératif et divisionnaire

2. La Classification Ascendante Hiérarchique (CAH)

La Classification Ascendante Hiérarchique (CAH) est une des méthodes hiérarchiques. Dans CAH, on cherche à ce que les individus regroupés au sein d'une même classe (homogénéité intra-classe) soient le plus semblables possibles tandis que les classes soient le plus dissemblables (hétérogénéité inter-classe).

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 existante entre les individus 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.

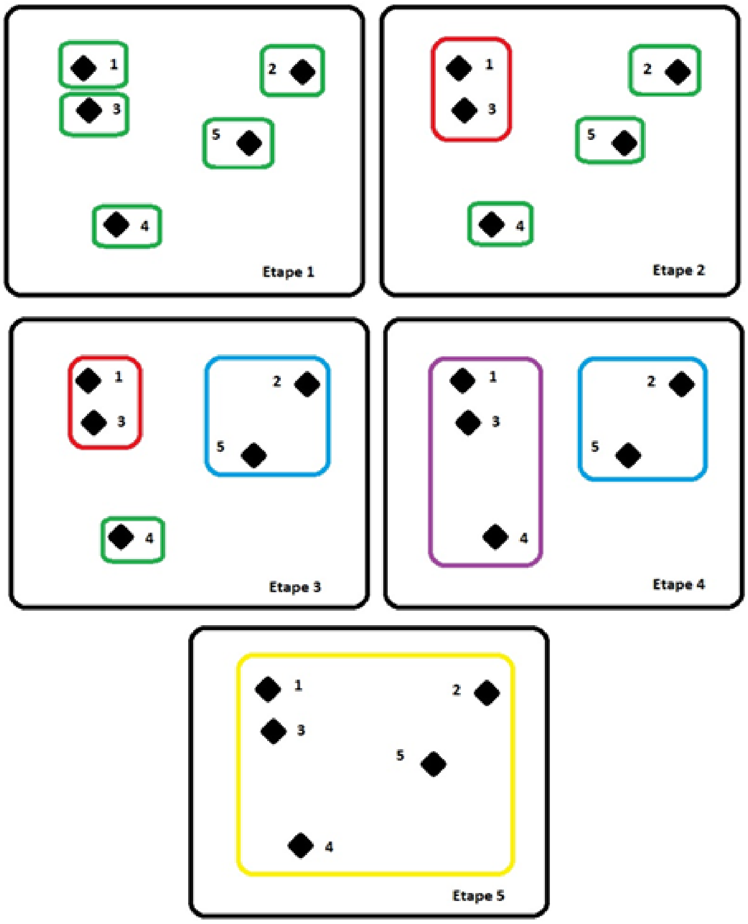


Figure 2. Exemple graphique de CAH.

3. Le calcul de la matrice des distances

La notion de ressemblance entre observations est évaluée par une distance entre individus.

Il existe de nombreuses distances mathématiques pour les variables quantitatives :

- Distance Euclidienne : $D = [\sum (val_1 - val_2)^2]^{1/2}$
- Distance de Manhattan : $D = \sum |val_1 - val_2|$
- Distance de Hamming : $D = \sum d_i$ avec $d_i = \begin{cases} 0, & X_i = \grave{X}_i \\ 1, & X_i \neq \grave{X}_i \end{cases}$

4. La construction de dendrogramme

Un dendrogramme est la représentation graphique d'une classification ascendante hiérarchique. Il se présente souvent comme un arbre binaire dont les feuilles sont les individus alignés sur l'axe des abscisses. Lorsque deux classes ou deux individus se rejoignent avec l'indice d'agrégation, des traits verticaux sont dessinés de l'abscisse des deux classes jusqu'à l'ordonnée, puis ils sont reliés par un segment horizontal.

Il faut choisir une méthode d'*agrégation* pour construire le dendrogramme. De nombreuses solutions existent (saut minimum, distance maximum, moyenne, Ward…). Chacune d'elle produira un dendrogramme différent. Cependant, à l'usage, on privilégiera le plus souvent la *méthode de Ward.* De manière simplifiée, cette méthode cherche à minimiser l'inertie intra-classe et à maximiser l'inertie inter-classe afin d'obtenir des classes les plus homogènes possibles.

Pour obtenir une *partition* de la population, il suffit de découper le dendrogramme obtenu à une certaine hauteur. En premier lieu, une analyse de la forme du dendrogramme pourra nous donner une indication sur le nombre de classes à retenir.

