Exercice 1

a) $ERM(EST\_1) = (8 \times 1 + 1 \times x)$

$= 8 + x$

$ERM(EST\_2) = (1 \times 1 + 3 \times x)$

$= 1 + 3x$

* $ERM(EST\_1) > ERM(EST\_2)$

$\Leftrightarrow \quad 8 + x > 1 + 3x$

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

$$\begin{cases} \text{Si } x < \frac{7}{2} & \text{Est\_2 meilleur} \\ \text{Sinon} & \text{Est\_1 meilleur} \end{cases}$$

b)

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) • $Y=T$

$$P(Y=T / X_1=1 \cap X_2=1) = P(Y=T) \times P(X_1=1/Y=T) \times P(X_2=1/Y=T)$$

$$= \frac{7}{17} \cdot \frac{5}{7} \cdot \frac{2}{7} \simeq 0,084$$

$$P(Y=T / X_1=1 \cap X_2=2) = \frac{7}{17} \cdot \frac{5}{7} \cdot \frac{5}{7} \simeq 0,210$$

$$P(Y=T / X_1=2 \cap X_1=1) = \frac{7}{17} \times \frac{2}{7} \cdot \frac{2}{7} \simeq 0,034$$

$$P(Y=T / X_1=2 \cap X_2=2) = \frac{7}{17} \times \frac{2}{7} \times \frac{5}{7} \simeq 0,084$$

• $Y=F$

$$P(Y=F / X_1=1 \cap X_2=1) = P(Y=F) \times P(X_1=1/Y=F) \times P(X_2=1/Y=F)$$

$$= \frac{10}{17} \times \frac{3}{10} \times \frac{6}{10} \simeq 0,106$$

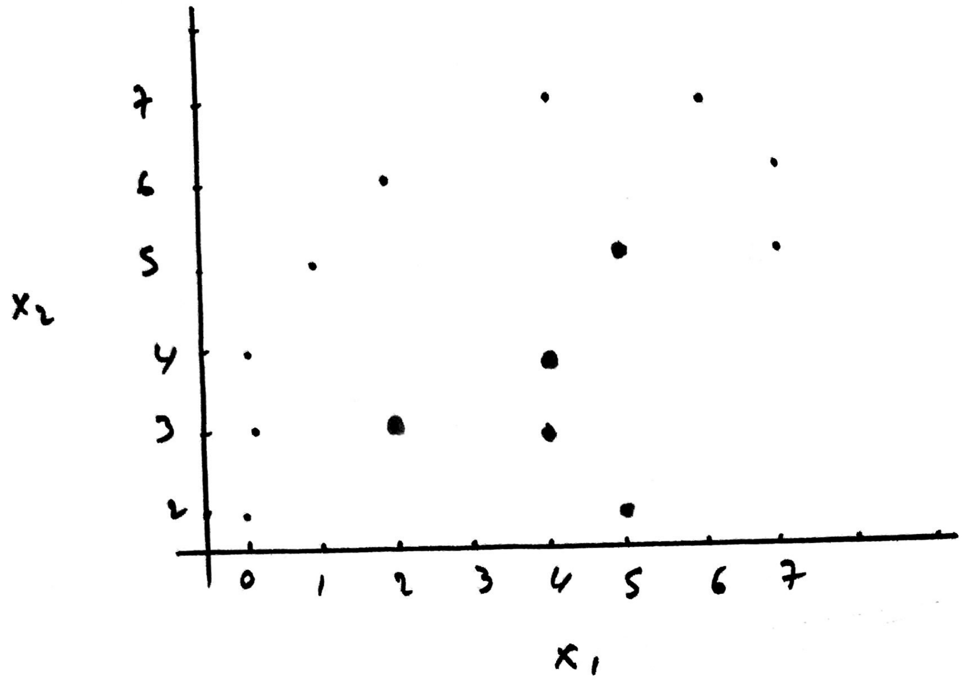
$$P(Y=F / X_1=1 \cap X_2=2) = \frac{10}{17} \times \frac{3}{10} \times \frac{4}{10} \simeq 0,071$$

$$P(Y=F / X_1=2 \cap X_2=1) = \frac{10}{17} \times \frac{7}{10} \times \frac{6}{10} \simeq 0,247$$

$$P(Y=F / X_1=2 \cap X_2=2) = \frac{10}{17} \times \frac{7}{10} \times \frac{4}{10} \simeq 0,165$$

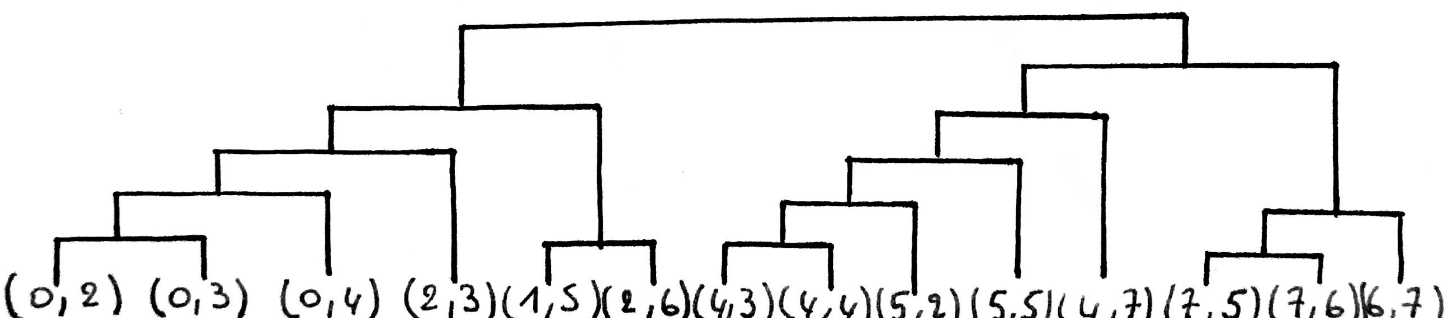
Exercice 2

a)



un K-means permettrait de faire un tel clustering.

b) dendogramme



la fonction choisie pour evaluer le coût est la distance

c) on peut utiliser un SVM linéaire pour séparer les 2 classes mais la séparation ne peut pas être parfaite

d) voir le shema

Exercice 3

a)
pour tester l'independance de 2 variables dans un tableau croisé d'effectif on calcule un chi2.
la p-value nous permet de savoir la probabilité d'avoir de telles valeurs aléatoirement.

p-value petit => très peu de chance => dependance
inversement p-value grand => independance

b) on peut essayer en tatonnant en commençant par 2 pts puis 3 ... jusqu'à ce que ce n'est plus possible

c) le risque c'est quand on connait les proba des evenements et ambiguité quand on les connait pas