FTML

Exercice 1

EST_1

TP	FN
6	1
8	2
FP	TN

EST_2

TP	FN
4	3
1	9
FP	TN

a). La fonction de risque empirique se traduit par.

$f_{EST}(x) = nFP + xFN$ $\{$ n le nombre de faux positifs

Pour l'estimateur 1, on a $f_{EST\_1}(x) = 8 + x$

l'estimateur 2, on a $f_{EST\_2}(x) = 1 + 3x$.

Comparons les.

$$8 + x \geqslant 1 + 3x \iff 7 \geqslant 2x$$
$$\iff \frac{7}{2} \geqslant x$$

L'estimateur 2 est avantagé sur $]-\infty, \frac{7}{2}]$

L'estimateur 1 est avantagé sur $[\frac{7}{2}, +\infty[$.

b). On pose $x = 2$.

Donc $f_{EST}(x) = nFP + 2FN$.

On se place dans le cas bayesien naïf

Y=T	1	2
X1	5	2
X2	2	5

Y=F	1	2
X1	3	7
X2	6	4

$P(Y=T / x=(1,1)) = \frac{7}{17} \times \frac{5}{7} \times \frac{2}{7} = \frac{10}{119}$

$P(Y=T / x=(1,2)) = \frac{7}{17} \times \frac{5}{7} \times \frac{5}{7} = \frac{25}{119}$

$P(Y=T / x=(2,1)) = \frac{7}{17} \times \frac{2}{7} \times \frac{2}{7} = \frac{4}{119}$

$P(Y=T / x=(2,2)) = \frac{7}{17} \times \frac{2}{7} \times \frac{5}{7} = \frac{10}{119}$

$P(Y=F / x=(1,1)) = \frac{10}{17} \times \frac{3}{10} \times \frac{6}{10} = \frac{9}{85}$

$P(Y=F / x=(1,2)) = \frac{10}{17} \times \frac{3}{10} \times \frac{4}{10} = \frac{6}{85}$

$P(Y=F / x(2,1)) = \frac{10}{17} \times \frac{7}{10} \times \frac{6}{10} = \frac{21}{85}$