4.1. Méthodes de liaison

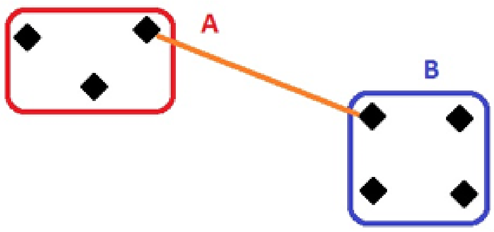
Supposant que nous avons deux clusters A et B avec plusieurs points chacun. Comment déterminer la distance entre ces deux clusters ? Voici les méthodes les plus populaires :

- **Single-linkage (lien simple, méthode du plus proche voisin)**

Le Single-linkage utilise la distance entre les éléments les plus proches dans le cluster (la distance la plus faible).

$$\forall\, A, B \text{ deux ensembles } \delta(A,B) = Min\,\{d(i,i'), i \in A\,, i' \in B\}$$

$$\delta^h(t, S_h \cup S'_h) = Min\{\delta^{h-1}(t,S_h), \delta^{h-1}(t,S'_h)\big)$$

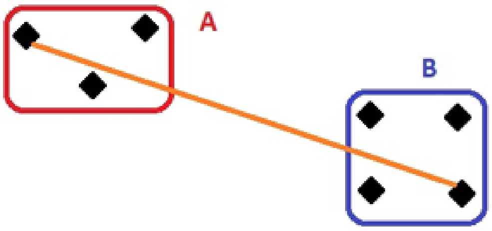


- **Liaison complète (méthode du voisin le plus éloigné)**

Le lien complet utilise la distance entre les éléments les plus éloignés.

$$\forall\, A, B \text{ deux ensembles } \delta(A,B) = Max\,\{d(i,i'), i \in A\,, i' \in B\}$$

$$\delta^h(t, S_h \cup S'_h) = Max\{\delta^{h-1}(t,S_h), \delta^{h-1}(t,S'_h)\big)$$

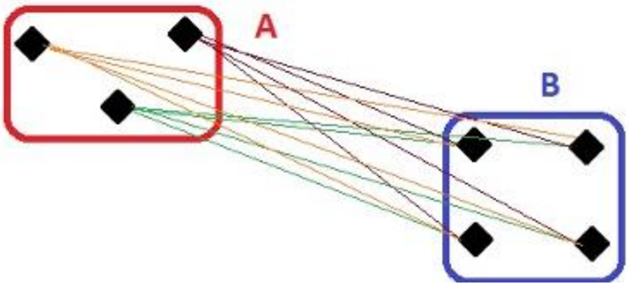


- **Average linkage (liaison moyenne, méthode de la distance moyenne)**

Le couplage moyen utilise la moyenne de toutes les distances par paire. La distance est calculée à partir de chaque combinaison et la moyenne à partir de celle-ci.

$$\forall\ A, B\ deux\ ensembles\ \ \delta(A,B)\ =\ \frac{\sum d(i,i')}{card(A)+card(B)}\ \{\ i \in A\ , i' \in B\}$$

$$\delta^h(t, S_h \cup\ S'_h) = \frac{card(S_h)\delta^{h-1}(t,S_h) + card(S'_h)\delta^{h-1}(t,S'_h)}{card(S_h) + card(S'_h)}$$



- **Ecart de Ward**

La méthode de Ward considère l'écart :

$$\delta(A,B) = \frac{P_A\, P_B}{P_A\ + P_B} d^2(g_A, g_B)$$

Où g_A est le centre de gravité de A, et g_B celui de B. On rappelle que gA est le point de coordonnées $(\bar{x}_{1,A}, \bar{x}_{2,A}, \dots\dots, \bar{x}_{p,A})$, où, pour tout j ∈ $\{1, \dots\dots, p\}$, $\bar{x}_{j,A}$ désigne la moyenne des valeurs observées du caractère X_j sur les n_A individus du groupe A. de même pour g_B.

