

Y-EST-1

	Réel	T	F
Préd	T	6	1
	F	8	2

Y-EST-2

	Réel	T	F
Préd	T	4	3
	F	1	9

$$\begin{aligned} \text{a) est 1: } LF(Y', Y) &= (8 \times 1 + 1 \times x) / 17 \\ &= \frac{8+x}{17} \end{aligned}$$

$$\begin{aligned} \text{est 2: } LF(Y', Y) &= (1 \times 1 + 3 \times x) / 17 \\ &= \frac{1+3x}{17} \end{aligned}$$

$$\begin{aligned} \text{est 2: } LF(Y', Y) &= (1 \times 1 + 3 \times x) / 17 \\ &= \frac{1+3x}{17} \end{aligned}$$

$$\frac{8+x}{17} = \frac{1+3x}{17}$$

$$\Leftrightarrow 8 - 1 = 2x$$

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

Donc pour $x > \frac{7}{2}$

Risque empirique(est 1) > Risque empirique(est 2)

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

Risque empirique(est 1) < Risque empirique(est 2)

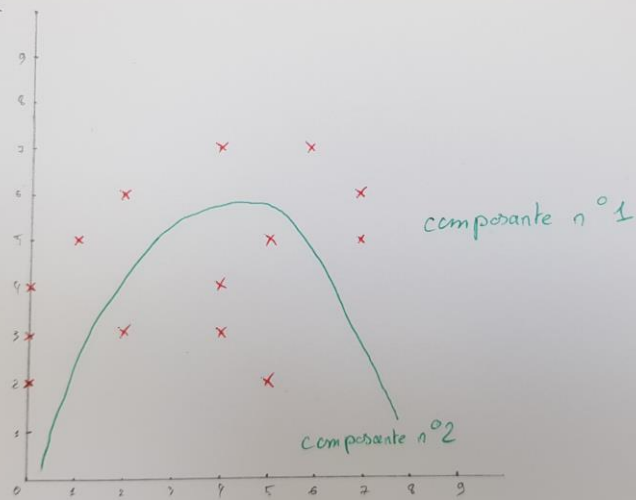
b) Un estimateur qui renvoie tout le temps VRAI
Son Risque Empirique maximal ~~est de 1~~ ~~sera de la taille~~
~~de l'échantillon (17 ici)~~ sera de 1.

~~de l'échantillon (17 ici)~~ sera de 1.

- c) → Entre X_1 et X_2 , il y a au max 4 combinaisons possibles, ce qui n'est pas optimal pour un bayésien naïf.
- De plus, les sorties (valeurs de Y) en fonction de la combinaison $X_1 X_2$ sont trop diverses (pour une même combinaison)

EXERCICE 2

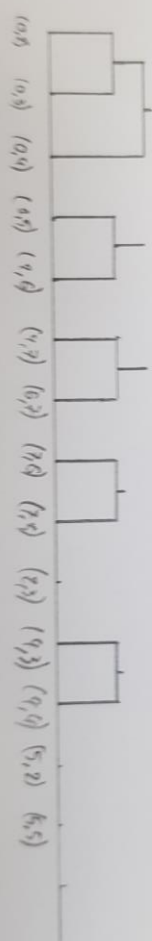
a)



Méthodes non supervisées pour les distinguer :

~~- K-Means~~
~~- Gaussian Mixture Models~~
~~- Expectation-Maximization~~

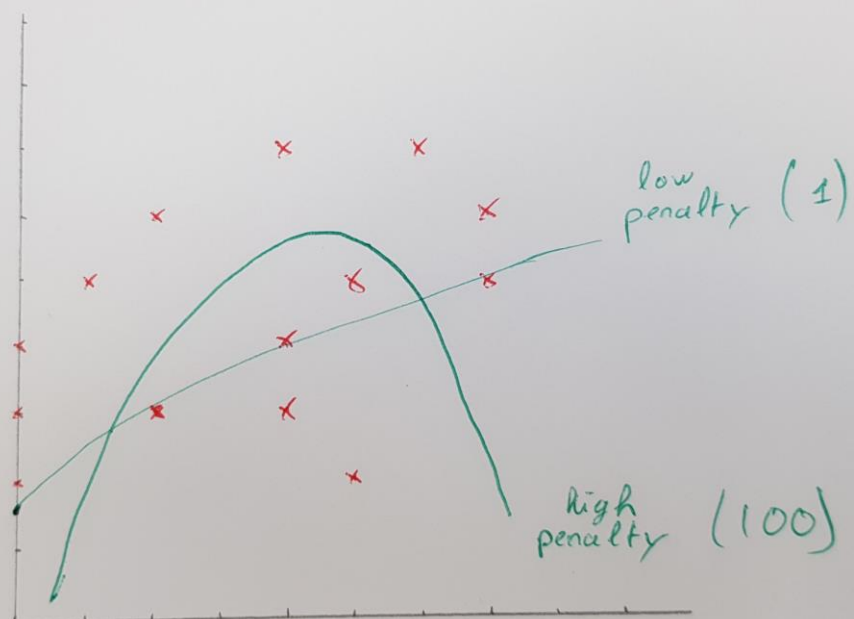
- Spectral Clustering
- Agglomerative Clustering
- Ward
- DBSCAN



- DBSCAN

- c) Il n'est pas possible de séparer parfaitement les deux classes avec une droite. Des points vont forcément se retrouver du mauvais côté.
un SVM linéaire ne peut donc pas séparer ces points

D)



c) Le Risque est basé sur des informations ~~véridi~~
Il peut être identifiable/calculé.

L'ambiguïté est basé sur le manque d'information