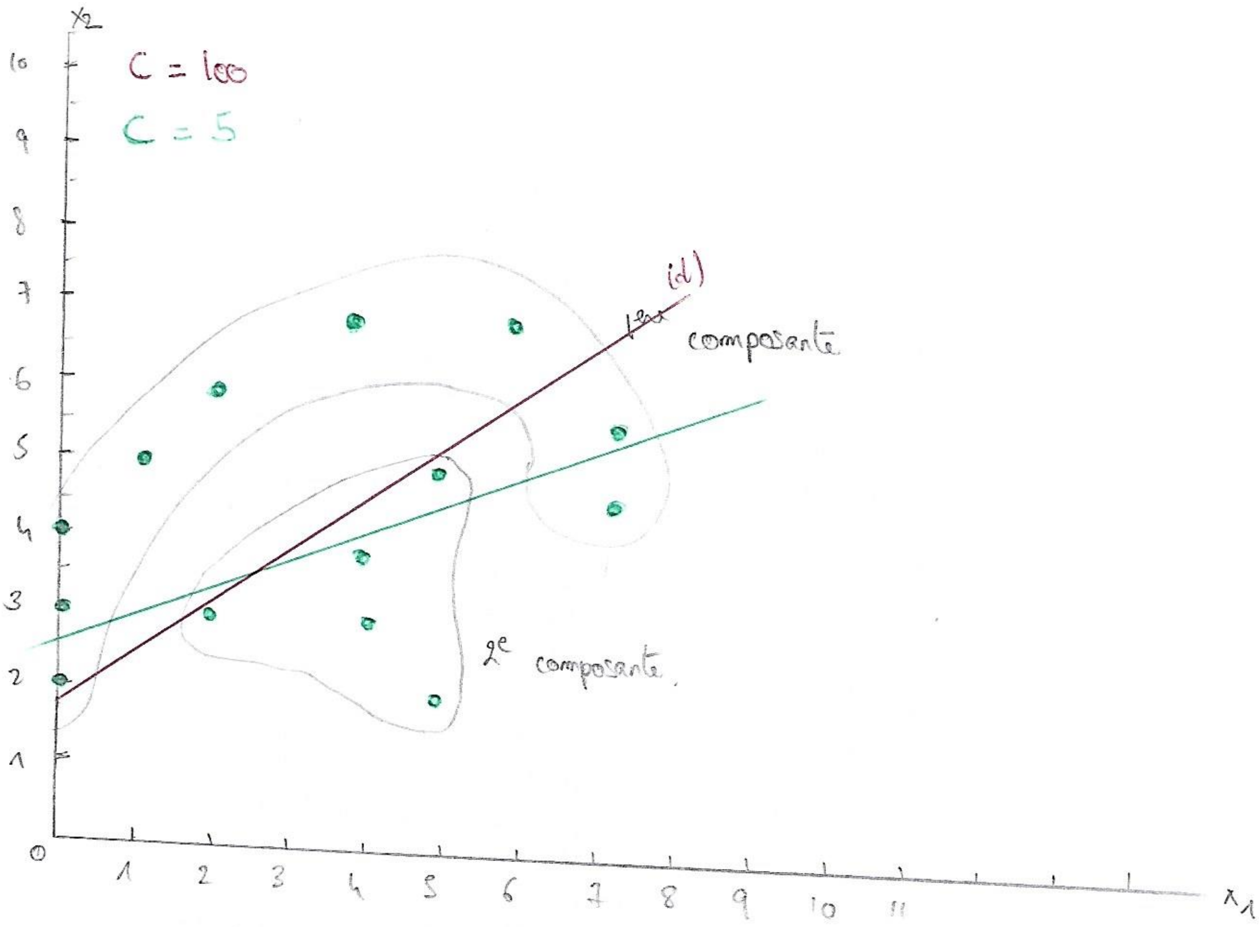
$P(Y=F / x(2,2)) = \frac{10}{17} \times \frac{7}{10} \times \frac{4}{10} = \frac{14}{85}$

Pour (1,1), on a 1T, 2F donc OK
(1,2), on a 4T, 1F donc OK
(2,1), on a 1T, 3F donc OK
(2,2), on a 1T, 3F donc OK.

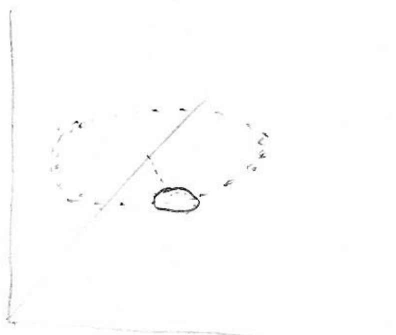
On obtient un risque nul

c).