Cette méthode prend en compte à la fois la dispersion à l'intérieur d'un groupe et la dispersion entre les groupes. A chaque itération, on agrège de manière à avoir un gain minimum d'inertie intra-classes

$$\delta^h(A, B \cup C) = \frac{(P_A + P_B)\delta^h(A,B) + (P_A + P_C)\delta^h(A,C) - P_C\delta^h(B,C)}{P_A + P_B + P_C}$$

- **Critère de la minimisation de l'inertie de la réunion de deux classes**

Le critère consiste à agréger pas à pas (partition par partition) deux classes si l'inertie (intra-classe) de leur réunion est minimale.

$$\delta^h(A, B \cup C) = \frac{(P_A + P_B)\delta^{h-1}(A,B) + (P_A + P_C)\delta^{h-1}(A,C) + (P_B + P_C)\delta^{h-1}(B,C) - P_Af(A) - P_Bf(B) - P_Cf(C)}{P_A + P_B + P_C}$$

A l'étape initiale

$$\delta^0(\{i\},\{i'\}) = \frac{P_i P_{i'}}{P_i + P_{i'}} \parallel i - i' \parallel^2$$

5. Exemple d'application

Nous avons demandé aux gens combien d'heures par semaine ils passaient sur les plateformes de médias sociaux et à la salle de sport.

	Réseaux sociaux	Salle de sport
Mohamed	**2**	**3**
Mouad	**5**	**2**
Younes	**5**	**3**
Ilyes	**1**	**4**
Fatima	**4**	**5**

Nous voulons maintenant effectuer une analyse de cluster hiérarchique. Dans la première étape, nous attribuons un cluster à chaque point. Nous avons donc autant de clusters que de personnes. Le but est maintenant de fusionner petit à petit de plus en plus de clusters, jusqu'à ce que finalement tous les points soient dans un seul cluster.

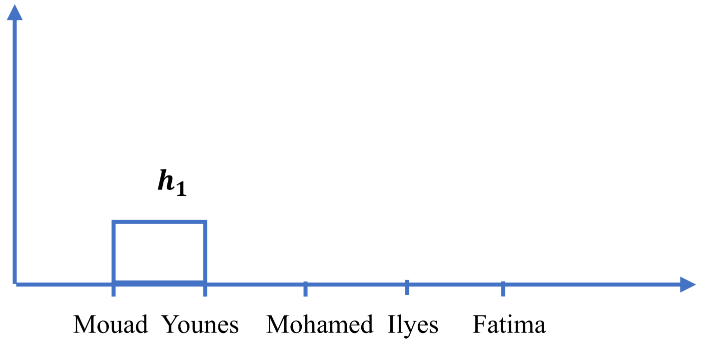
À chaque étape, les clusters les plus proches sont toujours fusionnés. Pour cela, nous devons déterminer deux choses :

- Comment la distance entre deux points est mesurée (dans cet exemple, on utilise la distance Euclidienne).
- Comment les points d'un cluster sont connectés (dans cet exemple, on utilise la méthode du lien simple).

Pour cela, nous devons d'abord calculer la matrice de distance. Dans la matrice de distance, nous entrons les clusters sur les deux dimensions, puis nous calculons les distances de chaque cluster à chaque autre cluster.

δ^0	Mohamed	Mouad	Younes	Ilyes	Fatima
Mohamed	**0**				
Mouad	**3.16**	**0**			
Younes	**3.00**	**1.00**	**0**		
Ilyes	**1.41**	**4.47**	**4.12**	**0**	
Fatima	**2.83**	**3.16**	**2.24**	**3.16**	**0**

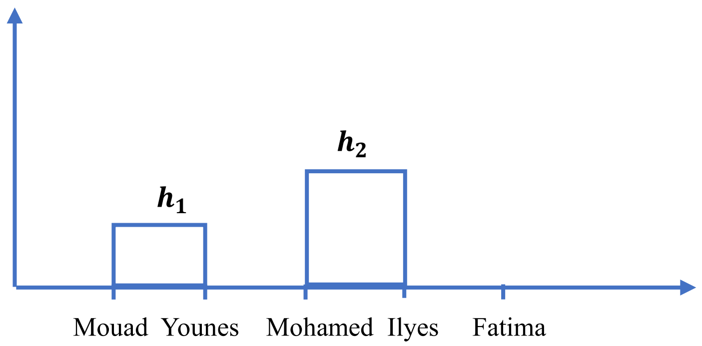
Nous pouvons maintenant fusionner les premiers clusters. Pour cela, nous regardons entre quels deux clusters nous avons la plus petite distance, c'est le cas entre Younes et Mouad. Pour cela, nous combinons maintenant Younes et Mouad en un seul cluster. Dans notre diagramme en arbre ou dendrogramme, nous pouvons dessiner la première connexion.



Nous devons maintenant mettre à jour notre matrice de distance. Nous avons décidé d'utiliser la méthode du lien simple. Ainsi, la distance entre deux clusters est donnée par les éléments qui sont les plus proches les uns des autres. Pour les clusters Mohamed, Ilyes et Fatima, du cluster Younes et Mouad respectivement, Younes est toujours la personne la plus proche.

