

Hugo  
Tritsch  
SCIA 2021  
Exercice 1

FTHL 2020

a)

EST-1

		Actual	
		T	F
Predict	T	6	8
	F	1	2

EST-2

		Actual	
		T	F
Predict	T	4	1
	F	3	9

EST-1:  ~~$E[y_i | y_i'] = 8x_1 +$~~

$ERH_1 = 8 \times 1 + x \times 1 = 8 + x$

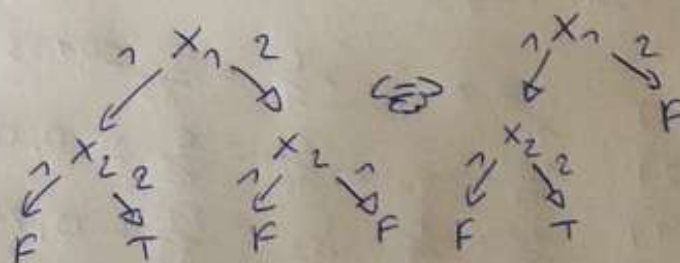
EST-2:  $ERH_2 = 1 \times 1 + 3 \times x = 1 + 3x$

plus  $x$  est grand plus EST-2 est moins bon par rapport à EST-1  
plus  $x$  est petit plus EST-2 est meilleur par rapport à EST-1

b)

$x = 2$   
 $ERH_1 = 8 + 2 = 10$

$ERH_2 = 1 + 6 = 7$



c) Il n'est pas nécessaire ici de se placer dans le cadre bayésien mais :

- indépendance trop forte entre les deux variables

-

d)

$Y = \text{True}$	1	2	$Y = \text{False}$	1	2
$X_1$	5	2	$X_1$	3	7
$X_2$	2	5	$X_2$	6	4

e) Pour  $Y = \text{True}$  : $X_1, X_2$ 

$$(1, 1) \quad \frac{7}{17} \times \frac{5}{7} \times \frac{2}{7} = \frac{10}{119} \approx 0,084$$

$$(1, 2) \quad \frac{7}{17} \times \frac{5}{7} \times \frac{5}{7} = \frac{25}{119} \approx 0,210$$

$$(2, 1) \quad \frac{7}{17} \times \frac{2}{7} \times \frac{2}{7} = \frac{4}{119} \approx 0,033$$

$$(2, 2) \quad \frac{7}{17} \times \frac{2}{7} \times \frac{5}{7} = \frac{10}{119} \approx 0,084$$

Pour  $Y = \text{False}$  $X_1, X_2$ 

$$(1, 1) \quad \frac{10}{17} \times \frac{3}{10} \times \frac{6}{10} = \frac{9}{85} \approx 0,105$$

$$(1, 2) \quad \frac{10}{17} \times \frac{3}{10} \times \frac{4}{10} = \frac{6}{85} \approx 0,071$$

$$(2, 1) \quad \frac{10}{17} \times \frac{7}{10} \times \frac{6}{10} = \frac{21}{85} \approx 0,247$$

$$(2, 2) \quad \frac{10}{17} \times \frac{7}{10} \times \frac{4}{10} = \frac{14}{85} \approx 0,165$$