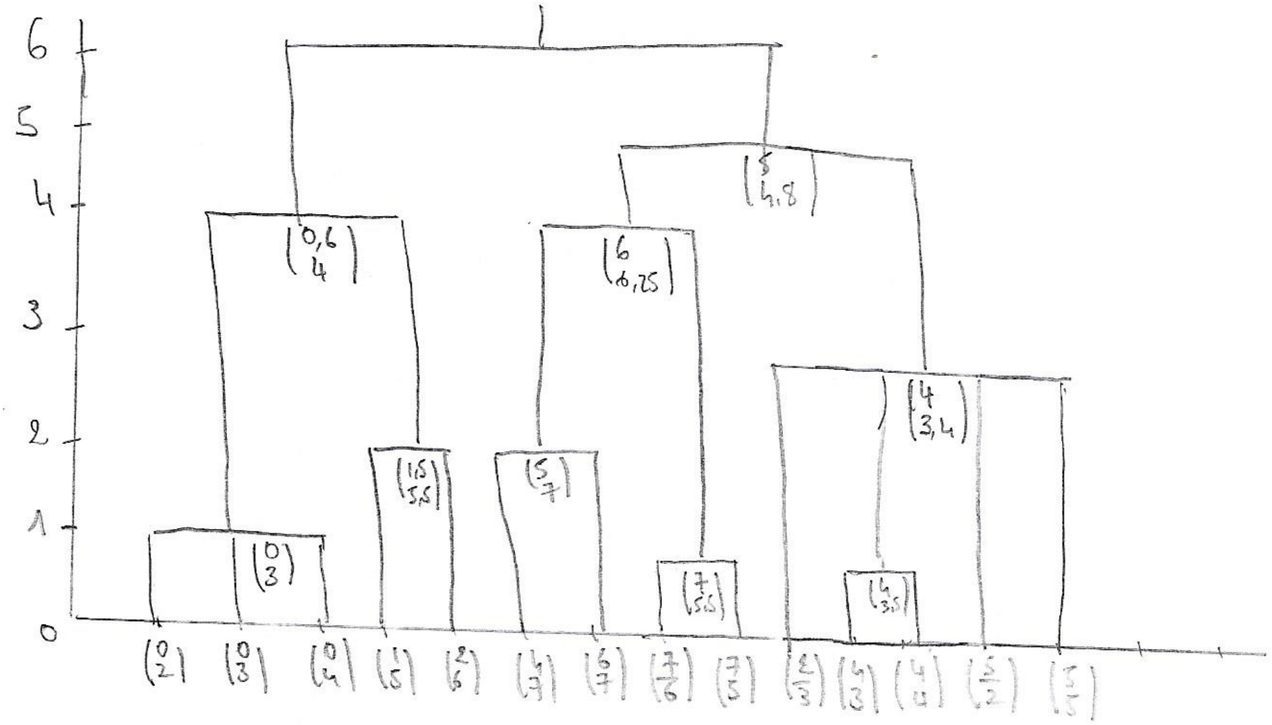
Exercice 2



a) Il semblerait qu'on obtient la 1ère composante qui est l'extérieur et la 2e composante qui est l'intérieur, si on réalise un kernel trick et qu'on réussit à le modéliser en 3D, on peut sûrement séparer les point par un SVM.



2b) On prend comme distance, la distance de manhattan.



2c) D'après ma réponse en 2a), il est possible de les séparer à l'aide d'un plan, dont l'équation est $ax + by + cz + d = 0$, donc ceci il est possible de les séparer à l'aide d'un SVM linéaire si on applique une augmentation de dimension.

Si on ne réalise pas d'augmentation, il est impossible de séparer parfaitement les deux composants à l'aide d'un SVM linéaire

2d) Voir graphe 2a).

3e) On propose la fonction suivante

$$f \left\{ \begin{array}{l} \mathbb{R}^2 \longrightarrow \mathbb{R}^3 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longrightarrow \begin{pmatrix} x_1 \\ x_2 \\ (x_1-4)^2 + (x_2-1)^2 \end{pmatrix} \end{array} \right.$$

On peut séparer selon $\boxed{z = 17}$

$x_1 - 4$	$x_2 - 1$	(z) Res.
-4	1	17
-4	2	20
-4	3	25
-3	5	25
-2	5	29
0	6	36
2	6	40
3	5	34
3	4	25
-2	2	8
0	2	4
0	3	9
1	1	2
1	4	17

a). On réalise un test de chi2 en posant au préalable la valeur minimum à atteindre, si la valeur du chi2 est inférieur au seuil, alors les deux variables sont indépendantes.

b) Il faut déterminer la dimension de Vapnik Chervonenkis de la famille qui pulvérise.

2c) Le risque est souvent quelque chose dont on a le contrôle, on décide nous-même des fonctions de pertes. Tandis que l'ambiguïté est lié aux donnés elles-mêmes, nous n'avons pas le contrôle dessus. A moins de réussir à faire soi-même son échantillon sans ambiguïté, et encore, ce n'est pas chose aisée.