

Algorithme de la classification ascendante hiérarchique

Phase préalable : Calcul des dissimilarités des objets 2 à 2

Entrées : $n(n - 1)/2$ dissimilarités

Regroupement des 2 éléments
les plus proches

jusqu'au regroupement
de tous les objets
en un seul groupe :
($n - 1$) étapes

Calcul d'une nouvelle matrice de
dissimilarités entre les éléments
(objets isolés ou groupes) restants

Stratégie d'agrégation

1^{ère} étape :

si d est une dissimilarité, on choisit e_i et $e_{i'}$ tels que $d(e_i, e_{i'})$ minimum $\Rightarrow G_1 = \{e_i, e_{i'}\}$

2^{ème} étape :

nouveau tableau de dissimilarités $(n - 1) \times (n - 1)$

\Rightarrow nécessité de définir une *méthode d'agrégation* entre un individu et un groupe d'individus ou entre deux groupes d'individus.

Méthodes d'agrégation

Lien minimum

$$\delta(A, B) = \min \{d(a, b), a \in A, b \in B\}$$

Lien maximum

$$\delta(A, B) = \max \{d(a, b), a \in A, b \in B\}$$

Distance des centres de gravité

$$\delta(A, B) = d(g_a, g_b)$$

...

Exemple

Agrégation selon le lien minimum

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<i>a</i>	0	23	35	43	50
<i>b</i>	23	0	21	32	45
<i>c</i>	35	21	0	11	25
<i>d</i>	43	32	11	0	17
<i>e</i>	50	45	25	17	0

$$G_1 = \{c, d\} \Rightarrow$$

	<i>a</i>	<i>b</i>	<i>e</i>	G_1
<i>a</i>	0	23	50	35
<i>b</i>	23	0	45	21
<i>e</i>	50	45	0	17
G_1	35	21	17	0

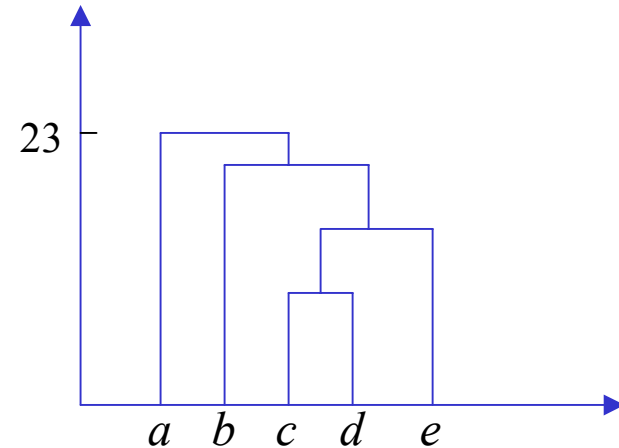
Tableau des dissimilarités

$$G_2 = \{e, G_1\} \Rightarrow$$

	<i>a</i>	<i>b</i>	G_2
<i>a</i>	0	23	35
<i>b</i>	23	0	21
G_2	35	21	0

$$G_3 = \{b, G_2\} \Rightarrow$$

	<i>a</i>	G_3
<i>a</i>	0	23
G_3	23	0



Exemple (suite)

Agrégation selon le lien maximum

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<i>a</i>	0	23	35	43	50
<i>b</i>	23	0	21	32	45
<i>c</i>	35	21	0	11	25
<i>d</i>	43	32	11	0	17
<i>e</i>	50	45	25	17	0

Tableau des dissimilarités

$$G_1 = \{c, d\}$$

⇒

	<i>a</i>	<i>b</i>	<i>e</i>	G_1
<i>a</i>	0	23	50	43
<i>b</i>	23	0	45	32
<i>e</i>	50	45	0	25
G_1	43	32	25	0

$$G_2 = \{a, b\}$$

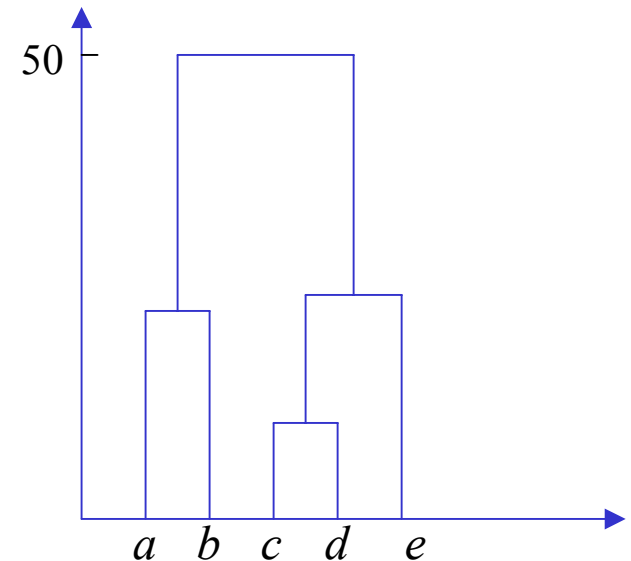
⇒

	<i>e</i>	G_1	G_2
<i>e</i>	0	25	50
G_1	25	0	43
G_2	50	43	0

$$G_3 = \{e, G_1\}$$

⇒

	G_2	G_3
G_2	0	50
G_3	50	0



Classification dans un espace euclidien

Inerties interclasse et intraclasse

Soit une classification en k groupes d'effectifs n_1, \dots, n_k , les individus étant des points d'un espace euclidien. Notons les groupes G_1, \dots, G_k , et g_1, \dots, g_k leurs centres de gravité (g est le centre de gravité du nuage).

$$\textit{Inertie totale} : \quad I_{tot} = \frac{1}{n} \sum_{i=1}^n d^2(e_i, g)$$

$$\textit{Inertie interclasse} : \quad I_{inter} = \frac{1}{n} \sum_{i=1}^k n_i \cdot d^2(g_i, g)$$

$$\textit{Inertie intraclasse} : \quad I_{intra} = \frac{1}{n} \sum_{i=1}^k \sum_{e \in G_i} d^2(e, g_i)$$

Critère d'agrégation selon l'inertie

Théorème de Huygens :

Inertie totale = Inertie inter-classe + Inertie intra-classe

Au fur et à mesure que les regroupements sont effectués, l'inertie intra-classe augmente et l'inertie interclasse diminue, car leur somme est une constante liée aux données analysées.