Nom : DER
Prénom : Vincent

FTML - Partiel du 16/07

(1)

Exercice 1 :

a). **Pour EST_1 :**

Matrice de confusion :

6	1
8	2

$$\Rightarrow \frac{1}{17} \times (1 \times 8 + x \times 1) = \frac{8+x}{17}$$

$\boxed{\text{Le risque empirique, pour EST\_1, est : } \frac{x+8}{17}}$. noté Rep 1

• **Pour EST_2 :**

Mat. de confusion :

4	3
1	9

$$\Rightarrow \frac{1}{17} \times (1 \times 1 + x \times 3) = \frac{3x+1}{17}$$

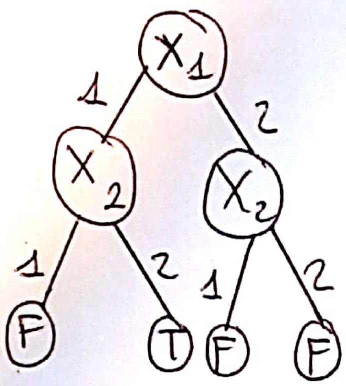
$\boxed{\text{Le risque empirique, pour EST\_2, est : } \frac{3x+1}{17}.}$, noté Rep 2

Comparaison : Soit $f : x \longmapsto x + 8 - 3x - 1$ ($\underline{\text{ie}}$: $f(x) = -2x + 7$)

x	$-\infty$		$7/2$		$+\infty$
$sg(f)$		$+$	0	$-$	

Ainsi : $\boxed{\text{Pour } x \in]-\infty ; \tfrac{7}{2}[\text{ ; Rep 1} > \text{Rep 2 et pour } x \in]\tfrac{7}{2} ; +\infty[\text{ : Rep1} < \text{Rep2}}$

b) Arbre de décision:



Ainsi: le risque empirique vaut: $\frac{1}{17} \times (3 \times 2 + 1 \times 1) = \frac{7}{17}$

c). Un arbre de décision nous a permis d'avoir des résultats optimaux...

d) Pour y = True:

	1	2
X_1	5	2
X_2	2	5

Pour y = False

	1	2
X_1	3	7
X_2	6	4

e)

- $X_1 = 1$ et $X_2 = 1$ (Δ)

$$\mathbb{P}(Y=T \mid X_1=1 \,\&\, X_2=1) = \frac{7}{17} \times \frac{5}{7} \times \frac{2}{7} = \frac{10}{119}$$

$$\mathbb{P}(Y=F \mid X_1=1 \,\&\, X_2=1) = \frac{10}{17} \times \frac{3}{10} \times \frac{6}{10} = \frac{9}{85}$$

- $X_1 = 1$ et $X_2 = 2$ $(\star)$

$$\mathbb{P}(Y=T \mid \star) = \frac{7}{17} \times \frac{5}{7} \times \frac{5}{7} = \frac{25}{119}$$

$$\mathbb{P}(Y=F \mid \star) = \frac{10}{17} \times \frac{3}{10} \times \frac{4}{10} = \frac{6}{85}$$

- $X_1 = 2$ et $X_2 = 1$ $(\$)$

$$\mathbb{P}(Y=T \mid \$) = \frac{7}{17} \times \frac{2}{7} \times \frac{2}{7} = \frac{4}{119}$$

$$\mathbb{P}(Y=F \mid \$) = \frac{10}{17} \times \frac{7}{10} \times \frac{6}{10} = \frac{21}{85}$$

- $X_1 = 2$ et $X_2 = 2$ (∇)

$$\mathbb{P}(Y=T \mid \nabla) = \frac{7}{17} \times \frac{2}{7} \times \frac{5}{7} = \frac{10}{19}$$

