

Exercice 1.

(1)

a) Pour l'estimateur 1: $\frac{1}{17} (1 \times 8 + x)$

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

Pour l'estimateur 2: $\frac{1}{17} (1 \times 1 + 3x)$

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

Si $x \leq 3$ l'estimateur 2 est meilleur
Sinon l'estimateur 1 est meilleur.

b) Etant donné les données que l'on a on peut créer un estimateur qui retourne True dans le cas $X_1 = 1$ et $X_2 = 2$, et qui retourne False sinon. On a alors une erreur de $\frac{7}{17}$.

c) On a pas besoin d'un bayésien naïf car :

- 1 - On a pas de cas ambigus où les fréquences sont égales
- 2 - On a peu de choix d'observations.

d)

$Y=T$	1	2
X_1	5	2
X_2	2	5

$Y=F$	1	2
X_1	3	7
X_2	6	4

e)

$$\begin{aligned}
 P(Y=T | X=1,1) &= P(Y=T) \times P(X_1=1 | Y=T) \times P(X_2=1 | Y=T) \\
 &= \frac{\cancel{7}}{17} \times \frac{5}{\cancel{7}} \times \frac{2}{7} \\
 &= \frac{10}{17 \times 7}
 \end{aligned}$$

$$\begin{aligned}
 P(Y=F | X=1,1) &= P(Y=F) \times P(X_1=1 | Y=F) \times P(X_2=1 | Y=F) \\
 &= \frac{\cancel{10}}{17} \times \frac{3}{\cancel{10}} \times \frac{6^3}{10^5} \\
 &= \frac{9}{17 \times 5}
 \end{aligned}$$

$$\Rightarrow \frac{50}{17 \times 7 \times 5} < \frac{63}{17 \times 5 \times 7} \rightarrow \text{Done on patient F.}$$

$$\begin{aligned}
 P(Y=T | X=1,2) &= P(Y=T) \times P(X_1=1 | Y=T) \times P(X_2=2 | Y=T) \\
 &= \frac{\cancel{7}}{17} \times \frac{5}{\cancel{7}} \times \frac{5}{7} \\
 &= \frac{25}{17 \times 7}
 \end{aligned}$$

$$\begin{aligned}
 P(Y=F | X=1,2) &= P(Y=F) \times P(X_1=1 | Y=F) \times P(X_2=2 | Y=F) \\
 &= \frac{\cancel{10}}{17} \times \frac{3}{\cancel{10}} \times \frac{4^2}{10^5} \\
 &= \frac{6}{17 \times 5}
 \end{aligned}$$

$$\Rightarrow \frac{4 \times 100 \times 125}{17 \times 7 \times 5} > \frac{42}{17 \times 5 \times 7} \rightarrow \text{Done on patient T}$$

$$\begin{aligned}
 P(Y=T | X=21) &= P(Y=T) \times P(X_1=2 | Y=T) \times P(X_2=1 | Y=T) \quad (2) \\
 &= \frac{\cancel{7}}{17} \times \frac{2}{\cancel{7}} \times \frac{2}{7} \\
 &= \frac{4}{17 \times 7}
 \end{aligned}$$

$$\begin{aligned}
 P(Y=F | X=21) &= P(Y=F) \times P(X_1=2 | Y=F) \times P(X_2=1 | Y=F) \\
 &= \frac{\cancel{10}}{17} \times \frac{7}{\cancel{10}} \times \frac{\cancel{6}3}{\cancel{10}5} \\
 &= \frac{21}{17 \times 5}
 \end{aligned}$$

$$\Rightarrow \frac{20}{17 \times 5 \times 7} < \frac{147}{17 \times 5 \times 7} \rightarrow \text{Donc on retient } F.$$

$$\begin{aligned}
 P(Y=T | X=22) &= P(Y=T) \times P(X_1=2 | Y=T) \times P(X_2=2 | Y=T) \\
 &= \frac{\cancel{7}}{17} \times \frac{2}{\cancel{7}} \times \frac{5}{7} \\
 &= \frac{10}{17 \times 7}
 \end{aligned}$$

$$\begin{aligned}
 P(Y=F | X=22) &= P(Y=F) \times P(X_1=2 | Y=F) \times P(X_2=2 | Y=F) \\
 &= \frac{\cancel{10}}{17} \times \frac{7}{\cancel{10}} \times \frac{\cancel{4}2}{\cancel{10}5} \\
 &= \frac{14}{17 \times 5}
 \end{aligned}$$

$$\Rightarrow \frac{50}{17 \times 7 \times 5} < \frac{98}{17 \times 5 \times 7} \rightarrow \text{Donc on retient } F$$