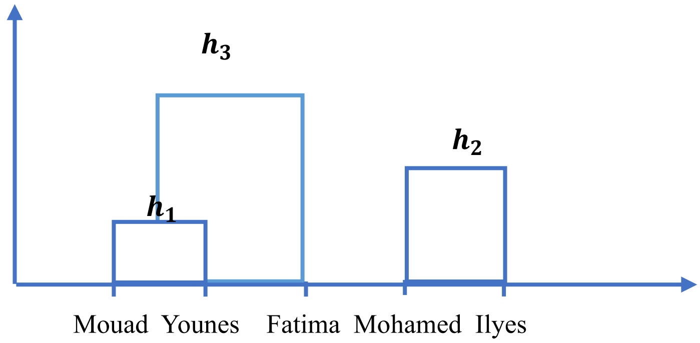
δ^1	Mohamed	Mouad, Younes	Ilyes	Fatima
Mohamed	**0**			
Mouad, Younes	**3.00**	**0**		
Ilyes	**1.41**	**4.12**	**0**	
Fatima	**2.83**	**2.24**	**3.16**	**0**

Maintenant, nous fusionnons à nouveau les clusters qui sont les plus proches. Ce sont Ilyes et Mohamed. Dans notre diagramme d'arbre ou dendrogramme, nous pouvons dessiner la deuxième connexion.

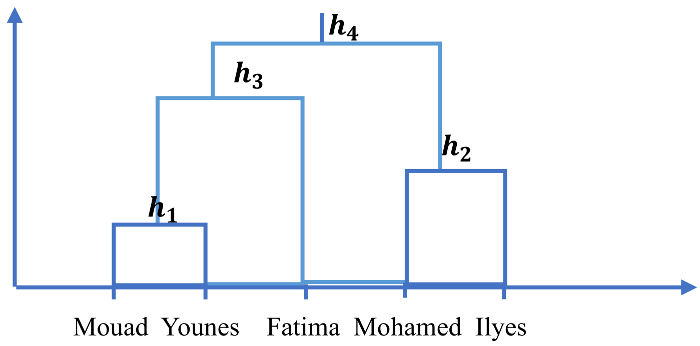


Maintenant, nous mettons à nouveau à jour la matrice de distance. Nous obtenons la plus petite distance entre le cluster Fatima et le cluster Mouad et Younes.

δ^2	Mouad, Younes	Mohamed, Ilyes	Fatima
Mouad, Younes	**0**		
Mohamed, Ilyes	**3.00**	**0**	
Fatima	**2.24**	**2.83**	**0**



Nous connectons donc ces deux clusters et dessinons la troisième connexion dans le diagramme en arbre. Il ne reste plus que deux clusters, et nous les fusionnons dans la dernière étape. Et nous obtenons notre dendrogramme terminé.



6. Etapes du regroupement hiérarchique

D'une manière générale, l'algorithme de clustering ou (re) groupement hiérarchique ascendant peut comprendre plusieurs itérations, dont le nombre maximal d'itérations est donne par le nombre et/ ou la nature d'objets à regrouper.

Cependant, on peut décider de l'arrêt du regroupement (convergence) par un seuil de regroupements selon le niveau d'agrégations désiré.

Chaque itération de l'algorithme de classification par clustering hiérarchique comprend :

1- Calcul de la matrice des distances entre objets, pour la première étape (ou mise a jour de la matrice des distances entre objets, pour les étapes suivantes) .
2- Déterminer la distance minimale (niveau d'agrégation) et répéter les groupes ou objets concernes.
3- Tester si le niveau d'agrégation n'a pas dépassé le seuil ou le nombre d'itérations est suffisant, sinon arrêt.
4- Regrouper les objets ou groupes concernées. Réitérer de l'étape 1.

7. Hiérarchie indicée :

Soient ***E*** un ensemble fini d'objets et ***H*** un ensemble de parties de ***E***. ***H*** est une hiérarchie si et seulement si :

1. $E \in H$.
2. $\forall x \in E, \{x\} \in H$.

3. $\forall h_1 \in H, \forall h_2 \in H, h_1 \cap h_2 = \emptyset$ ou $h_1 \subset h_2$ ou $h_2 \subset h_1$.

Une hiérarchie est dite indicée s'il existe une fonction de f $de\ H\ dans\ R^+$ telle que :

4. $\forall x \in E, f(\{x\}) = 0$.
5. $\forall h_1 \in H, h_2 \in H, h_1 \neq h_2, h_2 \subset h_1 \Rightarrow f(h_2) < f(h_1)$.

