

Nombre d'intercalaires : _____

Nom : BHAVSAR

Prénom : Ferdinand

Matière : FTML

Promo : 2021

Note : _____ Appréciation : _____

Groupe : _____ UID : _____

Nom de l'enseignant : _____ Visa de l'enseignant : _____

Exercice 1

a)

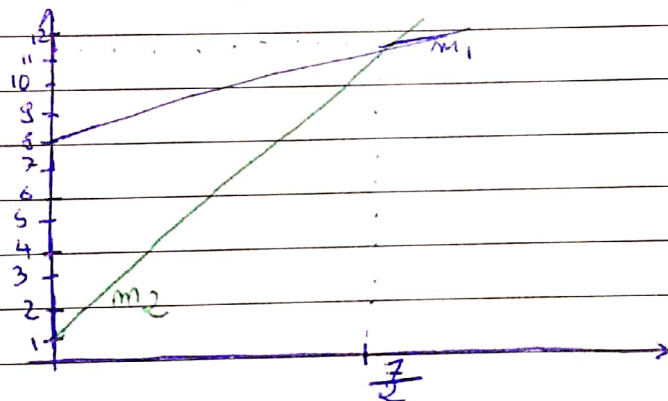
	FP	FN
Estimateur 1	8	1
Estimateur 2	1	3

Risque empirique :

$$m_1 = 8 + x$$

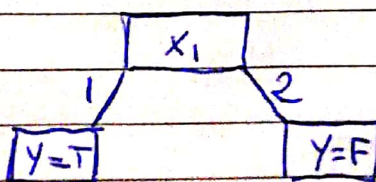
$$m_2 = 1 + 3x$$

Donc $\begin{cases} m_1 = m_2 ? \\ \Leftrightarrow 8 + x = 1 + 3x \\ \Leftrightarrow x = \frac{7}{2} \end{cases}$



m_2 a un risque moins grand pour $x < \frac{7}{2}$, m_1 a un risque moins grand pour $x > \frac{7}{2}$

b) Un modèle simple qui pourrait réduire le risque empirique est un dummy classifiée qui ressort toujours T car le coût de faux-positif est moins fort.
Soit un arbre binaire tel que :



Donne un risque de seulement 7

c)

d)	$x_1 = 1$	$x_1 = 2$		$x_2 = 1$	$x_2 = 2$
$y = T$	5	2	$y = T$	2	5
$y = F$	3	7	$y = F$	6	4

e) $P(Y=T/X) = \frac{P(X_1=1, Y=T)}{P(X_1=1, Y=T) + P(X_1=2, Y=T)}$

	$x_1 = 1$	$x_1 = 2$
$x_2 = 1$	$\frac{5}{7} \cdot \frac{2}{7} \cdot \frac{7}{17}$ $\frac{8}{17} \cdot \frac{8}{17}$	$\frac{2}{7} \cdot \frac{2}{7} \cdot \frac{7}{17}$ $\frac{9}{17} \cdot \frac{8}{17}$
$x_2 = 2$	$\frac{4}{7} \cdot \frac{5}{7} \cdot \frac{7}{17}$ $\frac{8}{17} \cdot \frac{9}{17}$	$\frac{7}{17} \cdot \frac{4}{17} \cdot \frac{8}{17}$ $\frac{9}{17} \cdot \frac{9}{17}$

