corrigé du rattrapage 2021
$2^{ème}$ Année STID

$7p^t$

Exercice 1

1/) 1/ Classification : Repartition des individus d'une population en s-ensembles de sorte que ~~deux~~ individus d'un même s-ensemble soient proche au sens d'une mesure de ressemblance

0,5

2/ Dissimilarité : est une application definie d'une population E ds $\mathbb{R}_+$ tq :

$d : E \times E \longrightarrow \mathbb{R}_+$ 1

- $d(x, y) = d(y, x)$
- $d(x, x) = 0$.

3) Partition : un ensemble de sous ensemble d'une population E

1 tq $A, B \in E$ alors $A \cap B = \phi$, $\forall A \in P$ $A \neq \phi$

La reunion de tous les ensembles de la partition forme E.

4) Hierarchie.

1 Une hierarchie h sur E est un ensemble de s-ensemble de E tq

- $\forall x \in E$ $\{x\} \in h$.
- $E \in h$ et $\in \in h$.
- ~~Si~~ $h_1, h_2 \in h$ Si $h_1 \cap h_2 \neq \phi$ alors soit $h_1 \in h_2$ ou $h_2 \subset h_1$.

5) Discrimination

0,5 Affectation d'un individu à une classe d'une partition

6) Matrice des variances totale, inter et intra classe

Soit $X = (x_{ij})$ une matrice de données de dim $(n \times p)$

Matrice de variance totale $\underset{(p,p)}{V} = \cancel{\frac{1}{n}} \sum_{i=1}^{n} p_i (x_{ij} - \cancel{\bar{X}^j})^2$

$\underset{(p,p)}{V} = \sum_{i=1}^{n} p_i \, {}^t\tilde{x}_i \cdot \tilde{x}_i = {}^t\tilde{X} D_p \tilde{X}$

p_i poid du $i^{ème}$ ind $\quad x_i = (x_{i1}, \ldots, x_{ip})$ $\quad D_p = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix}$

$\tilde{x}_i = (x_{i1} - \bar{X}^1, \ldots, x_{ip} - \bar{X}^p)$ vecteur centré. 0,5

$\tilde{X}$ matrice centrée

(0,5) Matrice des vars inter_classes
(0,5) Matrice de vars intra classes.

II) Inertie totale $I_E = \sum_{i=1}^{n} p_i\, d^2(x_i ; G)$ (0,5)

$E = \{x_1, \dots, x_n\}$ p_i poids de x_i G centre de gravité de E

Inertie intra_classe. $I_W = \sum_{i=1}^{k} P_i\, I_{C_i}$ (0,5)

$\{C_1, \dots, C_k\}$ une partition de k classe de E

P_i = poids de la classe C_i $P_i = p_i \times$ Cardinale de C_i

Inertie inter_classe $I_B = \sum_{i=1}^{n} p_i\, d^2(G_i, G)$ (0,5)

G_i centre de gravité de C_i

III) 1 _ s est une dissimilarité X

Exercice 2.

1) Distance euclidienne.

d	a	b	c	d	e
a	0	$\cancel{\sqrt{2}}$ 6	6	$3\sqrt{5}$	$\cancel{\sqrt{10}}$ 6
b		0	$\sqrt{26}$	$\cancel{\sqrt{29}}$ $3\sqrt{5}$	$\cancel{2}$ $\sqrt{65}$
c			0	$3\sqrt{5}$	$\cancel{\sqrt{10}}$ $3\sqrt{5}$
d				0	5
e					0

(2) (0,25)

La plus petite valeur est $\sqrt{2} = d(a,b)$, on forme la partie $\{a,b\}$

	$\{a,b\}$	c	d	e
$\{a,b\}$	0	6	$3\sqrt{5}$	$\sqrt{10}$
c		0	$3\sqrt{5}$	$\sqrt{10}$
d			0	5
e				

(0,25)

La plus petite valeur est $\sqrt{10} = d(\{a,b\}, e)$ on forme la partie de $\{a,b,e\}$.

ou $d(c,e) = \sqrt{10}$ $\{c,e\}$. (0,25)

