Le tab suivant est un tab d'ultrametrique

E \ E	1	2	3	4
1	0	6	0	6
2	6	0	5	2
3	0	5	0	5
4	6	2	5	0

Theorème : Il existe une bijection entre une hierarchie indicée et ultrametrique.

En d'autres termes, toute hierarchie indicée sur E correspond une ultrametrique et inversement.

On propose deux familles de méthodes pour obtenir une hierarchie indicée : - Les regroupements progressifs

- Les passages directes à l'ultrametrique

V. 4 / Les regroupements progressifs.

On a plusieurs étapes successives.

Sur le tableau des distances ou dissimilarité, on repère la plus petite valeur

Soit A la classe $\{(w_i ; w_{i'})\}$ tq $d(w_i, w_{i'})$ = plus petite valeur.

On calcule la distance ~~ou~~ dissimilarité de chaque ind à A, on obtient un tableau de dim $(n-1, n-1)$

On repère du nouveau tableau la dist ou diss la plus petite

On répète la procedure jusqu'à obtenir un regroupement complet

Rq si la + petite valeur correspond à une dist ou dissi entre deux autres ind il seront regroupés ds une m̃ classe, si cela correspond à dist ou dissi entre un ind et A on forme une classe B de 3 inds.

On a 3 manière de calculer les distances par utilisation d'indice ~~au Mesure d'agregation du lien minimum~~. d'agregation

Déf : On appelle indice d'agregation entre deux groupes d'individus une application $\delta : \mathbb{P}(E) \times \mathbb{P}(E) \longrightarrow \mathbb{R}_+$

$(h_1, h_2) \longmapsto \delta(h_1, h_2)$ tq $\delta(h_1, h_2) = \delta(h_2, h_1)$

1) Indice d'agrégation du lien minimum (Jardin et Sibson 1967)

$$\delta(h_1, h_2) = \min_{\substack{w_i \in h_1 \\ w_j \in h_2}} d(w_i, w_j)$$ d indice de dissimilarité.

Un ind w est affecté à une classe s'il est plus proche (au sens de la dissimilarité ou distance) d'un elt de cette classe que d'un elt d'une autre classe. avec $d(w, A) = \delta(\{w\}, A)$

2) Indice d'agrégation du lien maximum (Sorensen 1948)

$$\delta(h_1, h_2) = \max_{\substack{w_i \in h_1 \\ w_j \in h_2}} d(w_i, w_j)$$

un ind w est affecté à une classe ~~en~~ A s'il est plus près de touts les elts de cette classe que de tous les elts des autre classes.

(La plus petite distance de l'indice d'agrégation max)

3) Indice d'agrégation de l'inertie (indice de ward)

la dist utilisée $\delta(h_1, h_2) = \dfrac{Card(h_1)\, Card(h_2)}{Card(h_1) + Card(h_2)} d^2(g_{h_1}, g_{h_2})$ g_h centre de gravité de h.

g_h est le p^t moyen de h.

On utilisé parfois

4) indice d'agrégation de la moyenne (Sokal, Michenov 1958)

$$\delta(h_1, h_2) = \frac{1}{Card(h_1) . Card(h_2)} \sum_{\substack{w_i \in h_1 \\ w_j \in h_2}} d(w_i, w_j)$$

5) indice d'agrégation des centre de gravité:

$$\delta(h_1, h_2) = d(g_{h_1}, g_{h_2})$$

Exp: Soit E un ensemble à 4 ind $E = \{1, 2, 3, 4\}$, on donne le tab des distances entre elts de E.

E \ E	1	2	3	4
1	0	7	6	7
2	7	0	5	2
3	6	5	0	6
4	7	[illegible]	6	[illegible]

a) Par l'indice d'agrégation du lien minimal pour obtenir une hiérarchie
La distance minimale entre deux ind est 2 = d(2, 4)

le tabl. (3 x 3)

E \ E	1	{2,4}	3
1	0	$\delta(\{1\},\{2,4\}) = 7$	6
{2,4}	7	0	5
3	6	δ	0

$\delta = \min\{d(2,3), d(4,3)\}$.

On regroupe 3 et {2, 4} car d({3}, {2,4}) = 5 est la + petite.

on obtient le tabl. 2x2.

