FTML 2020

Exercice 1

a) Pour . Y_EST_1

Il y a – 8 faux positifs
– 1 faux négatifs

$$LF_{EST1} = \frac{8 \times 1 + 1 \times x}{17} = \boxed{\frac{8 + x}{17}}$$

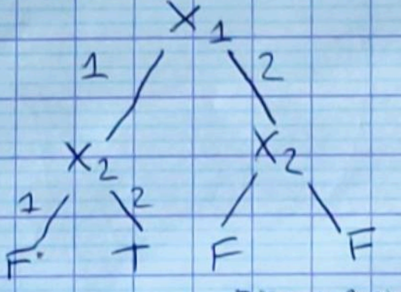
. Y_EST_2

Il y a – 1 faux positif
– 3 faux négatifs

$$LF_{EST2} = \frac{1 \times 1 + 3 \times x}{17} = \boxed{\frac{1 + 3x}{17}}$$

$$LF_{EST1} > LF_{EST2} \iff 8 + x > 1 + 3x$$
$$\iff \boxed{x < \frac{7}{2}}$$

b) On peut construire un arbre binaire de sorte à minimiser le risque empirique



Ce qui donne un risque empirique de $LF = \frac{1 \times 2 + 3 \times 1}{17} = \frac{5}{17}$

alors que pour $x = 2$,

$LF_{EST1} = \frac{10}{17}$ et $LF_{EST2} = \frac{7}{17}$

c) - Nous avons des informations sur les éventuelles corrélations entre les valeurs de x_1 et x_2, et les valeurs de Y. On peut donc avoir un raisonnement plus intelligent qu'un bayésien naïf.

d)

$Y = T$	$x = 1$	$x = 2$
x_1	5	2
x_2	2	5

$Y = F$	$x = 1$	$x = 2$
x_1	3	7
x_2	6	4

e) $P(Y = T \mid x_1 = 1 \text{ et } x_2 = 1) = \dfrac{P(x_1 = 1 \mid Y = T)\, P(x_2 = 1 \mid Y = T)\, P(Y = T)}{P(x_1 = 1 \text{ et } x_2 = 1)}$

$P(Y=T) = 7 \cancel{2}$

$$= \frac{\frac{5}{7} \times \frac{2}{7} \times \frac{7}{17}}{\frac{3}{17}} = \frac{10}{21} \simeq 0{,}48$$

$$P(Y = T \mid x_1 = 1 \text{ et } x_2 = 2) = \frac{\frac{5}{7} \times \frac{5}{7} \times \frac{7}{17}}{P(x_1 = 1 \text{ et } x_2 = 2)} = \frac{\frac{5}{7} \times \frac{5}{7} \times \frac{7}{17}}{\frac{5}{17}}$$

$$= \frac{5}{7} \simeq 0{,}71$$

$P(Y=T \mid x_1=2 \text{ et } x_2=1) = \dfrac{\frac{2}{7} \times \frac{2}{7} \times \frac{7}{17}}{\frac{5}{17}} = \dfrac{4}{35} \simeq 0{,}11$

$P(Y=T \mid x_1=2 \text{ et } x_2=2) = \dfrac{\frac{2}{7} \times \frac{5}{17} \times \frac{7}{17}}{\frac{4}{17}} = \dfrac{5}{14} \simeq 0{,}36$

On calcule les valeurs de $Y=F$ maintenant

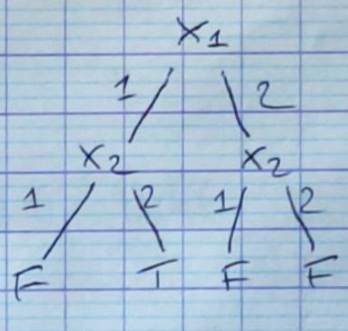
$P(Y=F \mid x_1=1 \text{ et } x_2=1) = \dfrac{\frac{3}{10} \times \frac{6}{10} \times \frac{10}{17}}{\frac{3}{17}} = \dfrac{3}{5} \simeq 0{,}6$

$\Rightarrow P(Y=F)$ est meilleur pour $x_1=1$, $x_2=1$

On compare les valeurs pour les autres valeurs de x_1, x_2 :

- $x_1=1$, $x_2=2 \Rightarrow Y=T$
- $x_1=2$, $x_2=1 \Rightarrow Y=F$
- $x_1=2$, $x_2=2 \Rightarrow Y=F$

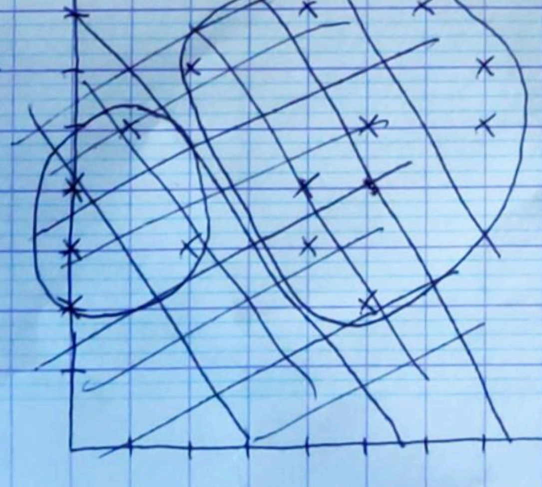
Cela donne donc l'estimateur suivant.

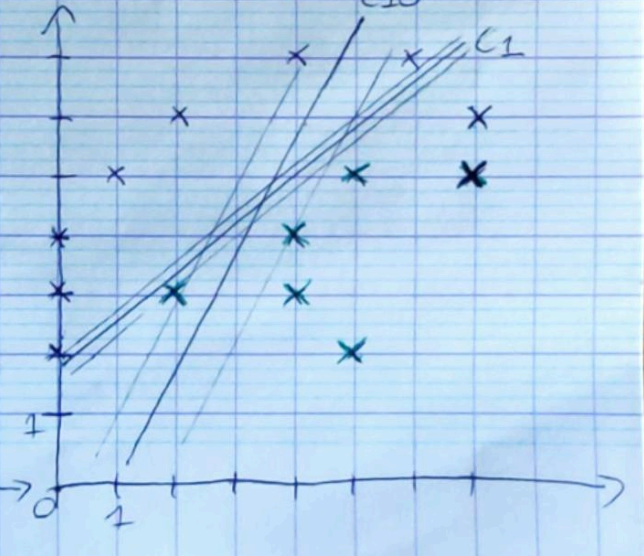


C'est donc le même estimateur que plus haut.

__Exercice 2__

a)





Une méthode non supervisée pouvant être efficace est l'agglomérative clustering, OPTICS, DBSCAN.

c) Il est possible de séparer les composantes (plus ou moins) avec un SVM linéaire, avec plus ou moins d'erreur selon le coefficient de pénalisation. Ce n'est cependant pas efficace et pas approprié

d) (voir graphe en a))

e) Une kernelisation polynomiale permettrait de séparer les deux classes

<u>Exercice 3</u>

a) Pour cela, on ~~calcul~~ calcule les effectifs croisés espérés, les effectifs croisés observés et l'écart relatif entre les 2.

c) La différence entre risque et ambiguïté est que le risque, on connait sa probabilité, contrairement à l'ambiguïté.