	$\{a,b,e\}$	c	d
$\{a,b,e\}$	0	$\sqrt{6}$	$3\sqrt{5}$
c		0	$3\sqrt{5}$
d			0

La plus petite valeur est

$\sqrt{6} = d(\{a,b,e\};\{c\})$

On forme la partie $\{a,b,e,c\}$

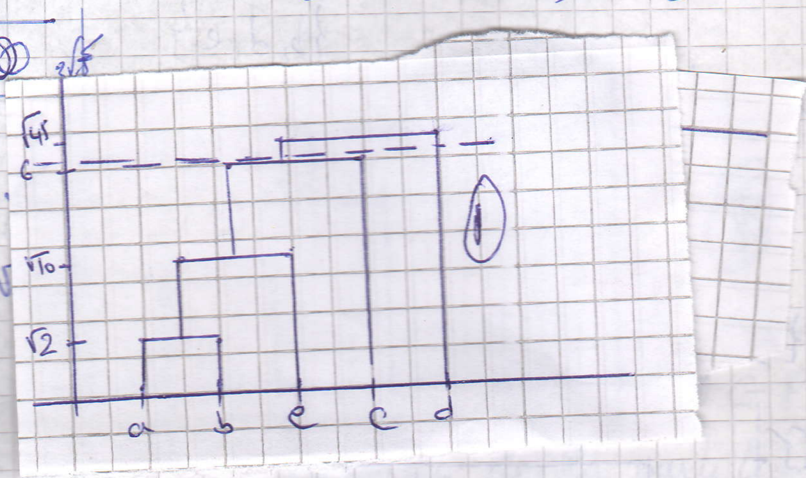
0,25

	$\{a,b,c,e\}$	d
$\{a,b,c,e\}$	0	$3\sqrt{5}$
d	$3\sqrt{5}$	0

0,75

$H = \{\{a\};\{b\};\{c\};\{d\};\{e\};\{a,b\};\{a,b,e\}$

$\{a,b,c,e\};\ E\}$ 1

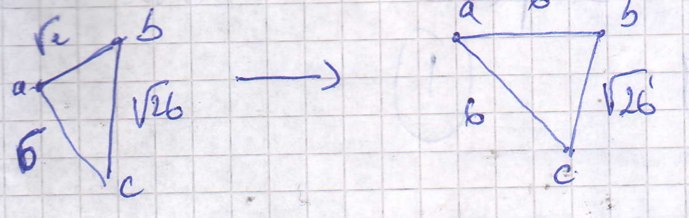


) classification à deux classes : on coupe entre $\sqrt{36}$ et $\sqrt{45}$

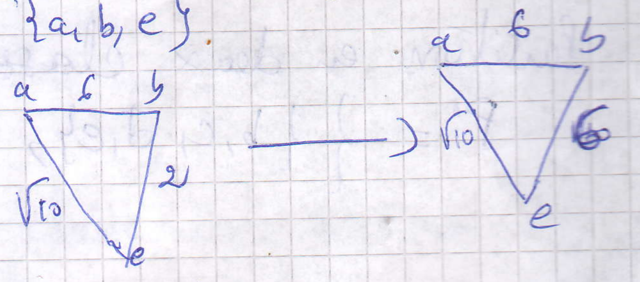
$P = \{\{a,b,e,c\};\{d\}\}$ 1

2) ultramétrique supérieure max $C_5^2 = 10$ triangle

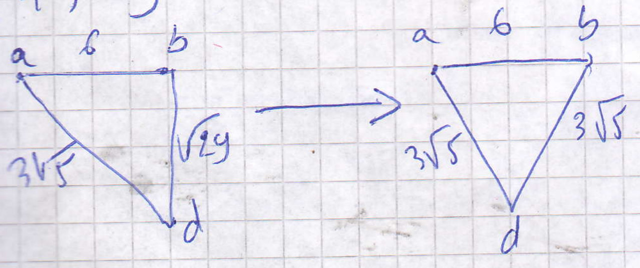
$\{a,b,c\}$



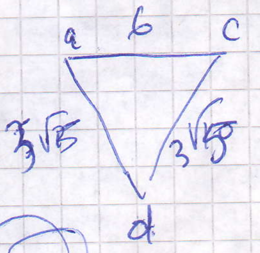
$\{a,b,e\}$



$\{a,b,d\}$