E \ E	1	{2,4,3}
1	0	6
{2,4,3}	6	0

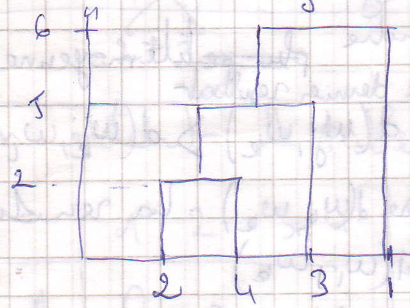
d(1,2) = 7 , d(1,4) = 7 et d(1,3) = 6

$\delta(1, \{2,3,4\}) = 6$ (dist min).

On regroupe 1 et {2, 4, 3} on obtient le tabl. (1, 1)

E \ E	{1, 2, 3, 4}
{1,2,3,4}	0

Ainsi l'hiérarchie indicée obtenue avec le lien min est



b) Par l'indice d'agrégation du lien maximal pour le m̂ exp.

étape 1: d({2}, {4}) = min = 2 on regroupe 2 et 4 ds une classe {2, 4}

	1	{2,4}	3
1	0		
{2,4}		0	
3			0

$\delta(1, \{2,4\}) = \max\{d(1,2); d(1,4)\} = \max\{7, 7\} = 7$

~~$\delta(3, \{2,4\}) = \max\{d(3,2); d(3,4)\} = \max\{5, 6\} = 6.$~~

~~3 est plus proche de {2,4} on regroupe 3 et {2,4} en une classe {2, 4, 3}~~

$\delta(3, 1) = 6.$

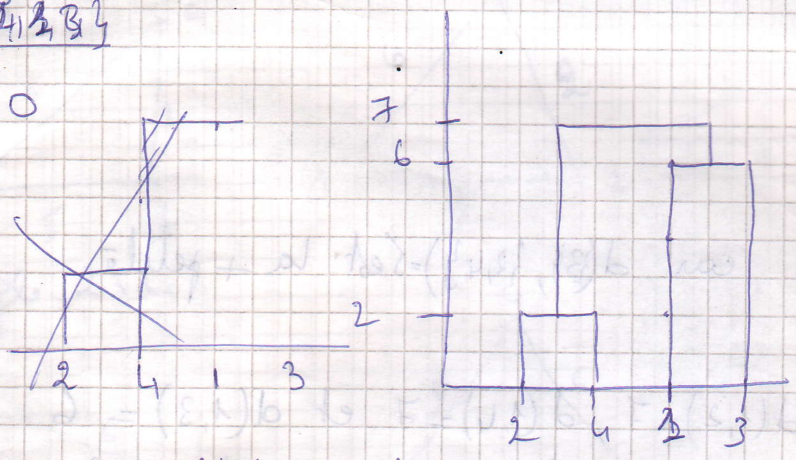
on regroupe {3, 1}

	{1,3}	{2,4}
{1,3}	0	7
{2,4}	7	0

$7 = \max d(A,B)$ $A=\{2,4\}$; $B\{1,3\}$

On regroupe $\{1,3\}$ et $\{2,4\}$.

	{2,4,1,3}
{2,4,1,3}	0



V.5 Passage à l'ultramétrique.

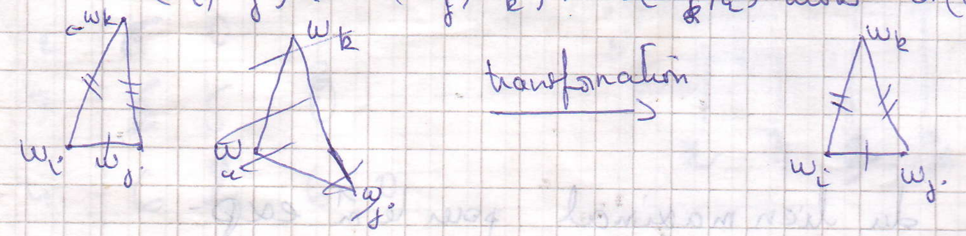
Le principe est de transformer la distance ~~en~~ (ou dissimilarité) en une ultramétrique pour obtenir une hiérarchie indicée, ce qui revient à rendre touts les triangles $w_i w_j w_k$ isocèles avec une base ≤ aux cotés.

On propose deux façons :

1 – Passage à l'ultramétrique inférieure maximale.

