ACHOUR SAMY

excercice 1:

$Y=F$ & $Y'=T \Rightarrow LF(Y,Y')=1$, $Y=T$ & $Y'=F \Rightarrow LF(Y,Y')=x$

a)

pour EST_1:

$$= \frac{1}{N} \sum_{i=1}^{17} (LF(Y_i, Y'_i))$$

$$= \frac{1}{N} \cdot (1 + x + \dots)$$

$$= \frac{1}{N} \cdot (8 + x) = \frac{8+x}{17}$$

pour EST_2:

$$= \frac{1}{N} \sum_{i=1}^{17} (LF(Y_i, Y'_i))$$

$$= \frac{1}{N} \cdot (1 + 3x) = \frac{1+3x}{17}$$

on peut dire que pour $x < \frac{9}{2}$, on obtiendra un risque empirique inférieur avec l'éstimateur 2.

pour $x > \frac{9}{2}$ $\rightarrow$ risque empirique inférieur avec l'éstimateur 1

$x = 2$

b) pas sûr d'avoir compris la question

Y	EST
T	T
F	F
T	T
F	F
T	T
F	F
F	F
T	T
F	F
F	F
T	T
F	F
T	T
F	F
F	F
F	F
T	T

→ risque empirique = 0

c)

d)

Y=T	1	2
x_1	5	2
x_2	2	5

Y=F	1	2
x_1	3	7
x_2	6	4

e)

$P(y=T) \times \prod P(X_i = k_i \mid x_1 = ? \; \& \; x_2 = ?)$

$\rightarrow = \frac{7}{17} \times P(x_1=1 \mid Y=T) \times P(x_2=1 \mid Y=T) = \frac{7}{17} \times \frac{5}{7} \times \frac{2}{7} = 0{,}0840$

$\rightarrow = \frac{7}{17} \times P(x_1=1 \mid Y=T) \times P(x_2=2 \mid Y=T) = \frac{7}{17} \times \frac{5}{7} \times \frac{5}{7} = 0{,}2100$

$\rightarrow = \frac{7}{17} \times P(x_1=2 \mid Y=T) \times P(x_2=1 \mid Y=T) = \frac{7}{17} \times \frac{2}{7} \times \frac{2}{7} = \cancel{0{,}0840} \; 0{,}0336$

$\rightarrow \frac{7}{17} \times P(x_1=2 \mid Y=T) \times P(x_2=2 \mid Y=T) = \frac{7}{17} \times \frac{2}{7} \times \frac{5}{7} = 0{,}0840$

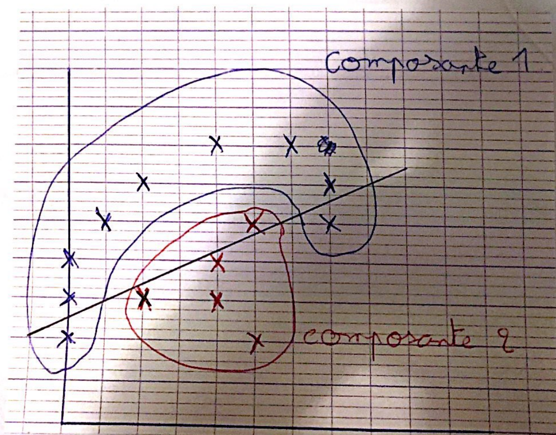
pour $P(y=F) \times \prod P(X_i = k_i \mid x_1 = ? \; \& \; x_2 = ?)$

$\rightarrow = \frac{10}{17} \times P(x_1=2 \mid Y=F) \times P(x_2=1 \mid Y=F) = \frac{10}{17} \times \frac{7}{10} \times \frac{6}{10} = 0{,}2470$

---

exercice 2

a)



pour distinguer ces deux composantes, on peut utiliser un Kmeans comme méthode non supervisée.

- spectral clustering
- Agglomerative clustering

b)

c) il n'est pas possible de separer les deux classes avec un SVM linéaire car il n'existe pas d'hyperplan pouvant separer les deux classes.

d) voir dessin question 1

e)

exercice 3:

a) afin de savoir si les deux variables concernées sont indépendantes, on utilise la methode du χ^2

b)

c) le risque peut être estimé alors que l'ambiguité ne nous permet pas d'estimer le risque qui lui est rattaché.