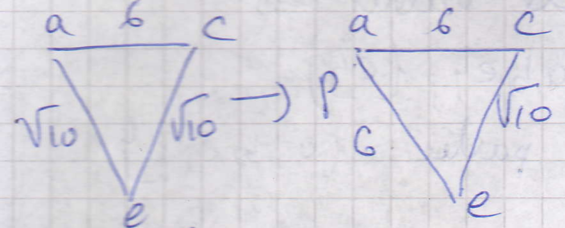
$\{a,c,d\}$



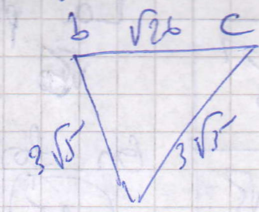
→ Pas de changement

2

$\{a, c, e\}$

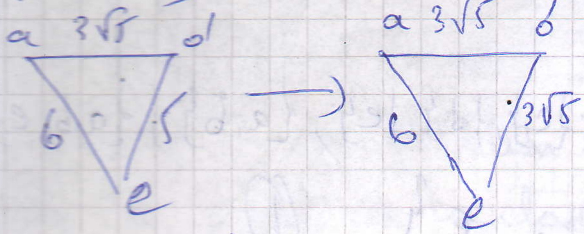


$\{b, c, d\}$

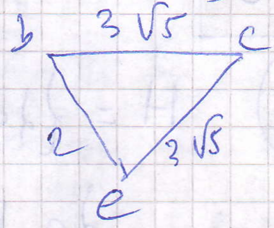


Pas de changement

$\{a, d, e\}$

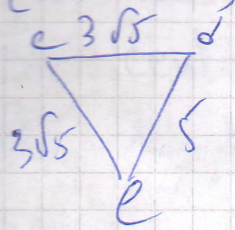


$\{b, c, e\}$



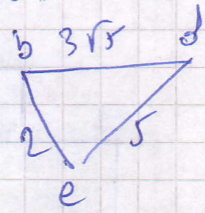
Pas de changement

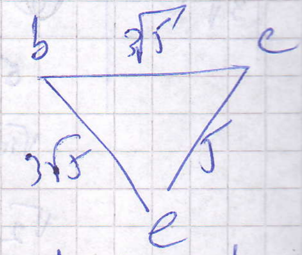
$\{c, d, e\}$



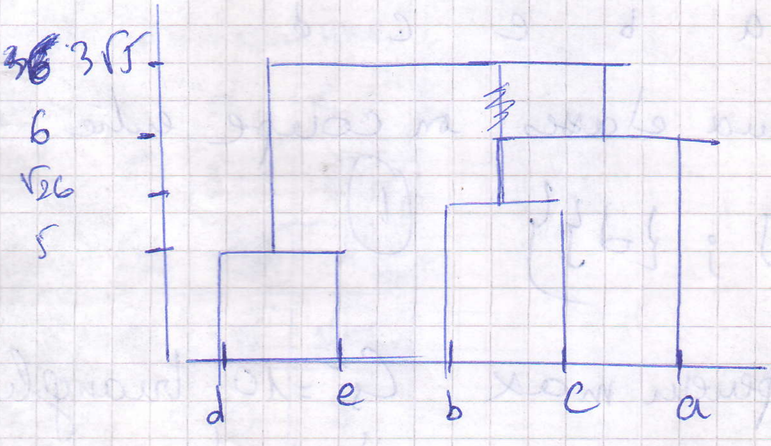
Pas de changement

$\{b, d, e\}$





$H = \{ \{a\} \{b\}, \{c\} \{d\} \{e\}, \{d, e\} \{b, c\}, \{a, b, c\} ; \{a, b, c, d, e\}$ (1)



(1)

Partition en deux classe : on coupe entre $6 = \sqrt{36}$ et $3\sqrt{5} = \ldots$

$P' = \{ \{b, c, d, e\}, \{a\} \}$ (1)