

EXAM FTML 16/07/2020

Exercice 1

a. Matrices de confusion :

EST_1 :

	T	F
T	6	1
F	8	2

EST_2 :

	T	F
T	4	3
F	1	9

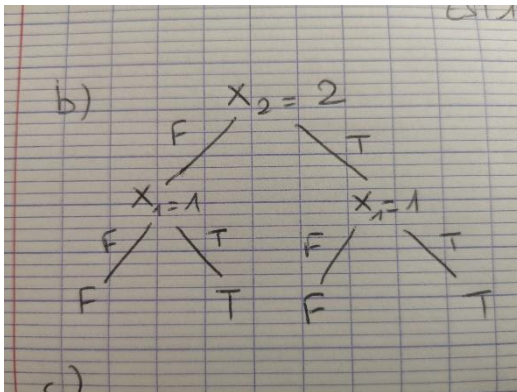
$$\text{Risque empirique : } ERM(EST1) = \frac{1}{17}(8 * 1 + 1 * x) = \frac{8+x}{17}$$

$$ERM(EST2) = \frac{1}{17}(1 * 1 + 3 * x) = \frac{1+3x}{17}$$

Donc $EST1 > EST2$ si $x < \frac{7}{2}$

$EST1 < EST2$ si $x \geq \frac{7}{2}$

b.



c. En optimisant les faux négatifs avec un arbre de décision, on peut obtenir un ERM fiable.
Il aurait fallu se placer dans un bayésien si elles étaient alignées dans la majorité des cas.

d.

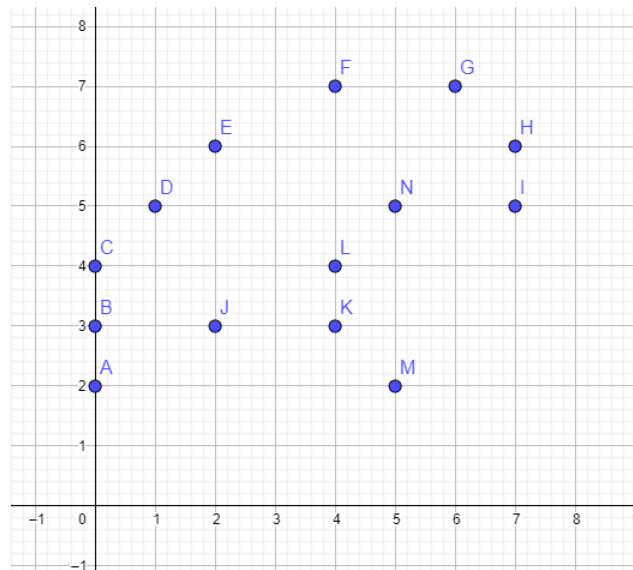
X1	T	F
1	5	3
2	2	7

X2	T	F
1	2	6
2	5	4

e.

X1/X2	1	2
1	Ypred = F	Ypred = T
2	Ypred = F	Ypred = F

Exercice 2



- Toutes les méthodes de classification permettraient de les séparer, en particulier SpectralClustering, AgglomerativeClustering, DBSCAN, SVM avec kernel rbf, DecisionTree, KNN.
-
- Il n'est pas possible de les séparer parfaitement avec un SVM linéaire.
-
-

Exercice 3

- On peut utiliser le test du χ^2 .
- Il faut trouver la dimension de Vapnik-Chervonenkis.
- Le risque implique une probabilité connue ou devinable alors que l'ambiguïté implique des probabilités totalement inconnues.