$$(X_1, X_2) \rightarrow F$$

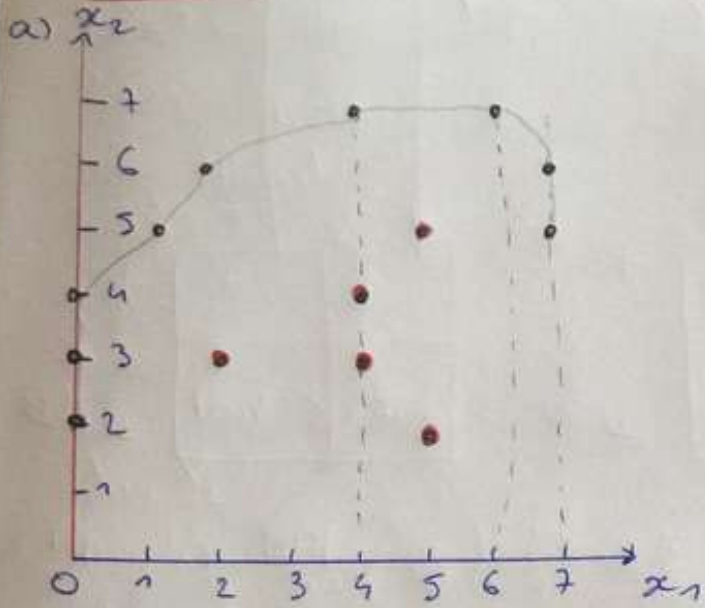
$$(1, 1) \rightarrow F$$

$$(1, 2) \rightarrow T$$

$$(2, 1) \rightarrow F$$

$$(2, 2) \rightarrow F$$

## Exercice 2

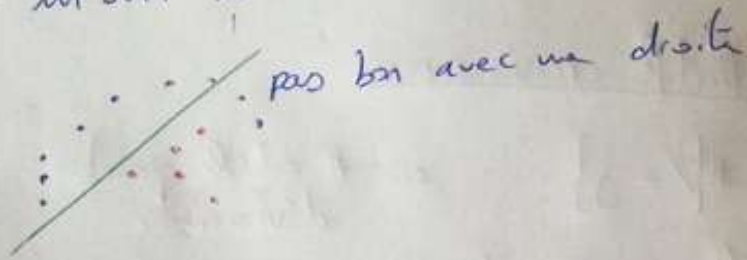


La classification hiérarchique ascendante semble à même de les distinguer.

b)



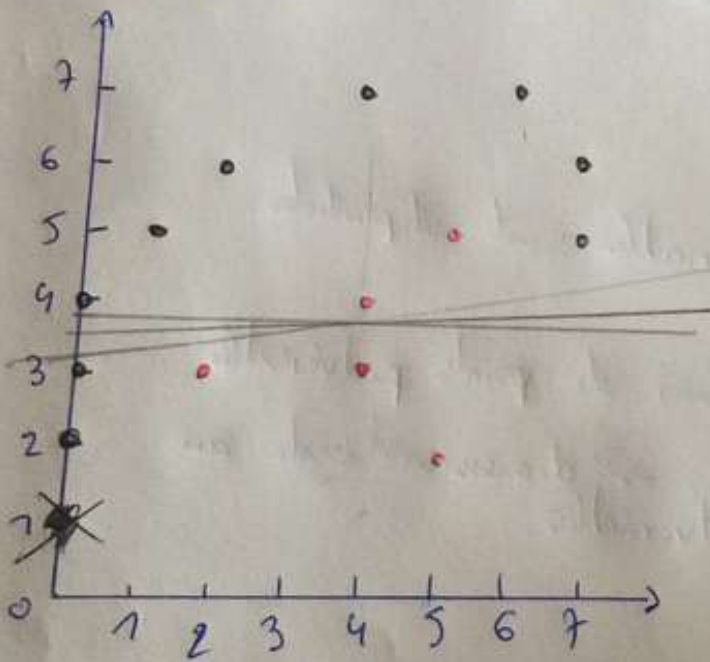
c) il n'est pas possible de distinguer les deux classes avec un SVM linéaire



d)

$$y' > y \Rightarrow LF(y'', y') = 1$$

$$y' < y \Rightarrow LF(y, y') = 1$$



e) Kernel RBF  $K(x, x') = e^{-\frac{\|x - x'\|^2}{2\sigma^2}}$

### Exercice 3

a)

on utilise le test du  $\chi^2$

1<sup>ère</sup> étape calcul les effectifs croisés espérés

$$E_{i,j} = \frac{1}{n} \sum \# \{X=i\} \# \{Y=j\}$$

où  $\#$  est le nombre d'événements

2<sup>ème</sup> étape calcul effectifs croisés observés

$$O_{i,j} = \# \{X=i \text{ et } Y=j\}$$

3<sup>ème</sup> étape calcul l'écart relatif entre les 2 variables

$$T = \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

Lecture des résultats

Si  $T$  est faible alors les variables sont indépendantes

b) Pour trouver le nombre maximum de points pulvérisables  
On peut calculer Vapnik-Chervonenkis  $\rightarrow$  sa dimension est égale au nombre maximum d'éléments pulvérisables.

c)