Pour tout triplet $A=\{w_i, w_j ; w_k\}$ reduire la plus grande distance (ou dissimilarité) entre de p^ts de A $\boxed{\text{La plus grande} \overset{\text{rendre}}{=} \text{~~plus petite~~ moyenne du dernier resultat}}$

si $d(w_i, w_j) > d(w_j, w_k) > d(w_k, w_i)$ alors $d(w_i, w_j) \overset{=}{=} d(w_j, w_k) > d(w_k, w_i)$

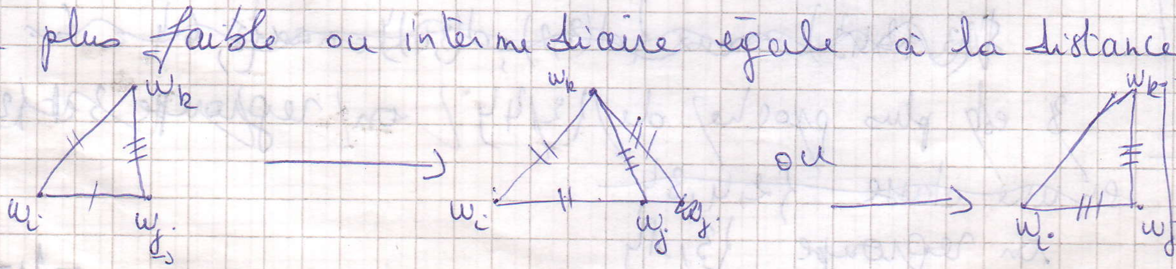


reduire $d(w_j, w_k)$. la rendre $= d(w_i, w_k)$

(plus grand coté = 2^e plus grand)

2 – Passage à l'ultramétrique supérieure maximale.

Pour tout triplet $A=\{w_i, w_j, w_k\}$ rendre la distance (ou dissimilarité) la plus faible ou intermédiaire égale à la distance supérieure maximale



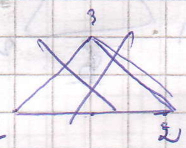
La plus faible ou int = plus grande du dernier resultat

Exemple :

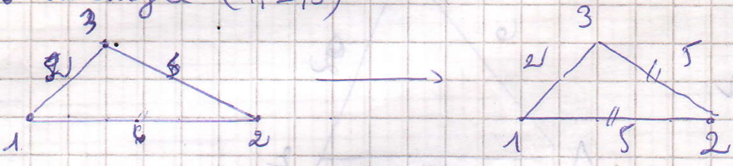
Soit l'ensemble $E = \{1, 2, 3, 4\}$; le tableau de distances est :

$E \backslash E$	1	2	3	4
1	0			
2	~~6~~ 5	0		
3	2	5	0	
4	6	1	6	0

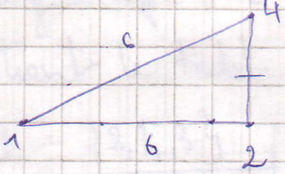
par la stratégie de passage à l'ultrametrique inferieure maximale pour obtenir une hierarchie indicée



le triangle (1, 2, 3)

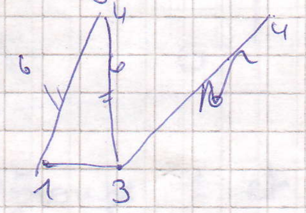


Les triangle 1 2 4



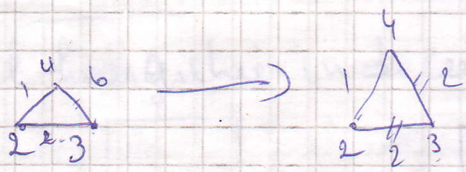
à ne pas transformer.

le triangle 1 3 4



" "

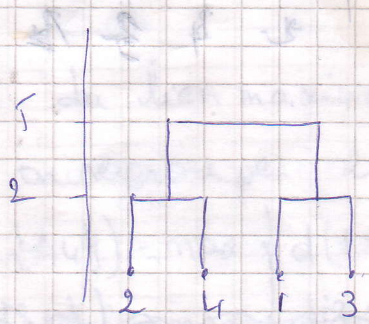
le triangle 3 2 4



On obtient le table de distance ultrametrique.

