

# FTML

Ex 1

X <sub>1</sub>	X <sub>2</sub>	Y	Y-EST-1	Y-EST-2
1	2	T	T	T
2	2	F	T	F
1	1	T	F	F
1	2	F	T	T
1	2	T	T	T
2	2	F	T	F
2	1	F	T	F
2	2	T	T	T
2	1	F	T	F
2	1	F	T	F
1	2	T	T	T
1	1	F	F	F
2	1	T	T	F
1	1	F	F	F
2	2	F	T	F
2	1	F	T	F
1	2	T	T	T

1 de pénalisation  
 x Faux positif  
 Faux négatif

a)

		Donnée		
		Positif	Négatif	
Prédit	Positif	6	8	EST-1
	Négatif	4	1	EST-2
Prédit	Positif	1	2	EST-1
	Négatif	3	9	EST-2

Risque

$$EST_1 = 1x + 8$$

$$EST_2 = 3x + 1$$



$$1x + 8 - (3x + 1) = 0$$

$$2x = 7$$

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

Si  $x < 3.5$  alors l'estimateur 1 a un coût moins élevé que EST-2.

Si  $x > 3.5$  alors l'estimateur EST-2 aura le coût le moins élevé.

b)  $x = 2$

Estimateur - Custom

Risque EST\_CUSTOM = 0

T  
F  
T  
F  
T  
F  
F  
T  
F  
F  
T  
F  
T  
F  
F  
F  
T

c) Ici, il n'y a pas de forte indépendance les hypothèses,

Il n'y a pas de relation entre  $X$  et  $Y$  exprimée sous forme de probabilité conditionnelle.



d)

$Y = \text{True}$	$X_1 = 1$	$X_2 = 2$
$X_1$	5	2
$X_2$	2	5

$Y = \text{False}$	$X_1 = 1$	$X_2 = 2$
$X_1$	3	7
$X_2$	1	1

e)  $P(Y=T) = 7/17$   
 $P(Y=F) = 10/17$

$$P(Y=T | X_1=1, X_2=1) = \frac{5/7 \times 2/7 \times 7/17}{3/17} = \frac{17}{3} \times \frac{70}{833}$$

$$P(Y=F | X_1=1, X_2=4) = \frac{\frac{3 \times 6}{100} \times \frac{10}{17}}{\frac{17}{17}} = \frac{17}{3} \times \frac{18}{170}$$

$$P(Y=T | X_1=1, X_2=2)$$

$$= \frac{5/7 \times 2/7 \times 7/17}{3/17} = \frac{17}{3} \times \frac{70}{833}$$

Exercice 3

Par le temps de fin...

a. Tableau d'effectif croisé, 2 variables continues indépendantes?

Pour vérifier l'indépendance de deux variables à l'aide d'un tableau croisé d'effectif on peut utiliser le test du  $\chi^2_2$ .

Pour que le test du  $\chi^2_2$  soit pertinent il faut que le jeu de données soit suffisamment grand (environ 20 obs./total).

S: la p-value résultante du  $\chi^2_2$  est inférieure 0.01 une relation de dépendance significative peut être déterminée.

b. Pour trouver le nombre max. de points pluvénisables à la surface d'une sphère je procéderai de la manière suivante:

Dessiner un cercle en 2D sur ma feuille et trouver le maximum de points pluvénisables graphiquement. Une fois que j'obtiens ce maximum je passe en 3D.



supérieure en 3D sur une sphère et je continue à partir du max trouvé précédemment à pulvériser mes points. Lorsque je peux pulvériser  $m$  points et qu'à  $m+1$  c'est impossible je sais que  $m$  est le nombre max. de points pulvérisables.

c) Diff. entre risque et ambiguïté?

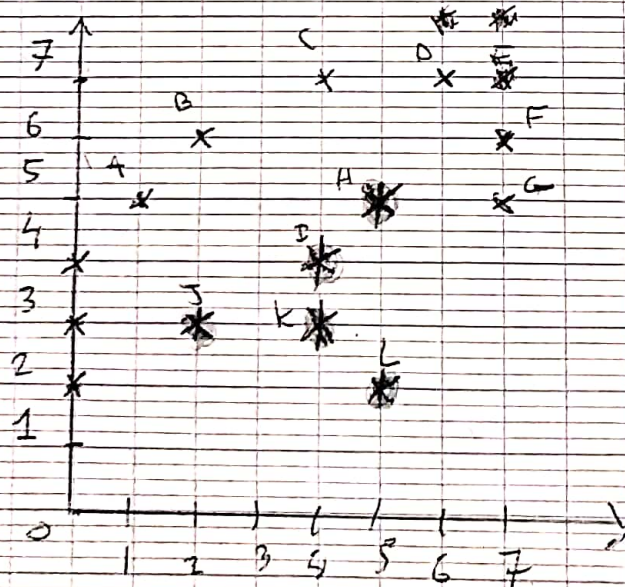
L'ambiguïté c'est lorsque je n'ai aucune idée de la probabilité d'un événement peut prendre.

Tandis que dans le risque on connaît les probabilités comme lorsqu'on lance un dé, une pièce etc..

## Exercice 2

$x_1$	0	0	0	1	2	4	6	7	7	2	4	4	5	5
$x_2$	2	3	4	5	6	7	7	6	5	3	3	4	2	3

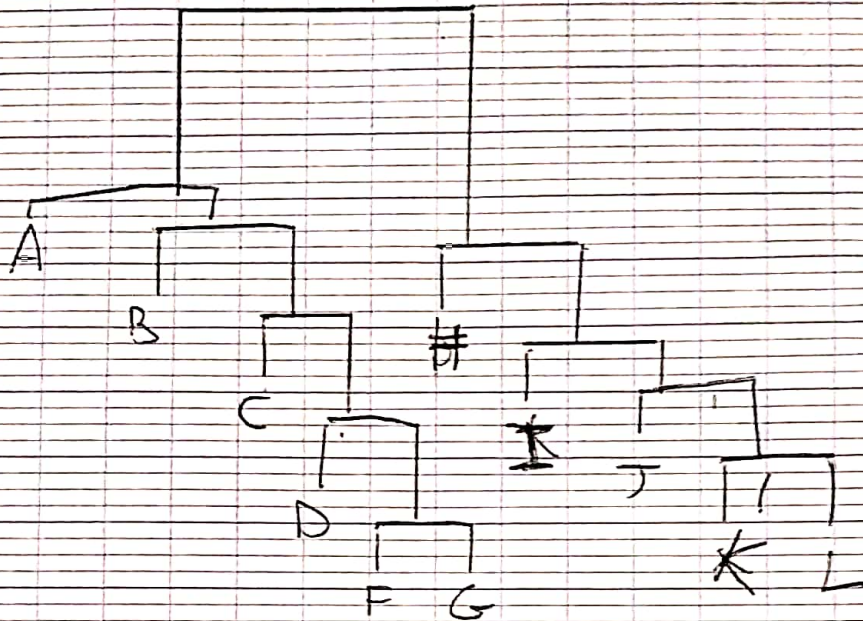
a)



Un agglomérative (clustering) devrait parvenir à les distinguer  
DBSCAN et spectral clustering devraient aussi fonctionner



b)



La fonction de coût étant une distance  $d(A, B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$

c) Il n'est pas possible de les séparer avec un SVM linéaire parfaitement.

d)

$$v(x_1, x_2) = \begin{cases} 1 & \text{si } x_1 \neq x_2 \\ 0 & \text{si } x_1 = x_2 \end{cases}$$