La fonction f est appelée indice d'agrégation. Elle est maximale pour *E*.

Exemple 2 :

	I_1	**I_2**	**I_3**	**I_4**
X	0	1	3	4
Y	1	0	4	4

- Appliquer une CHA avec le critère de la minimisation de l'inertie de la réunion de 2 classes.
- Donner toutes les partitions trouvées.
- Calculer l'inertie totale du nuage et les inerties intra-classes (des classes) des partitions trouvées.

Méthodes par partitionnement

L'algorithme des centres mobiles

K-means (k-moyennes en français ou bien Algorithme des centres mobiles) est un algorithme de clustering. Le clustering est un type d'apprentissage non supervisé qui consiste à regrouper les éléments de jeu de donnée en groupes, appelés clusters. Le but est de faire ressortir les structures cachées dans les données en regroupant les éléments qui se ressemblent. L'algorithme des k-moyennes regroupe les points en k clusters. Cela suppose qu'il faut avoir une idée du nombre de clusters pour appliquer cet algorithme.

1. Principe de fonctionnement

L'élément central de cet algorithme est le centroïde. Un centroïde est un point du jeu de données que l'on choisira comme le centre d'un cluster. C'est en fonction du centroïde que nous définirons l'appartenance à un cluster.

Un autre élément important dans cet algorithme est la distance. Une distance est une application qui associe un couple de vecteurs à un nombre réel positif. Il existe plusieurs types de distances dont la plus connue est la distance euclidienne.

L'algorithme k-means définit le centre de gravité d'un cluster comme la valeur moyenne des points au sein du cluster. Il procède comme suit :

- Tout d'abord, il sélectionne au hasard k objets, dont chacun représente initialement une moyenne ou un centre de cluster.
- Pour chacun des objets restants, un objet est affecté au cluster auquel il est le plus similaire, en fonction de la distance entre l'objet et la moyenne du cluster. L'algorithme k-means améliore ensuite itérativement la variation intra-cluster.
- Pour chaque cluster, il calcule la nouvelle moyenne à l'aide des objets affectés au cluster lors de l'itération précédente. Tous les objets sont ensuite réaffectés à l'aide des centroïdes mis à jour en tant que nouveaux centres de cluster.
- Les itérations se poursuivent jusqu'à ce que l'affectation soit stable, c'est-à-dire que les clusters formés au tour en cours soient les mêmes que ceux formés au tour précédent.

La méthode k-means n'est pas garantie de converger vers l'optimum global et se termine souvent à un optimum local. Les résultats peuvent dépendre de la sélection aléatoire initiale des centroïdes de cluster. Pour obtenir de bons résultats en pratique, il est courant d'exécuter plusieurs fois l'algorithme k-means avec différents centroïdes initiaux.

2. Algorithme des centres mobiles

- Initialisation :
  - Il faut tout d'abord déterminer combien de groupes on souhaite trouver « K ».
  - Ensuite, on commence par choisir, au hasard, k centroïdes. Qui seront les centres des clusters de départ.
- Boucle :
- On construit k clusters :

Chaque point est dans le cluster du centroïde qui lui est le plus proche.

- On calcule les nouveaux centroïdes :

Pour chacun des clusters qu'on vient de former, on calcule la moyenne. Celle-ci devient le nouveau centroïde (n'est pas nécessairement un point du jeu de donnée).

- On recommence jusqu'à ce qu'à ce qu'il y ait convergence :

La convergence correspond au fait que les centroïdes ne changent pas après une mise à jour. Ici, on dit que l'algorithme converge quand plus rien ne bouge d'une itération à l'autre, c'est-à-dire quand le centroïde reste immobile même après avoir recalculé pour chaque point son cluster.