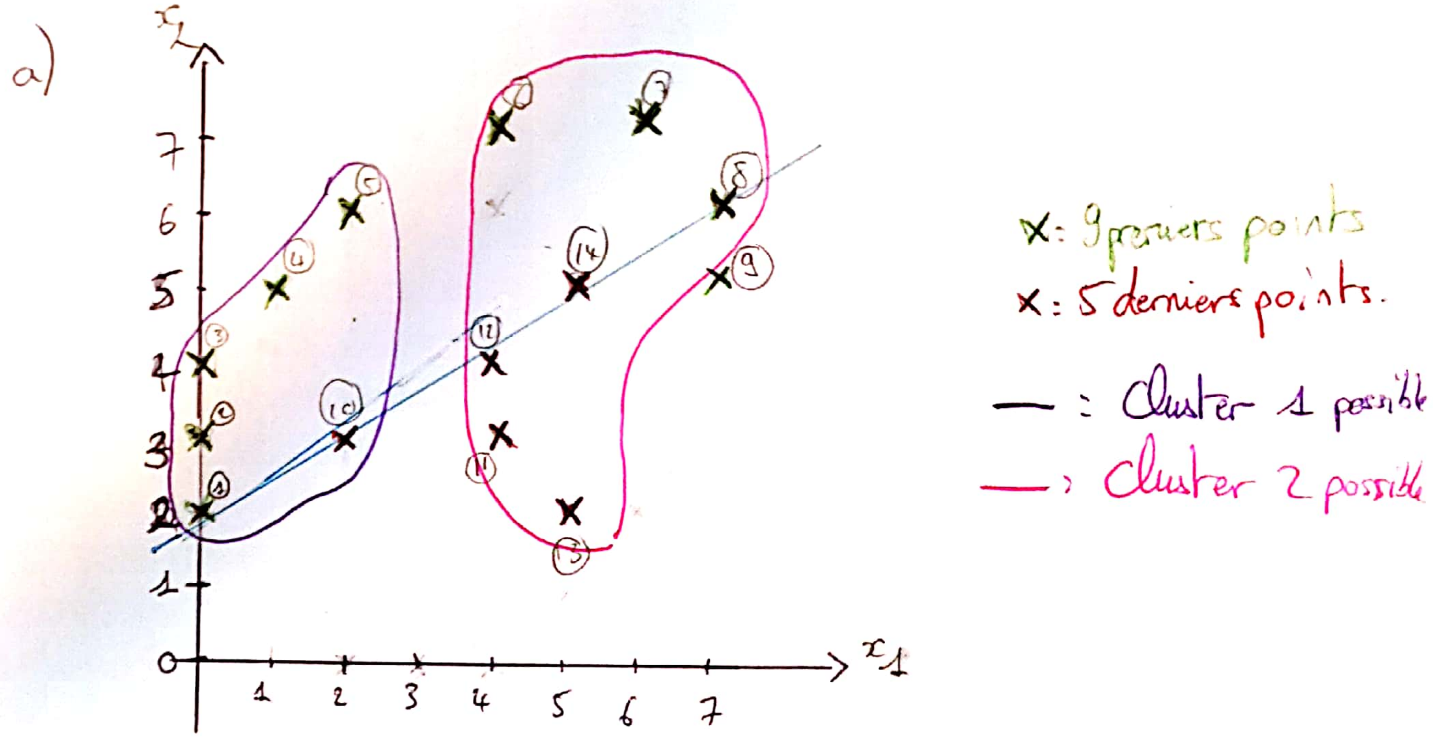
$$\mathbb{P}(Y=F \mid \nabla) = \frac{10}{17} \times \frac{7}{10} \times \frac{4}{10} = \frac{14}{85}$$

De ce fait, nous obtenons ce tableau résumant l'estimateur bayésien naïf optimal :

	1	2
1	F	T
2	F	F

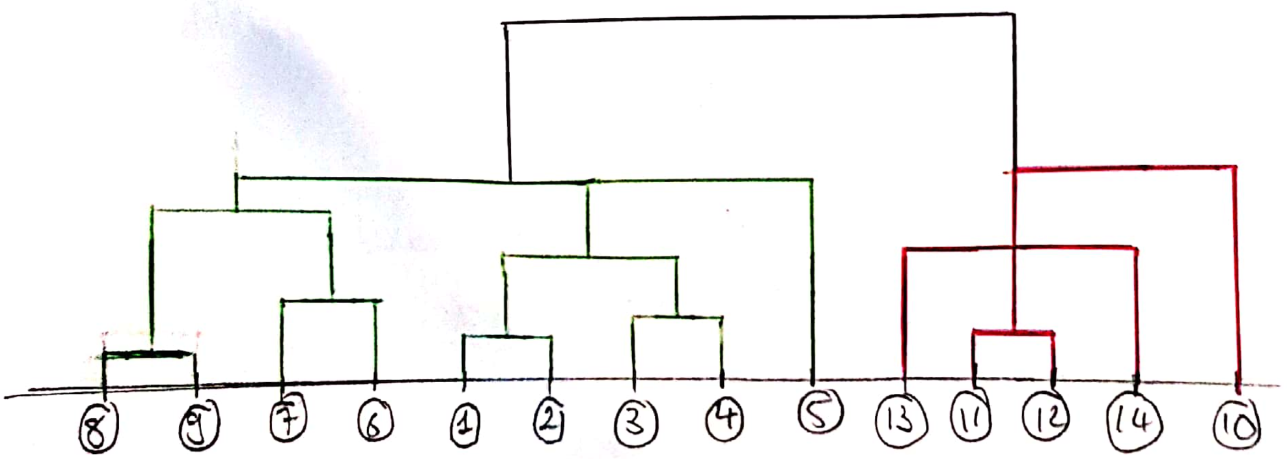
Exercice 2 :

a)



- On pourrait utiliser un kmeans ou bien une classification hiérarchique.

b) Dendogramme d'une classification hiérarchique ascendante.



c)

d) cf question a (voir ~~courbe~~ droite bleu).

e)

Exercice 3:

a) Pour vérifier l'indépendance des deux variables (noté A et B):

1) On calcule les effectifs croisés espérés:

$$E_{i,j} = \frac{1}{n} \times \sum_{i,j} card(A=i) \times card(B=j)$$

2) On calcule les effectifs croisés observés:

$$O_{i,j} = Card((A=i) \cap (B=j))$$

3) Enfin, on calcule l'écart entre les 2 variables. L'indépendance se traduit par une faible valeur de cet écart.

$$\varepsilon = \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$