Guillaume
BLASSEL

Devoir de FTML

Exercice 1:

a) On obtient les matrices de confusion suivantes:

Estimateur 1:

	T	F
T	6	1
F	8	2

Donc $ERM(Est1) = \dfrac{8 \times 1 + x \times 1}{17}$

Estimateur 2:

	T	F
T	4	3
F	1	9

Donc $ERM(Est2) = \dfrac{3x + 1}{17}$

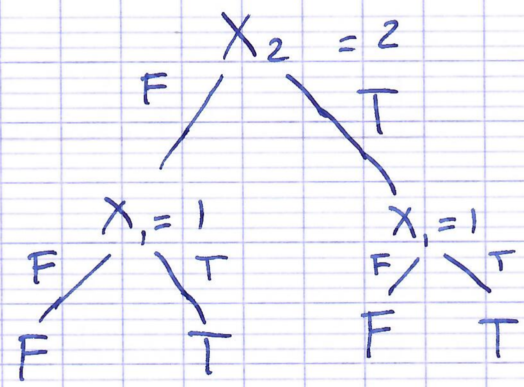
Pour que $Est1 > Est2$

: $\dfrac{8+x}{17} > \dfrac{3x+1}{17}$

$x < \dfrac{7}{2}$

Sinon si $x > \dfrac{7}{2}$, on a $Est2 > Est1$

b)



On peut dire que cet estimateur minimise le risque empirique avec 1 de pénalité pour un faux positif, et 2 pour un faux négatif

c) Avec un arbre de décision, on peut obtenir un bon ERM.
Si ça n'avait pas été le cas on aurait pu envisager un bayésien naïf.

d)

X_1	T	F
1	5	3
2	2	7

X_2	T	F
1	2	6
2	5	4

Pour $Y = T$:

x_1 \ x_2	1	2
1	$\dfrac{\frac{5}{7} \times \frac{2}{7} \times \frac{7}{17}}{\frac{5+3}{17} \times \frac{2+6}{17}}$	$\dfrac{\frac{5}{7} \times \frac{5}{7} \times \frac{7}{17}}{\frac{5+3}{17} \times \frac{5+4}{17}}$
2	$\dfrac{\frac{2}{7} \times \frac{2}{7} \times \frac{7}{17}}{\frac{2+7}{17} \times \frac{2+6}{17}}$	$\dfrac{\frac{2}{7} \times \frac{5}{7} \times \frac{7}{17}}{\frac{2+7}{17} \times \frac{5+4}{17}}$

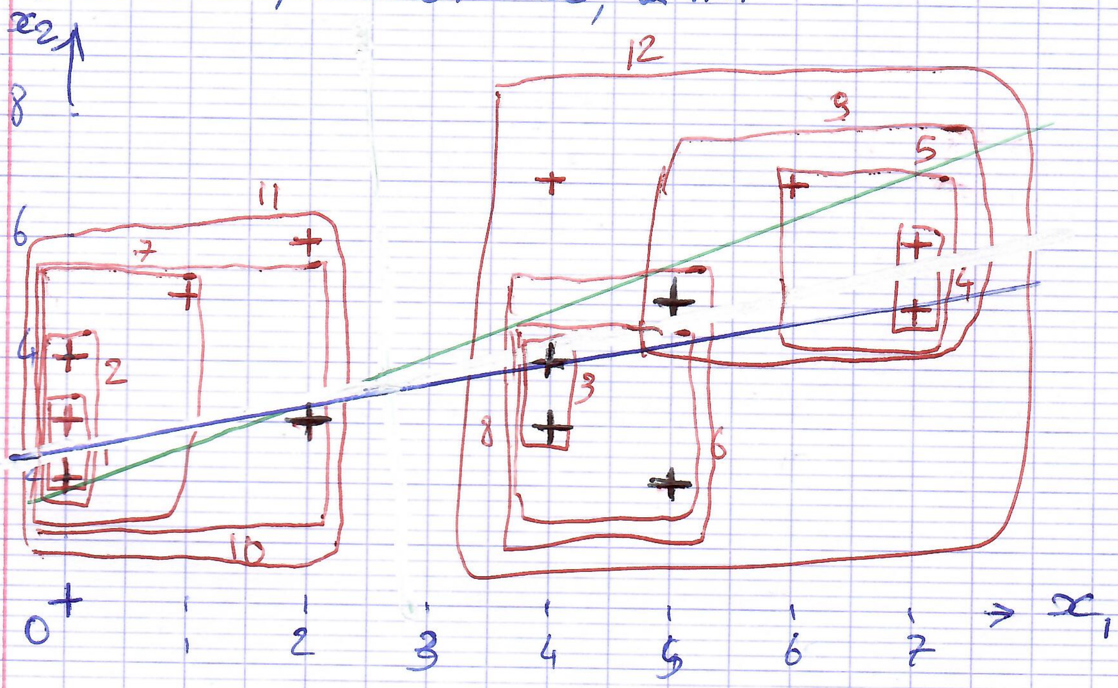
$y = F$ x_1 \ x_2	1	2
1	$\dfrac{\frac{3}{10} \times \frac{6}{10} \times \frac{10}{17}}{\frac{3+5}{17} \times \frac{6+2}{17}}$	$\dfrac{\frac{3}{10} \times \frac{4}{10} \times \frac{10}{17}}{\frac{5+3}{17} \times \frac{5+4}{17}}$
2	$\dfrac{\frac{7}{10} \times \frac{6}{10} \times \frac{10}{17}}{\frac{7+2}{10} \times \frac{6+2}{10}}$	$\dfrac{\frac{7}{10} \times \frac{4}{10} \times \frac{10}{17}}{\frac{7+2}{10} \times \frac{4+5}{10}}$

Conclusion :

x_1 \ x_2	1	2
1	$y = F$	$y = T$
2	$y = F$	$y = F$

Exercice 2 :

a) Ils sont en vois, il faudrait utiliser par exemple : RBF SVM, decision tree, knn



c) Malheureusement c'est impossible, comme on peut le voir sur le schéma il y aura forcément des outliers.

d) Les SVM sont en vert et bleu.
Pour le vert, la fonction de perte est de 10 si les noirs sont mal classés, et 0 si les rouge sont mal classés.
Pour le bleu, la fonction de perte est uniforme, il y a donc 2 mal classés de chaque coté

Exercice 3:

a) On peut tester l'indépendance grace au test du Chi2.

b) Pour connaitre le nombre maximum de points pulvérisable à la surface d'une sphère, on peut utiliser la dimension de Vapnik-Chervonenkis

c) Un risque est le fait de connaitre déjà à l'avance de probabilités, par exemple jouer avec dés. Même s'il y a incertitude, elle est connue.
L'ambiguïté c'est quand on a aucune idée des probabilités impliquées.