Au final on a

X_1	X_2	Y
1	1	F
1	2	T
2	1	F
2	2	F

Ce qui nous donne la
 \rightarrow même chose qu'à la
 question b)

Exercice 2

a) Cf extrait excel

Ici on a plutôt envie de séparer un "noyau" d'un cercle.
D'après les schémas proposés par sklearn des différentes méthodes de clustering, les plus adaptées semblent être :

- Spectral Clustering
- DBSCAN
- Agglomerative Clustering
- OPTICS

b)

c) Non il n'est pas possible de séparer parfaitement ces deux classes avec un kernel linéaire.

d)

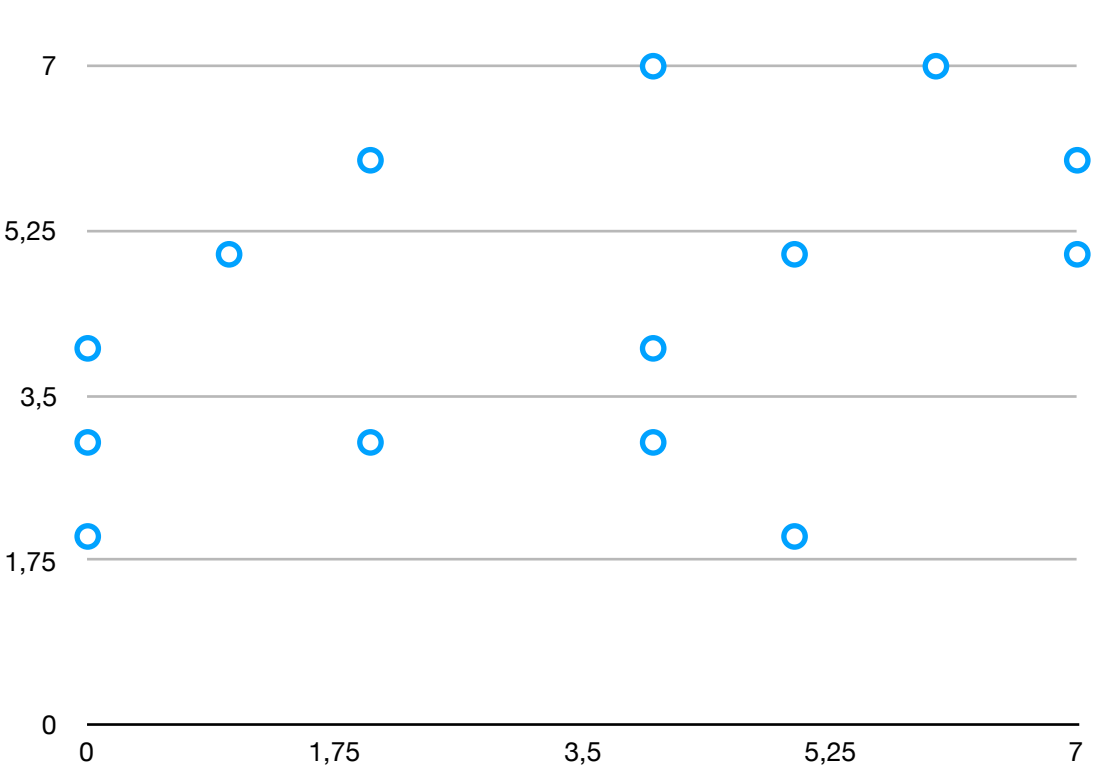
e) On peut les séparer grâce à un kernel de la forme $\frac{x^2}{a} + \frac{y^2}{b} = c$ (soit une ellipse)

Exercice 3.

a) On calcule leur p -value

Si celle-ci < 0.1 , les variables sont dépendantes

b) On calcule la dimension de Vapnik-Chervonenkis pour l'estimateur que l'on veut utiliser.



Points

0	2
0	3
0	4
1	5
2	6
4	7
6	7
7	6
7	5
2	3
4	3
4	4
5	2
5	5