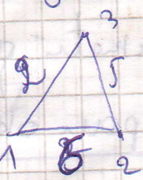
	1	2	3	4
1	0			
2	5	0		
3	2	5	0	
4	6	2	6	0

La hierarchie indicée

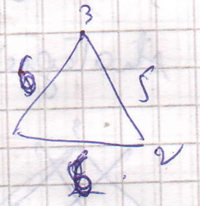


Pour le m̄ tab avec l'ultrametrique superieure maximale .

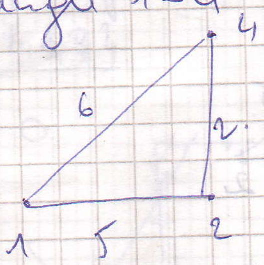
le triangle 123

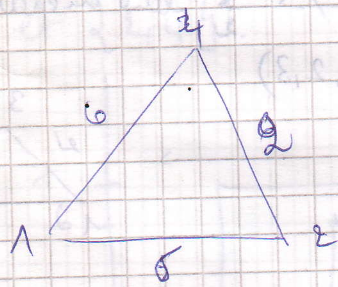


~~rien à faire~~

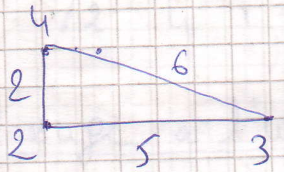


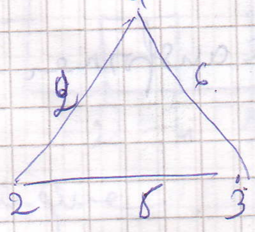
le triangle 124



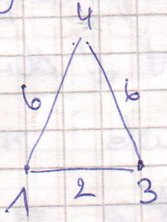


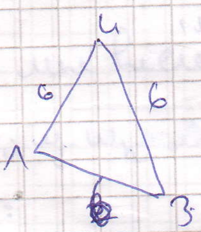
le triangle 234





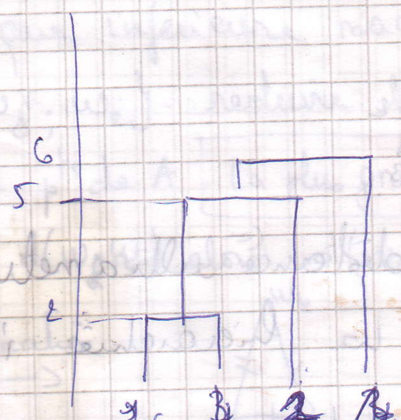
le triangle 134





le tableau .

	1	2	3	4
1	0			
2	6	0		
3	6	5	0	
4	6	2	6	0



Soit une partition P d'un ensemble $E = \{w_1, \dots, w_n\}$ en k classe.

$P = \{C_1 ; \dots , C_k\}$.

On defini l'inertie d'un ensemble A par

$$I_A = \frac{1}{n} \sum_{i=1}^{n} d^2(w_i, G) \quad G \text{ centre de gravité de } A$$

Inertie intra_classe : interieur des classes.

Inertie Intraclasses :

$$I_W(E) = \sum_{i=1}^{k} P_i \, I_{A_i} \qquad P_i = \frac{\text{card}(C_i)}{n}$$

Inertie Interclasses

$$I_B(E) = \sum_{i=1}^{k} p_i \, d^2(G_i ; G)$$

<u>Proposition</u>

$I_A = I_W + I_B$?

$$\cancel{I_W = \frac{1}{n} \sum_{i=1}^{k}} p_i \sum_{j=1}^{n_i} d^2(w_{j_i} ; g_i)$$

$$I_W = \sum_{i=1}^{k} P_i \, I_{C_i} = \sum_{i=1}^{k} \frac{\text{card}(C_i)}{n} \cdot \frac{1}{\text{card}(C_i)} \sum_{j=1}^{\text{card}(C_i)} d^2(w_j, g_{C_i}).$$

$$= \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{\text{card}(C_i)} d^2(w_j, g_{C_i})$$

$$d^2(w_j, g_{C_i}) = d^2(w_i, g) + d^2(g, g_{C_i})$$

$$I_W = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{\text{card}(C_i)} \left[d^2(w_j, g) + d^2(g, g_{C_i}) \right]$$

$$= \frac{1}{n} \sum_{i=1}^{k} \sum_{j=}^{\text{card} C_i} d^2(w_j, g) + \frac{1}{n} \sum_{i=1}^{k} \text{card}(C_i) \, d^2(g, g_{C_i})$$