3. **Illustration de l'algorithme des centres mobiles**

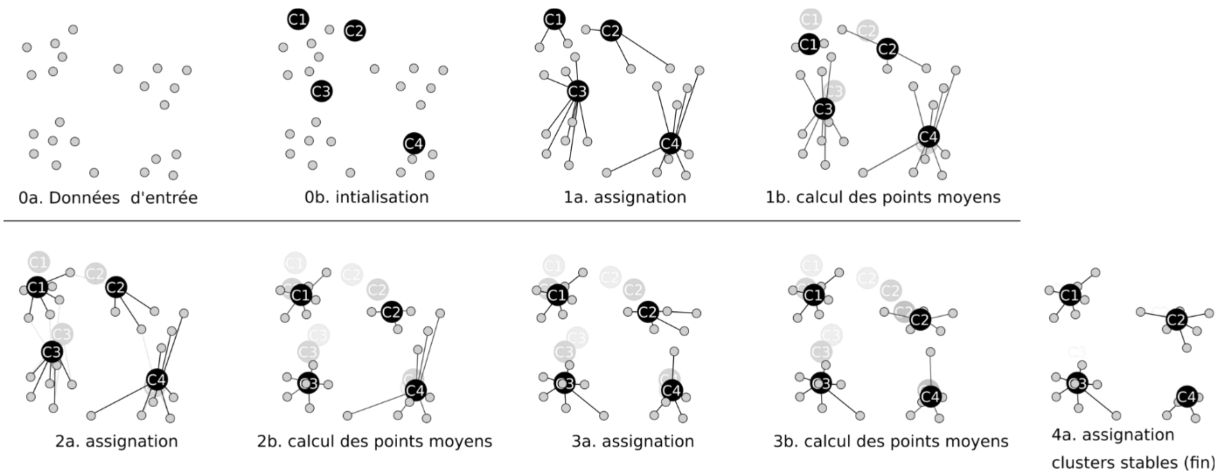


Figure 3. Exemple graphique de K-means.

4. **Affectation et critère à optimiser**

La fonction d'affectation f(L)=P

L : L'ensemble des centres(L_1, L_2,…,L_k)

P : {C_1, C_2, …, C_k}

X ∈ C_i $si \quad d^2(X,L_i) < d^2(X,L_j) \quad j \neq i$

On prend le plus petit indice en cas d'égalité.

Représentation : g(P)=L , L_i est le centre de gravite de la classe i (C_i)

Critère a optimiser : $W = \sum_{i=1}^{k} d^2(X, L_i)$

$$W = \sum_{i=1}^{k} \sum_{X \epsilon C_i} d^2\,(X, L_i)$$

5. **Avantages et inconvénients du k-means**

L'algorithme du k-means converge en général très rapidement : il n'est pas rare qu'il atteigne la convergence au bout de 10 itérations, même avec beaucoup de points.

Cependant, il faut faire attention à l'initialisation des centroïdes : comme on les place aléatoirement, le résultat donné peut être différent quand on relance l'algorithme. Il est donc conseillé de relancer plusieurs fois l'algorithme pour sélectionner la partition dont l'inertie intra-classe sera la plus petite.

Un algorithme de ce type est dit non-déterministe. Les résultats qu'il produit peuvent varier car les opérations calculatoires effectuées s'appuient sur une variable initiale aléatoire.

Le k-means n'est pas capable de déterminer le nombre de classes optimal : on est obligé de le lui spécifier au départ.

6. Exemple d'application

- **Exemple 1 :**

Dans une étude industrielle, on a étudié 2 variables : X_1 et X_2, sur 6 individus $\omega_1, \ldots, \omega_6$. Les données recueillies sont :

	X_1	X_2
ω_1	−2	2
ω_2	−2	−1
ω_3	0	−1
ω_4	2	2
ω_5	−2	3
ω_6	3	0

- Créer des clusters par l'algorithme des centres mobiles avec, pour centres initiaux, L_1^0 de coordonnées (−1, −1) et L_2^0 de coordonnées (2,3).

➢ Le tableau des distances[1] entre les individus et ces centres est

d^2	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
L_1^0	**2**	**1**	**1**	18	17	17
L_2^0	17	32	20	**1**	**16**	**10**

Exemple de calcul : $d^2(\omega_1, L_1^0) = (-2 - (-1))^2 + (2 - (-1))^2 = 2$

D'où les deux groupes :

$C_1 = \{\omega_1, \omega_2, \omega_3\}$, $C_2 = \{\omega_4, \omega_5, \omega_6\}$. *W= 2+1+1+1+16+10=31*

➢ *On considère deux nouveaux centres,* L_1^1 *et* L_2^1*, lesquels sont les centres de gravité des deux groupes* C_1 *et* C_2*. Donc* L_1^1 *a pour coordonnées* $\frac{-2-2+0}{3}$, $\frac{2-1-1}{3}$ *= (−1.33, 0) et* L_2^1 *a pour coordonnées* $\frac{2-2+3}{3}$, $\frac{2+3+0}{3}$ *= (1, 1.67)*

