

Exercice 1

a) EST_1

EST_2

		Prédictions	
		T	F
Réelle	T	6	1 ^{FN}
	F	8 ^{FP}	2

		Prédictions	
		T	F
Réelle	T	4	3 ^{FN}
	F	1 ^{FP}	9

X1	X2	Y	Y_EST_1	Y_EST_2
1	2	T	T	T
2	2	F	T	F
1	1	T	F	F
1	2	F	T	T
1	2	T	T	T
2	2	F	T	F
2	1	F	T	F
2	2	T	T	F
2	1	F	T	F
2	1	F	T	F
1	2	T	T	T
1	1	F	F	F
2	1	T	T	F
1	1	F	F	F
2	2	F	T	F
2	1	F	T	F
1	2	T	T	T

Risque empirique 1: $1 \times 8 + x \times 1 = x + 8 = R1$

Risque empirique 2: $1 \times 1 + x \times 3 = 3x + 1 = R2$

Soit x^* le moment où $R1 = R2$: $x^* + 8 = 3x^* + 1$

$$\Leftrightarrow 2x^* = 7$$

$$\Leftrightarrow x^* = \frac{7}{2}$$

Si $x < x^* = \frac{7}{2}$, alors le meilleur estimateur est le 2.

Si $x > x^* = \frac{7}{2}$, alors le meilleur estimateur est le 1.

b) EST_3:

X2 \ X1	1	2
1	F	T
2	F	F

$X=11: 1T-2F$

$X=12: 4T-1F$

$X=21: 1T-4F$

$X=22: 1T-3F$

Risque empirique
non prédit T
 $1 \times 2 = 2$

$1 \times 1 = 1$

$1 \times 4 = 4$

$1 \times 3 = 3$

Risque empirique
non prédit F
 $2 \times 1 = 2 \Rightarrow F$ (arbitraire: le plus fréquent)

$2 \times 4 = 8 \Rightarrow T$

$2 \times 1 = 2 \Rightarrow F$

$2 \times 1 = 2 \Rightarrow F$

Son risque empirique est de $2 + 1 + 2 + 2 = 7 = R3$.

c) - il n'y a pas de cas d'égalité des fréquences

d)

$Y=T$			$Y=F$		
	1	2		1	2
X_1	5	2	X_1	3	7
X_2	2	5	X_2	6	4

e) $f(X=11)=F$

$$P(Y=T | X=11) \propto P(Y=T) P(X_1=1 | Y=T) P(X_2=1 | Y=T) = \frac{7}{17} \times \frac{5}{7} \times \frac{2}{7} = \frac{10}{119} \approx 0,084$$

$$P(Y=F | X=11) \propto P(Y=F) P(X_1=1 | Y=F) P(X_2=1 | Y=F) = \frac{10}{17} \times \frac{3}{10} \times \frac{6}{10} = \frac{18}{170} \approx 0,106$$

$f(X=12)=T$

$$P(Y=T | X=12) \propto P(Y=T) P(X_1=1 | Y=T) P(X_2=2 | Y=T) = \frac{7}{17} \times \frac{5}{7} \times \frac{5}{7} = \frac{25}{119} \approx 0,210$$

$$P(Y=F | X=12) \propto P(Y=F) P(X_1=1 | Y=F) P(X_2=2 | Y=F) = \frac{10}{17} \times \frac{3}{10} \times \frac{4}{10} = \frac{12}{170} \approx 0,071$$

$f(X=21)=F$

$$P(Y=T | X=21) \propto P(Y=T) P(X_1=2 | Y=T) P(X_2=1 | Y=T) = \frac{7}{17} \times \frac{2}{7} \times \frac{2}{7} = \frac{4}{119} \approx 0,034$$

$$P(Y=F | X=21) \propto P(Y=F) P(X_1=2 | Y=F) P(X_2=1 | Y=F) = \frac{10}{17} \times \frac{7}{10} \times \frac{6}{10} = \frac{42}{170} \approx 0,247$$

$f(X=22)=F$

$$P(Y=T | X=22) \propto P(Y=T) P(X_1=2 | Y=T) P(X_2=2 | Y=T) = \frac{7}{17} \times \frac{2}{7} \times \frac{5}{7} = \frac{10}{119} \approx 0,084$$

$$P(Y=F | X=22) \propto P(Y=F) P(X_1=2 | Y=F) P(X_2=2 | Y=F) = \frac{10}{17} \times \frac{7}{10} \times \frac{4}{10} = \frac{28}{170} \approx 0,165$$

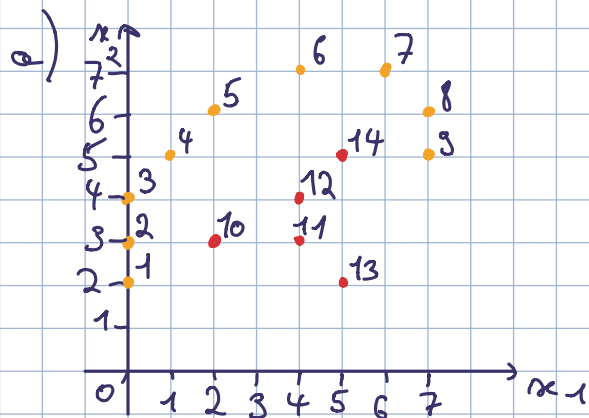
↳

EST-4:

$X_1 \backslash X_2$	1	2
1	F	T
2	F	F

C'est le même que EST-3.