$P(Y=F/X)$

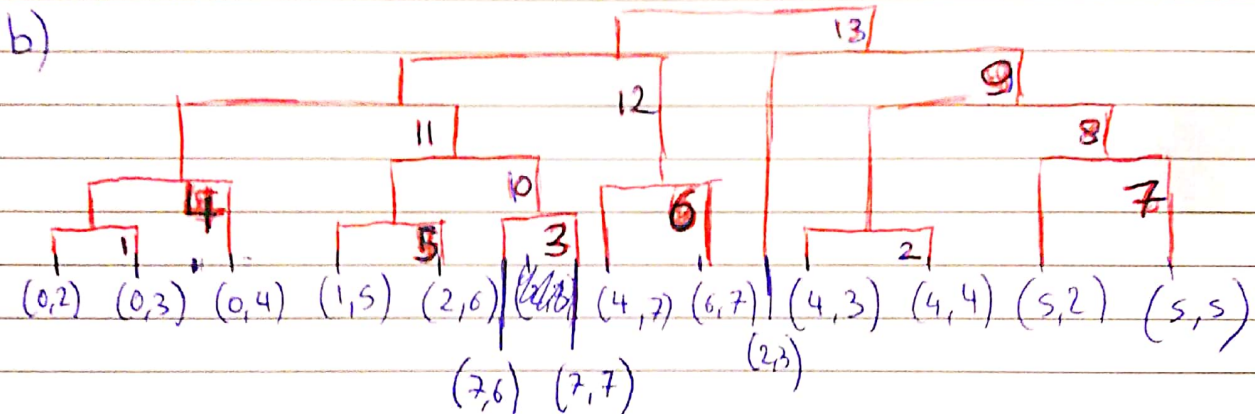
$\frac{\frac{5}{17} \cdot \frac{5}{17} \cdot \frac{7}{17}}{\frac{8}{17} \cdot \frac{9}{17}}$

	$x_1 = 1$	$x_1 = 2$
$x_2 = 1$	$\frac{3}{10} \cdot \frac{6}{10} \cdot \frac{10}{17}$ $\frac{9}{17} \cdot \frac{8}{17}$	$\frac{3}{10} \cdot \frac{6}{10} \cdot \frac{10}{17}$ $\frac{9}{17} \cdot \frac{8}{17}$
$x_2 = 2$	$\frac{5}{10} \cdot \frac{4}{10} \cdot \frac{10}{17}$ $\frac{8}{17} \cdot \frac{9}{17}$	$\frac{3}{10} \cdot \frac{4}{10} \cdot \frac{10}{17}$ $\frac{9}{17} \cdot \frac{9}{17}$

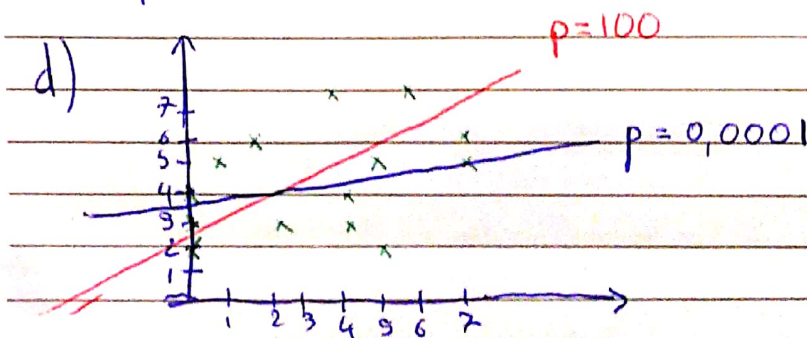
Exercice 2



Les méthodes supervisees possibles sont le k -nearest neighbors, le k -means et le clustering hiérarchique agglomératif. Ce dernier semble le plus approprié car la forme du jeu de données peut entraîner certains points $(5,5)$ à être mal classés.



c) Il n'est pas possible de séparer parfaitement cet ensemble par un SVM linéaire, car il n'est pas linéairement séparable.



e) Une fonction de kernelisation possible serait: $x_3 = x_1^2 + x_2^2$

Exercice 3

1)

2) Pour trouver le nombre maximum de points pulvérisables à la surface d'une sphère il faudrait d'abord trouver la borne inférieure, pour cela il faudrait :

- Montrer que $1, 2, 3, \dots, n$ points sont pulvérisables en montrant qu'ils sont pulvérisables.
- Ensuite montrer qu'on ne peut pas pulvériser à $n+1$
- On a notre borne inférieure

Par la suite on montre

Ensuite on montre que c'est aussi notre borne supérieure,

3) Le risque est la l'erreur liée à la donnée, l'erreur naturelle, alors que l'ambiguïté ~~est~~ est un biais ~~à~~ créé lorsque l'on fait ~~des~~ tirer des conclusions sur le modèle, par exemple pour réduire le risque.