Exercice 1

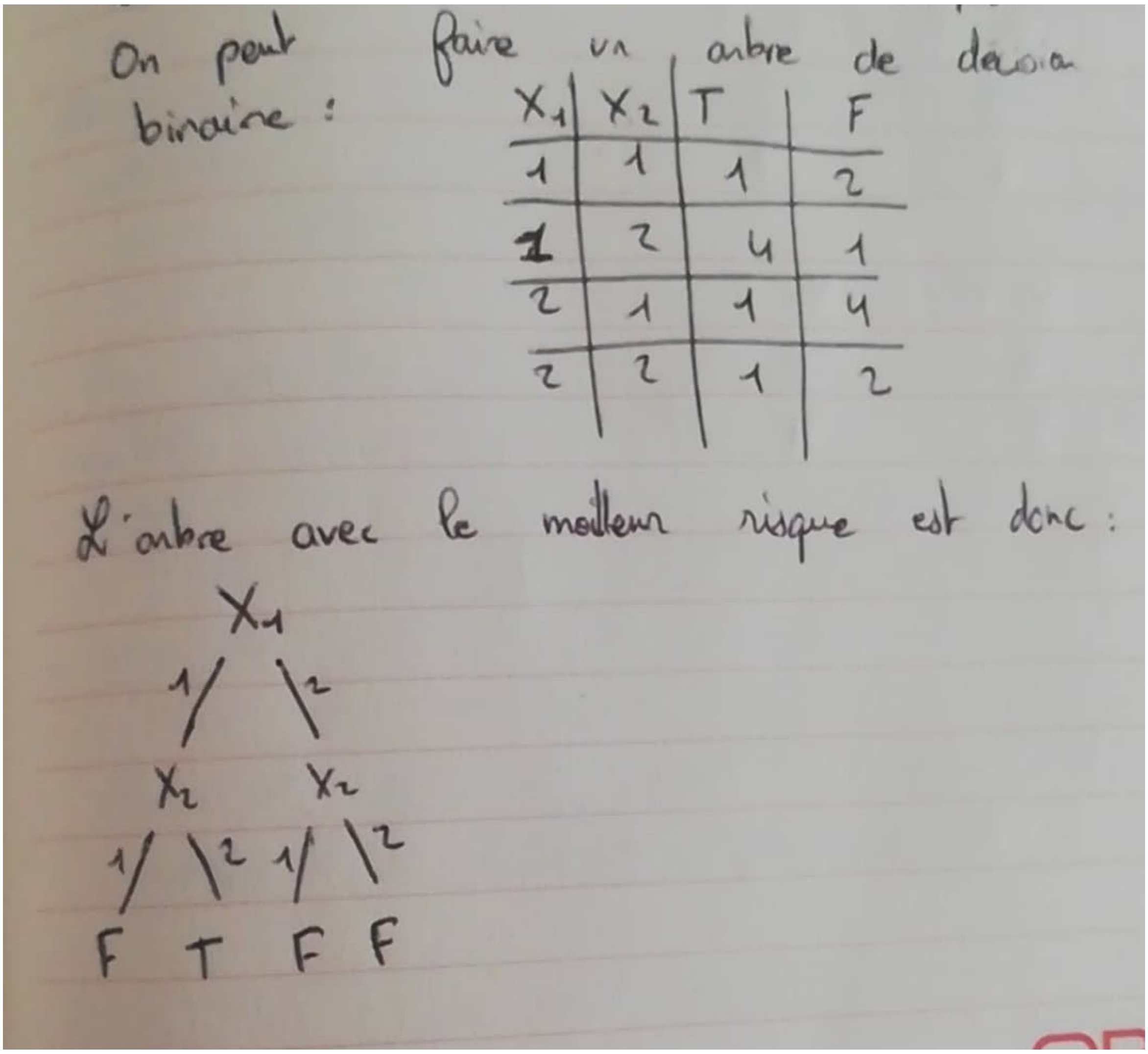
a) $R_{Est1}= 117(8+x)$

$R_{Est2}= 117(1+3x)$

$8+x=1+3x2x=7x=3.5$

Le risque est égal pour les deux estimateurs à $x=3.5$. Avant ça, l'estimateur 2 est meilleur. Après ce point, l'estimateur 1 devient meilleur à cause du cout de x.

b)


On peut faire un arbre de décision binaire :

X_1	X_2	T	F
1	1	1	2
1	2	4	1
2	1	1	4
2	2	1	2

L'arbre avec le meilleur risque est donc :

c)

- On possède assez d'informations sur les corrélations entre les différentes valeurs
- Vu que nous avons assez de données, on peut utiliser une meilleure méthode comme un arbre binaire qui devrait (en théorie) donner de meilleurs résultats qu'une approche Bayesienne naïve

d)

X_1	$Y=T$	$Y=F$
1	5	3
2	2	7

X_2	$Y=T$	$Y=F$
1	2	6
2	5	4

e) On cherche a voir, pour un cas donné, quelle sera la prédiction de l'estimateur bayesien naif.

