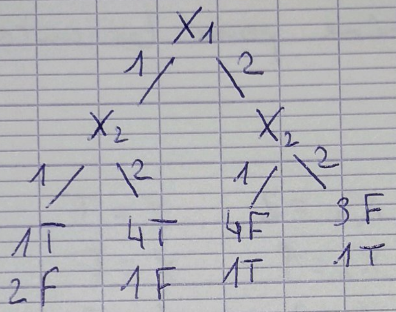
Machine Learning

Exercice 1

a) Le risque empirique de EST_1 est de $\frac{6+x}{17}$

EST_2 est de $\frac{1+3x}{17}$

b)



Estimateur

X_1	X_2	Ypred
1	1	F
1	2	T
2	1	F
2	2	F

c) Car nous avons assez de données pour pouvoir essayer sens.

d)

X_1	T	F
1	5	3
2	2	7

X_2	T	F
1	2	6
2	5	4

e) $P(Y=T) = 7 \qquad P(Y=F) = 10$

$$P(Y=T \mid X_1=1, X_2=1) = \frac{P(X_1=1 \mid Y=T) \times P(X_2=1 \mid Y=T) \times P(Y=T)}{P(X_1=1, X_2=1)}$$

$$= \frac{\frac{5}{7} \times \frac{2}{7} \times \frac{7}{13}}{\frac{3}{17}} = \frac{10}{21} = 0{,}476$$

Similairement nous avons :

$$P(Y = F \mid X_1 = 1, X_2 = 1) = \frac{17}{3} \times \frac{16}{170} = 0.6$$

Et au vu du fait que $0,6 > 0,476$ cela signifie que :

$$P(X = F \mid X_1 = 1, X_2 = 1) > P(X = T \mid X_1 = 1, X_2 = 1)$$

Donc $Y = F$ si $X_1 = 1$ et $X_2 = 1$

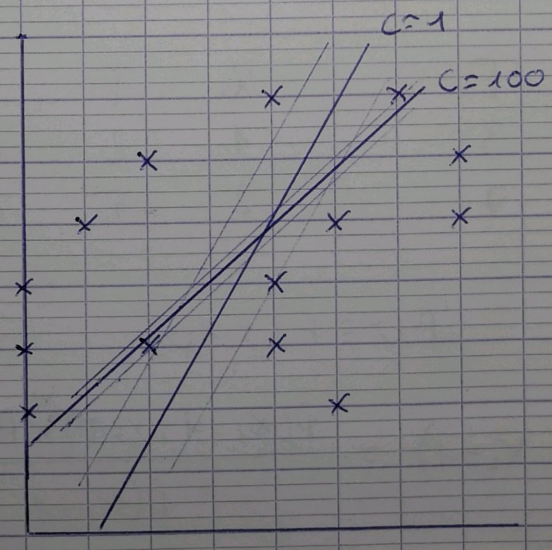
En faisant de même pour les autres valeurs nous obtenons :

$X_1 = 1$, $X_2 = 2$ $\quad Y = T$
$X_1 = 2$ $\quad X_2 = 1$ $\quad Y = F$
$X_1 = 2$ $\quad X_2 = 2$ $\quad Y = F$

Cet estimateur est donc le même que celui trouvé plus haut

Exercice 2

a)



Nous pourrions utiliser pour les distinguer les méthodes suivantes :

- Agglomerative Clustering

- DBScan

- OPTICS

Exercice 2

b)

c) Non il est impossible de les séparer avec un SVM linéaire

d) Cf graphique

e) Kernel polynomial

Exercice 3

a) Afin de savoir si deux variables sont indépendantes il suffit de faire un test statistique et d'examiner leurs probabilités individuelles. Si ces probabilités ne changent pas lorsque les evenements se rencontrent alors ces variables sont indépendantes

b)

c) Le risque rentre en compte quand il s'agit de probabilités qué l'on peut plus ou moins calculer. Alors que avec l'ambiguité nous n'en avons aucune idée.