Exercice 1

a) Le risque empirique associé à EST_1 :

D(f) = 8 + x

Le risque empirique associé à EST_2 :

D(f) = 1 + 3x

Si on considère l'équation :

$8 + x < 1 + 3x \Leftrightarrow 7 < 2x \Leftrightarrow x > 7/2$

Cela veut dire que EST_1 a l'avantage si x > 7/2 et EST_2 a l'avantage dans l'autre cas.

b) Voici mon estimateur :

MY_EST
T
F
T
F
T
F
F
T
F
F
T
F
T
F
F
F
T

d)

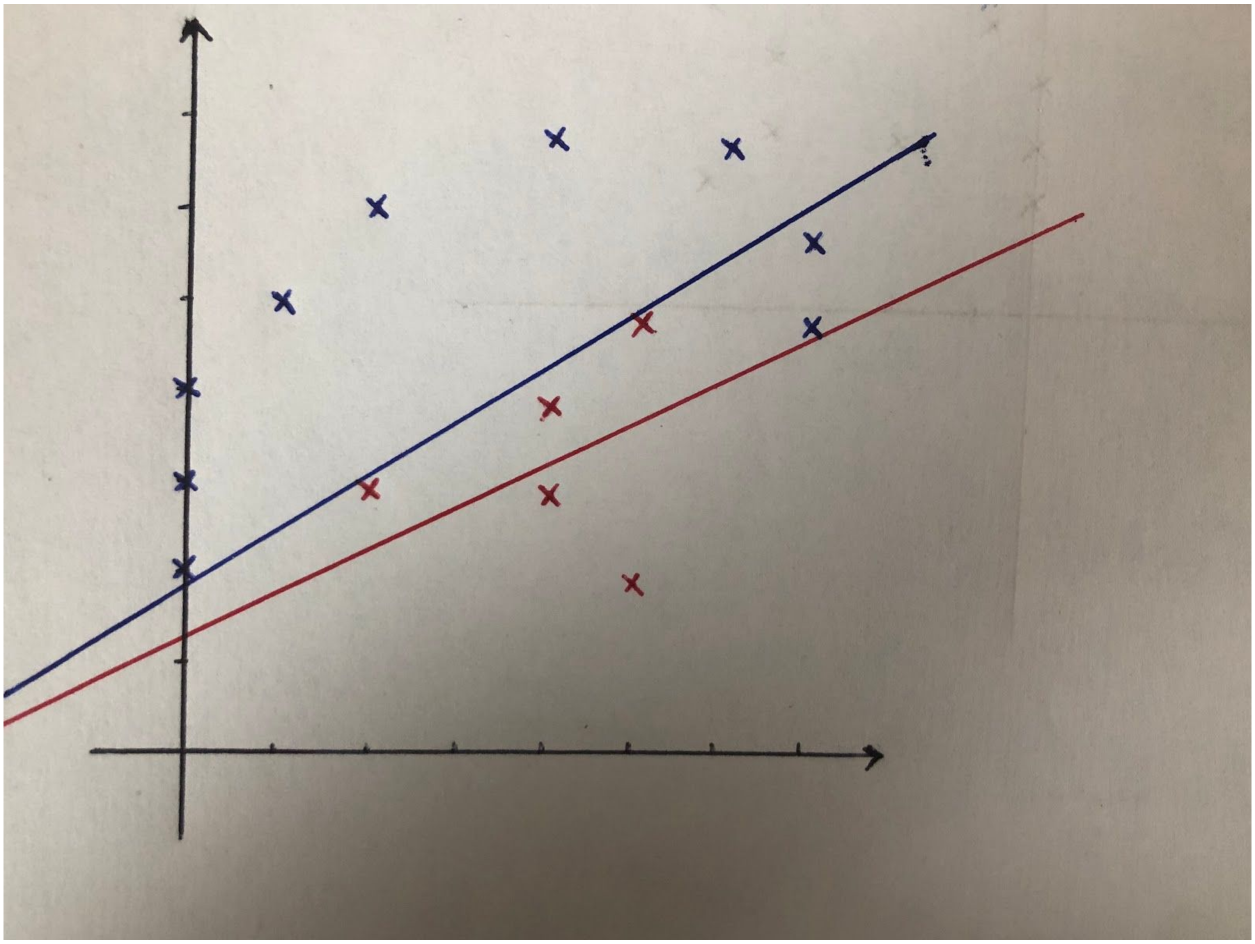
Y = T	1	2
X_1	5	2
X_2	2	5

Y = F	1	2
X_1	3	7
X_2	6	4

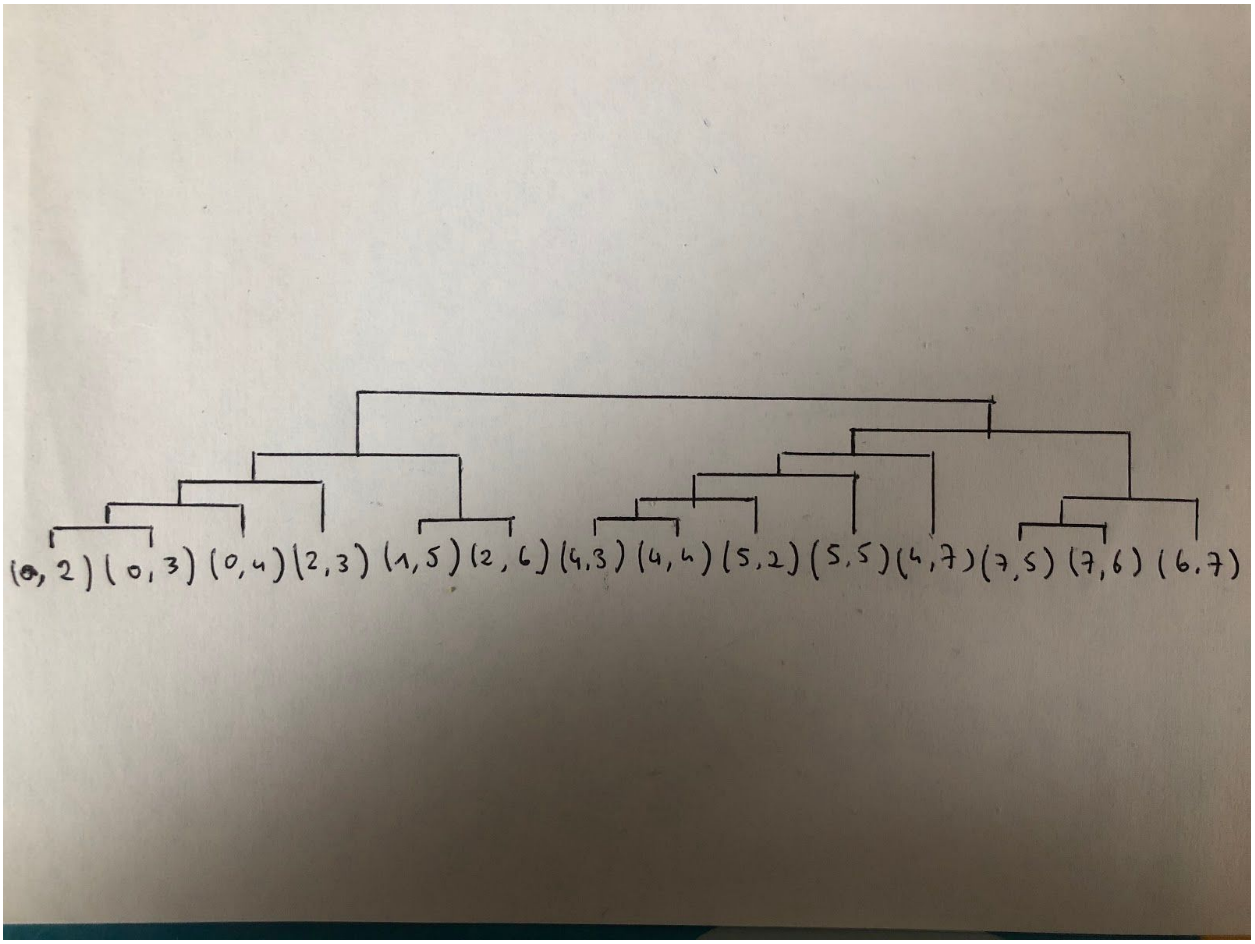
e) P(Y=T|X1=1, X2=1) = 7/17 * 5/7 * 2/7
P(Y=T|X1=1, X2=2) = 7/17 * 5/7 * 5/7
P(Y=T|X1=2, X2=1) = 7/17 * 2/7 * 2/7
P(Y=T|X1=2, X2=2) = 7/17 * 2/7 * 5/7
P(Y=F|X1=1, X2=1) = 10/17 * 3/10 * 6/10
P(Y=F|X1=1, X2=2) = 10/17 * 3/10 * 4/10
P(Y=F|X1=2, X2=1) = 10/17 * 7/10 * 6/10
P(Y=F|X1=2, X2=2) = 10/17 * 7/10 * 4/10

Exercice 2

a) Les méthodes non supervisées qui pourraient distinguer ces deux classes seraient des algorithmes de clustering, je pense plus particulièrement au k-means et au CAH.



b) Pour le dendrogramme, j’ai choisi une fonction de distance maximum pour l’intégration.



c) Un SVM linéaire ne serait pas la méthode optimale pour séparer ces deux classes. Une façon simple de s'en rendre compte est qu'il est difficile de séparer ces deux classes à l'aide d'une droite

d) Voir figure 1. J'ai choisi un niveau de pénalisation grand pour la droite bleue et plus petit pour la droite rouge.

e) Vu qu'on a une classe qui est à l'intérieur de l'autre, je propose d'utiliser un kernel gaussien pour évoluer vers un modèle plus linéaire qui pourra ensuite être géré à l'aide d'un SVM.

Exercice 3

a) Si l'on dispose du tableau croisé d'effectif, on peut faire un chi2 dessus pour obtenir la p-value. Cette dernière peut ensuite être interprétée en fonction des cas. Si elle est très petite, ça veut dire que les deux variables sont dépendantes, sinon elle ne le sont pas.

b) Si je voulais tester le nombre de points pulvérisables à la surface d’une sphère je commencerais par une petit nombre puis j'augmenterais au fur et à mesure.

c) La différence entre risque et ambiguïté est que dans le cas du risque, on se base sur des probabilités que l’on connaît pour déterminer les chances de se tromper alors que dans le cas de l'ambiguïté on n’est pas sûr qu’on s’est trompé ou pas.