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

$|Y=T) P(X_1=1,X_2=1) = 57 * 27 * 717\ 317 = 1021$

Similairement pour $Y=F$, on trouve:

$P(Y=F|X_1=1,X_2=1) = P(Y=F) * P(X_1=1|Y=F) * P(X_2=1$

$|Y=F) P(X_1=1,X_2=1) = 310 * 610 * 1017\ 317 = 610$

Etant donné que $P(Y=F|\ldots) > P(Y=T|\ldots)$, notre estimateur bayesien donnera dans ce cas un résultat de $Y=F$.

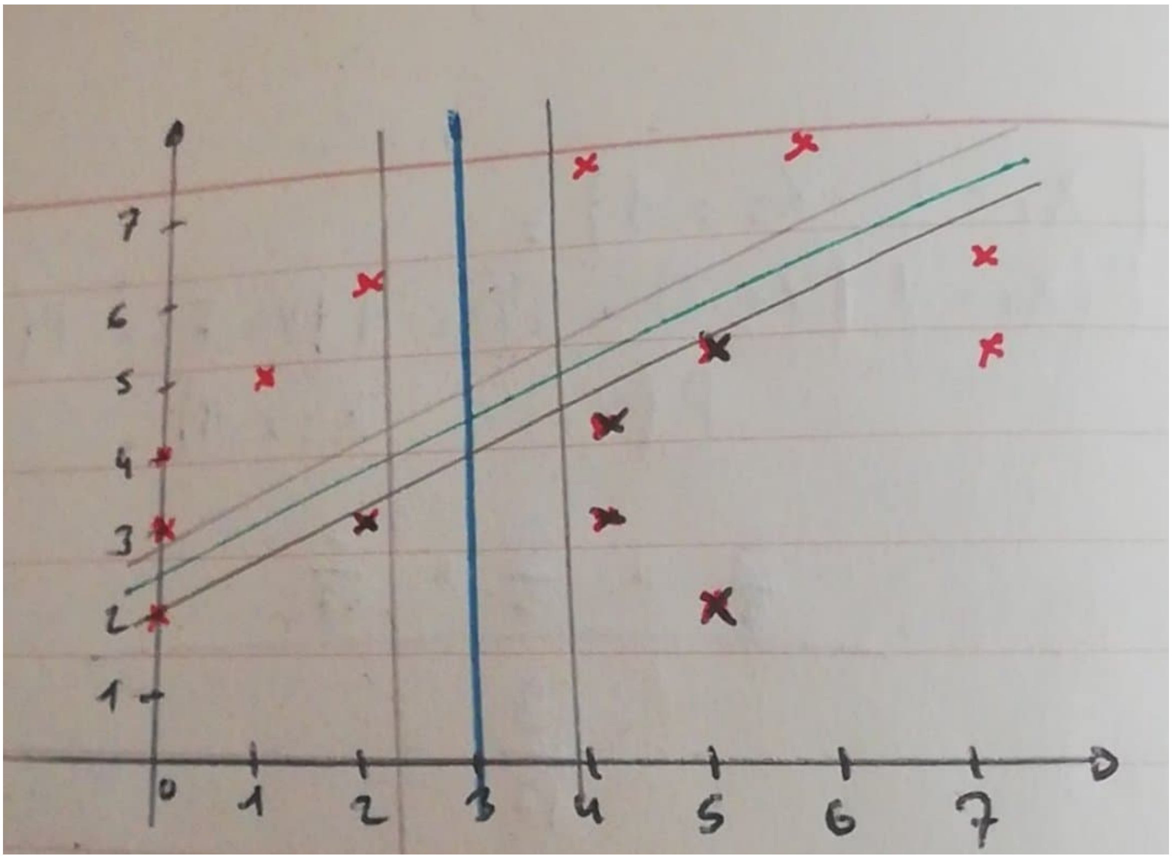
De manière similaire, on trouve pour les autres cas:

- $P(Y=T|X_1=1,X_2=2) > P(Y=F|X_1=1,X_2=2) \Rightarrow T$
- $P(Y=F|X_1=2,X_2=1) > P(Y=T|X_1=2,X_2=1) \Rightarrow F$
- $P(Y=F|X_1=2,X_2=2) > P(Y=T|X_2=2,X_2=2) \Rightarrow F$
  Dans ce cas, l'estimateur bayesien naïf donne les même résultats que l'arbre de décision binaire!

On voit en effet les même prédictions pour les même valeurs de X_1 X_1 et X_2

Exercice 2

a)



Les méthodes les plus appropriées dans ce genre de cas ("cercles imbriqués") seraient:

- AgglomerativeClustering
- SpectralClustering
- OPTICS
- DBSCAN

b)

c) Non ce genre de problème ne peut pas se résoudre avec une SVM utilisant un kernel linéaire.

d) On voit dans le dessin a) en bleu une SVM avec un coefficient bas et en vert une SVM avec un coefficient plus élevé (marge plus strictes)

e) On peut probablement séparer les deux classes en utilisant un kernel polynomial

Exercice 3

a) Faire un test statistique de type Chi² ou Student permet d'obtenir une probabilité d'indépendance de données.

b) On peut répeter les étapes de pulvérisation normales jusqu'à atteindre une étape bloquante

c) Le risque est connu a l'avance alors que l'ambiguité est inconnue. On peut anticiper le risque mais difficilement (voir pas du tout) l'ambiguité