Exercice 2

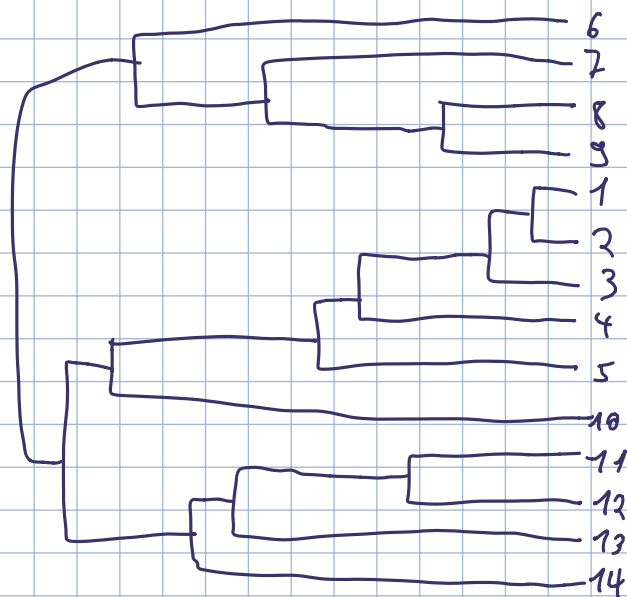


K-means : NON : les centres des deux composantes sont proches

Clustering Ascendant Hiérarchique : OUI : les points d'une même composante sont assez proches

b) Mesures utilisées :

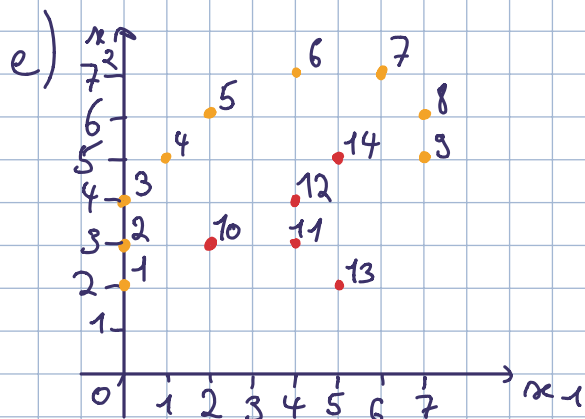
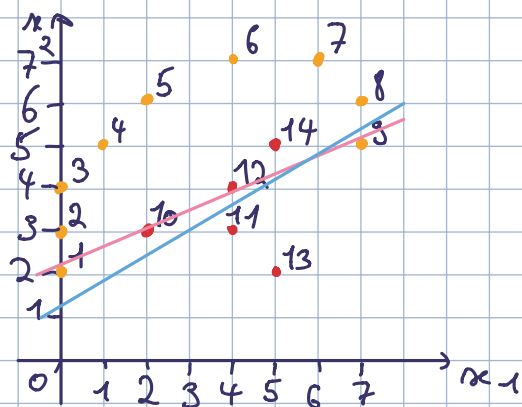
- distance entre 2 points : distance euclidienne
- distance avec un cluster : distance au point le plus proche



c) Non : ce n'est pas séparable par une droite

d) • Pénalité très élevée ; 1000 000 : le moins d'erreurs possibles

• Pénalité très faible ; 0,1 : erreurs peu coûteuses



Comme ça ressemble à un cercle, on cherche par centres les variables telle que le centre de l'intérieur soit en $(0,0)$:

Pour l'intérieur $\bar{x} = \frac{2+4+4+5+5}{5} = 4$ et $\bar{y} = \frac{3+3+4+2+5}{5} = \frac{17}{5}$

Ensuite, on prend les valeurs absolues, ce qui donne :

$$f(x, y) = |x - 4| + |y - \frac{17}{5}|$$

Pour chaque (x_i, y_i) , on le projète avec $(x_i, y_i, f(x_i, y_i))$.

(note : à des coefficients près)

Exercice 3

a) On pose l'hypothèse qu'elles sont indépendantes et une limite pour le valeur- p (généralement 0.05).

On calcule le valeur- p d'un test du χ^2 .

Si cette valeur est inférieure à la limite, on rejette l'hypothèse : elles sont dépendantes.

b) C'est la dimension VC. On peut en calculer une borne inférieure en calculant la dimension VC des sommets de polygones réguliers (qui est 7 pour un pentagone par exemple).

Pour le calculer exactement, on commence par prendre 2 points sur le cercle (on les choisit comme on veut) et on regarde s'ils sont réalisables (c'est-à-dire que toute manière de les associer est réalisable).

On continue ce procédé jusqu'à ce que ça ne soit plus réalisable.

c)