➢ *Le tableau des distances carré entre les individus et ces centres est*

d^2	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
c_1^1	***4.44***	***1.44***	***2.76***	*15.05*	***9.42***	*18.75*
c_2^1	*9.12*	*16.16*	*8.12*	***1.10***	*10.76*	***6.81***

D'où les deux groupes :

C_1= {ω_1, ω_2, ω_3, ω_5}, C_2 = {ω_4, ω_6}. W= 4.44+1.44+2.76+1.1+9.42+6.81=25.97

- *On considère deux nouveaux centres L_1^2, et, L_2^2 lesquels sont les centres de gravité des deux groupes C_1 et C_2. Donc L_1^2a pour coordonnées $\frac{-2-2+0-2}{4}$, $\frac{2-1-1+3}{4}$ = (−1.5, 0.75) et L_2^2 a pour coordonnées $\frac{2+3}{2}$, $\frac{2+0}{2}$ = (2.5, 1)*

- *Le tableau des distances entre les individus et ces centres est*

d^2	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
L_1^2	***1.82***	***3.31***	***5.29***	*3.84*	***5.29***	*20.79*
L_2^2	*21.25*	*24.20*	*10.24*	***1.25***	*24.20*	***1.25***

D'où les deux groupes :

C_1 = {ω_1, ω_2, ω_3, ω_5}, C_2 = {ω_4, ω_6}. W=1.82+3.31+5.29+1.25+5.29+1.25=18.21

On retrouve le même clustering que l'étape précédente, on arrête l'algorithme.

- **Exemple 2 :**

	I_1	**I_2**	**I_3**	**I_4**
X	0	1	3	4
Y	1	0	4	4

Appliquer l'algorithme de Kmeans avec K=2 et $L_1 = \binom{1}{1}\ et\ L_2 = \binom{2}{2}$

- **Méthodes non paramétrique**
  Classification par morphologie mathématique

Les méthodes morphologiques s'intéressent a la forme du nuage de points sans calcul d'aucun paramètre visuel.

Les opérateurs de base de ces méthodes sont les opérateurs de la morphologie mathématique.

Une opération essentielle est la discrétisation (passer d'un espace continu a un espace discret).

Soit $y_1, y_2, \dots, y_n$ un échantillon de n individus.

$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \cdot \\ \cdot \\ y_{ip} \end{pmatrix}$, L'origine est translaté au point $\theta' = \begin{pmatrix} \min y_{i1} \\ \min y_{i2} \\ \cdot \\ \cdot \\ \min y_{ip} \end{pmatrix}$

Exemple

$\boldsymbol{y_1}$	1	3	5	1	**0**	2
$\boldsymbol{y_2}$	2	4	2	**1**	1	4

$O = \binom{0}{0}, O' = \binom{0}{1}$

1. **La transformation diagonale :**

$$y'_{ij} = \frac{y_{ij} - min\ y_{ij}}{max\ y_{ij} - min\ y_{ij}} . R$$

Ceci permet de situer les observations dans un hypercube de cote R (résolution).

$O' = \binom{0}{1} \Rightarrow$

$\boldsymbol{y_1}$	1	3	5	1	0	2
$\boldsymbol{y_2}$	1	3	1	0	0	3

R=6

j=1 $\max y_{ij} = 5\ , min\ y_{ij} = 0$

j=2 $\max y_{ij} = 3\ , min\ y_{ij} = 0$

j=1 $y_{ij}' = \frac{y_{ij} - 0}{5 - 0} . R = y_{ij}\ \frac{5}{6}$

j=2 $y_{ij}' = \frac{y_{ij} - 0}{3 - 0} . 6 = 2. y_{ij}$

$\boldsymbol{y_1}'$	1.2	3.6	6	1.2	0	2,4
$\boldsymbol{y_2}'$	2	6	2	0	0	6

Chaque axe du nouvel espace est découpé en R (ou R+1) intervalle adjacent égaux de longueur unité.

0 1 2 3 4 5 6 7

		1	1			
	1					1
1	1					

0 1 2 3 4 5 6 7

- **La discritisation :**

Cette discrétisation définie un ensemble hypercube (carre) Hi de cote unité dont chacun est repéré par les p parties entiers.

Puisque l'observation (le point) y'_i est située dans l'hypercube des coordonnées. Cet hypercube est mis a 1.

$$H_i = \begin{pmatrix} \text{int}(y'_{i1}) \\ \text{int}\,(y'_{i2}) \\ . \\ . \\ \text{int}\,(y'_{ip}) \end{pmatrix}$$

Aprés discrétisation :

X= {(0,0),(1,0),(1,2),(2,6),(3,6),(6,2)}

Remarque :

Le chois de R de la discrétisation est décisif pour les traitements de la morphologie mathématique. En effet R est équivalent a une résolution si R est très petit toutes les données vont se trouvées dans un seul hypercube (la forme du nuage sera perdu). D'un autre cote si R est trop grand les données vont être trop dispersées dans l'ensemble discret.

