BOSSUT
Geoffrey
bossut.geoffrey

FTML - Partiel 16/07/2020

Exercice 1

(a)

On a : EST_1 :

Pred Y / Y' (réalité)	T	F
T	6	1
F	8	2

EST_2

Y Pred / Y' Réalité	T	F
T	4	3
F	1	9

$$LF(yest_i, y_i') = \begin{cases} \text{si } yest_i > y_i' \text{ alors } 1 \\ \text{sinon si } yest_i < y_i' \text{ alors } x \\ \text{sinon } 0 \end{cases}$$

$$ERM(est) = \frac{1}{N} \times \sum_{i \in N} LF(yest_i, y_i')$$

d'où $ERM(EST\_1) = (8 \times 1 + 1 \times x) \times \frac{1}{17} = \frac{x+8}{17}$

et $ERM(EST\_2) = (1 \times 1 + 3 \times x) \times \frac{1}{17} = \frac{3x+1}{17}$

EST_1 est avantagé si

$x + 8 < 3x + 1$

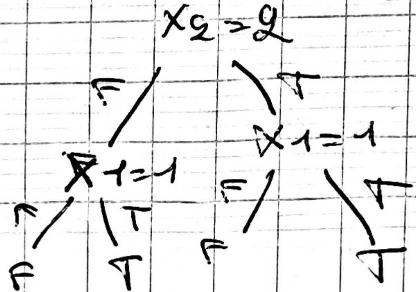
$8 - 1 < 2x$

$\frac{7}{2} < x$

On a $x = 2$

(b) on veut le moins de faux négatifs possibles (le moins de prédictions F pour T attendu)

On a l'arbre de décision suivant :



avec les effectifs :

2	pour 2, 2 et F
1	pour 1, 2 et F
2	1, 1 et F
3	2, 1 et F

3 pour	1, 2 et T
1	1, 1 et T
1	2, 1 et T
1	2, 2 et T

(c) L'hypothèse de non dépendance de X_1 et X_2 est forte, de plus on peut utiliser un estimateur probabiliste qui traite X_1 et X_2 dépendamment (comme ~~[illegible]~~ par $X_1 = 1$ et $X_2 = 2$) par exemple on valeurs vu en (b) pour minimiser l'erreur si $x = 2$.

(d)

X_1	T	F
1	5	3
2	2	7

X_2	Y	F
1	2	6
2	5	4

(e) L'estimateur Bayesien naïf estime les probabilités par les fréquences, mais surtout il traite indépendamment les variables les unes des autres, ainsi :

$$P(Y=T \mid (X_1 = a \;\&\&\; X_2 = b)) = \frac{P(Y=T) \times P(X_1 = a \mid Y=T) \times P(X_2 = b \mid Y = T)}{P(X_1 = a) \times P(X_2 = b)}$$

et de même pour $P(Y=F)$, on a donc :

$$P(Y=T \mid (X_1 = 1 \;\&\&\; X_2 = 1)) = \frac{\overbrace{\frac{7}{17} \times \frac{5}{7} \times \frac{2}{7}}^{\frac{10}{119}}}{P(X_1 = a) \times P(X_2 = b)}$$ on veut comparer

à $$P(Y=F \mid (X_1 = 1 \;\&\&\; X_2 = 1)) = \frac{\overbrace{\frac{10}{17} \times \frac{3}{10} \times \frac{6}{10}}^{\frac{18}{170}}}{P(X_1 = a) \times P(X_2 = b)}$$ d'où :

$$P(Y=F \mid (X_1 = 1 \;\&\&\; X_2 = 1)) > P(Y=T \mid \ldots)$$

d'où l'estimateur bayesien naïf prédira F pour l'observation $X_1 = 1 \;\&\&\; X_2 = 1$

~~Y=T x2 2 5 / 1 2 x1 valeurs 5 1~~

$Y=T$ \ X_2	valeurs	2 (1)	5 (2)
X_1			
5	1	$\frac{5}{7} \times \frac{2}{7} \times \frac{7}{17}$	$\frac{5}{7} \times \frac{5}{7} \times \frac{7}{17}$
2	2	$\frac{2}{7} \times \frac{2}{7} \times \frac{7}{17}$	$\frac{2}{7} \times \frac{5}{7} \times \frac{7}{17}$

Suite 1e

Y' = F	x_2	6	4
X1	valeurs	1	2
3	1	$\frac{3}{10} \times \frac{6}{10} \times \frac{10}{17}$	$\frac{3}{10} \times \frac{4}{10} \times \frac{10}{17}$
7	2	$\frac{7}{10} \times \frac{6}{10} \times \frac{10}{17}$	$\frac{7}{10} \times \frac{4}{10} \times \frac{10}{17}$

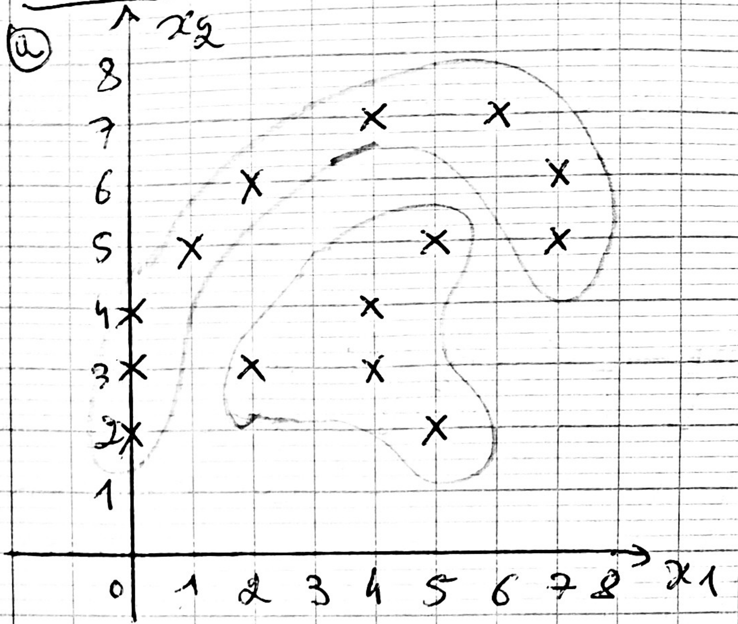
estimations Bayésien naïf :

Soit

x_1 \ x_2	1	2
1	Ypred = F	Ypred = T
2	Ypred = F	Ypred = F

Exercice 2

ⓐ



Classifieurs non supervisés :

- kNN
- Decision Tree
- ...

ⓑ On peut utiliser l'algorithme d'agglomération distance ou kmeans (à noter que kmeans dépend de l'initialisation et n'est pas déterministe).

ⓒ En l'état actuel non sauf si [redacted]
l'on ajoute un axe qui est une combinaison linéaire des axes
X_1 et X_2.

~~On peut également effectuer une ACP.~~

ⓓ

ⓔ

Ex3

(a) Test du chi2 permet de tester l'indépendance de deux variables

(b)

(c) ~~[illegible]~~ risque objectif : probabilités objectives connues

- ambiguïté : probabilités subjectives inconnues (on n'a aucune idée de quelles sont les probabilités)