2. **Transformation Morphologique :**

2.1. Dilatation :

$$\boldsymbol{X} \bigoplus \boldsymbol{S} = \bigcup_{s \in S} (\boldsymbol{X})_s$$

$(\boldsymbol{X})_{\boldsymbol{s}}$: la translation de X par S = {t=x+s / $x \in \boldsymbol{X}$}

$\boldsymbol{S}$: est l'élément structurant généralement plus petit par rapport a l'ensemble X

Exp :

$$\boldsymbol{S} = \begin{matrix} \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \end{matrix}$$ L'origine

$$\boldsymbol{S} = \{(\mathbf{0},\mathbf{0}),(\mathbf{1},\mathbf{0}),(\mathbf{0},\mathbf{1}),(-\mathbf{1},\mathbf{0}),(\mathbf{0},-\mathbf{1})\}$$

Suite de l'exemple

$S = \{(0,0),(1,1)\}$

$X \bigoplus S = (X)_{s1} \cup (X)_{s2}$

$(X)_{s1} = X$

$(X)_{s2} = \{(1,1),(2,1),(2,3),(7,3),(3,7),(4,7)\}$

$X \bigoplus S = \{(0,0),(1,0),(1,2),(6,2),(2,6),(3,6),(1,1),(2,1),(2,3),(7,3),(3,7),(4,7)\}$

Propriétés :

La dilatation est une opération croissante :

$\boldsymbol{X} \subset \boldsymbol{Y} \;\; \Rightarrow \; X \oplus S \;\subset\; Y \oplus S$

Opération extensive

$\boldsymbol{X} \subset \;(X \oplus S\;)$

Distributivité :

$(X \cup Y) \oplus S = (X \oplus S) \cup (Y \oplus S)$

Relation d'itération :

$(X \oplus S_1) \oplus S_2 = X \oplus (S_1 \oplus S_2)$

2.2. L'érosion

C'est l'opération duale de la dilatation par rapport a la complémentation définie comme suivant :

$\{X \ominus S = \cap_{s \in S}\,(X)\_{s}$ ou $X \ominus S = \cap_{s \in S}\,(X)_s$

Propriétés :

Operation croissante:

$X \subset Y \;\Rightarrow X \ominus S \subset Y \ominus S$

Operation anti-extensive:

$(X \ominus S) \subset X$

Distributivité

$(X \cap Y) \ominus S = (X \ominus S) \cap (Y \ominus S)$

$X \ominus (S_1 \cup S_2) = (X \ominus S_1) \cap (X \ominus S_2)$

Relation d'iteration:

$(X \ominus S_1) \ominus S_2 = X \ominus (S_1 \ominus S_2)$

Remarque:

Pour chaque ensemble discret noté sous forme binaire on doit mentionner obligatoirement le point origine.

L'erosion est une opération de filtrage qui ne garde que les éléments de X représentant l'image contenue dans le filtre (élément structurant). Plus l'élément structurant est grand plus l'effet de filtrage est important.

- **L'ouverture**

  $\boldsymbol{X \circ S = (X \ominus S) \oplus S}$

  Propriétés

- **Croissante :** $\boldsymbol{X \subset Y \Longrightarrow (X \circ S) \subset (Y \circ S)}$
- **Opération anti-extensive : (X**$\circ$ $\boldsymbol{S)} \subset \boldsymbol{X}$
- **Relation d'idempotance :** $\boldsymbol{(X \circ S) = (X \circ S) \circ S}$ (quelque soit le nombre de fois qu'on répète l'opération, le resultat reste le même)

- **La fermeture**

  $X \cdot S = (X \oplus S) \ominus S$

  Propriétés

- **Croissante :** $\boldsymbol{X \subset Y \Longrightarrow (X\,.\,S) \subset (Y.S)}$
- **Opération extensive :** $\boldsymbol{X \subset (X.S)}$
- **Relation d'idempotence :** $\boldsymbol{(X.S) = (X.S).S}$

Remarque :

Puisque l'ouverture fait ressortir les regroupements d'individu de l'ensemble X (selon le choix de l'élément structurant) elle peut être utilisée dans la classification comme étant une méthode de détection de région modales.