Exercice 1

a) Est_1 :

GT \ Pred	T	F
T	6	1
F	8	2

Est_2 :

GT \ Pred	T	F
T	4	3
F	1	9

$Loss_1 = 1 \times \#FauxPositifs + x \times \#FauxNegatifs$

$= 8 + x \times 1$

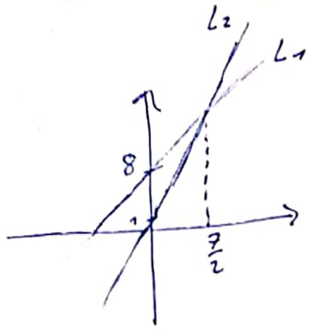
$= x + 8$

$Loss_2 = 1 \times \#FauxPositifs + x \times \#FauxNegatifs$

$= 1 + 3x$

$= 3x + 1$

$Loss_1 \leq Loss_2 \iff x+8 \leq 3x+1 \iff 7 \leq 2x \iff \boxed{\frac{7}{2} \leq x}$



b) Données

X	# Y = T	# Y = F
11	1	2
12	4	1
21	1	4
22	1	3

Si l'on prend un estimateur déterministe,

$\rightarrow Loss_{11}(Ypred = T) = 2 = Loss_{11}(Ypred = F) = 2$

$\rightarrow Loss_{12}(Ypred = T) = 1 < Loss_{12}(Ypred = F) = 8$

$\rightarrow Loss_{21}(Ypred = T) = 4 > Loss_{21}(Ypred = F) = 2$

$\rightarrow Loss_{22}(Ypred : T) = 3 > Loss_{22}(Ypred : F) = 2$

Le meilleur choix pour un estimateur déterministe nécessite donc

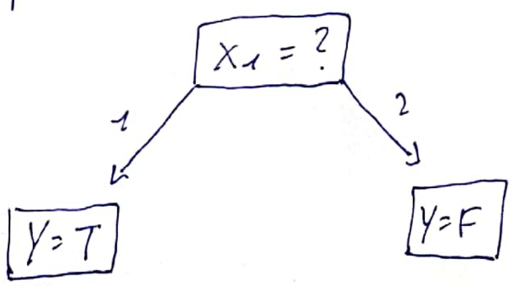
$f(1\ 2) = T$

$f(2\ 1) = F$

$f(2\ 2) = F$

Total loss = 7

On peut prendre un arbre de décision tel que



Scanné avec CamScanner

c) L'espace des données est suffisamment petit :
↳ 4 valeurs possibles pour X et 2 pour Y

d)

$Y=T$	$X_i=1$	$X_i=2$
X_1	5	2
X_2	2	2

$Y=F$	$X_i=1$	$X_i=2$
X_1	3	7
X_2	6	4

e) $P(Y=T \mid X=11) = \frac{1}{\Omega_{11}} \times P(X_1=1 \mid Y=T) \times P(X_2=1 \mid Y=T) = \frac{1}{\Omega_{11}} \times \frac{5}{7} \times \frac{2}{4} = \frac{1}{\Omega_{11}} \times \frac{5}{14}$

$P(Y=F \mid X=11) = \frac{1}{\Omega_{11}} \times \frac{3}{10} \times \frac{6}{10} = \frac{1}{\Omega_{11}} \times \frac{9}{50}$:

$\frac{5}{14} \simeq 0,357 > \frac{9}{50} \simeq 0,18$

$P(Y=T \mid X=12) \simeq \frac{5}{7} \times \frac{1}{2} = \frac{5}{14}$ $\qquad$ $\frac{5}{14} \simeq 0,357 > \frac{6}{50} \simeq 0,12$

$P(Y=F \mid X=12) \simeq \frac{3}{10} \times \frac{4}{10} = \frac{6}{50}$

$P(Y=T \mid X=21) \simeq \frac{2}{7} \times \frac{1}{2} = \frac{1}{7}$ $\qquad$ $\frac{1}{7} \simeq 0,143 < \frac{21}{50} = 0,42$

$P(Y=F \mid X=21) \simeq \frac{7}{10} \times \frac{6}{10} = \frac{21}{50}$

$P(Y=T \mid X=22) \simeq \frac{2}{7} \times \frac{1}{2} = \frac{1}{7}$ $\qquad$ $\frac{1}{7} \simeq 0,143 < \frac{14}{50} = 0,28$

$P(Y=F \mid X=22) \simeq \frac{7}{10} \times \frac{4}{10} = \frac{14}{50}$

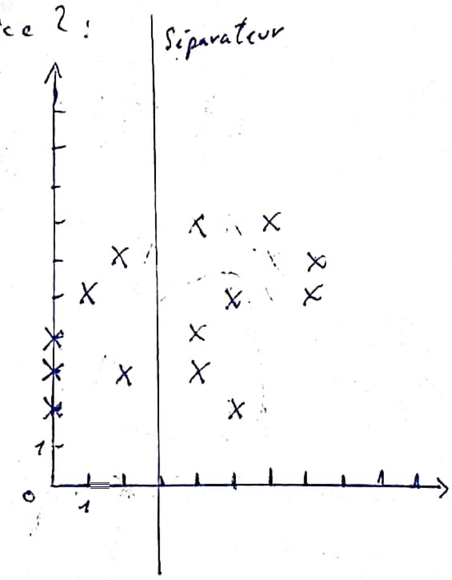
Au Final :

$X_1 \backslash X_2$	1	2
1	$Y=T$	$Y=T$
2	$Y=F$	$Y=F$

Cet estimateur Bayésien naïf est rigoureusement équivalent.

Exercice 2 :

a)



Un Séparateur Linéaire ?

b) Je ne me souviens pas avoir vu ça. Et ne sait pas ce qu'est le "coût d'intégration".

J'aurais du prendre (7, 5), mais je n'ai plus le temps

c) (0, 2) → (7, 6) $y = ax + b$ → $a = \frac{6-2}{7-0} = \frac{4}{7}$ $y = \frac{4}{7}x + 2$

$b = 2$

$\frac{4}{7} \times 2 + 2 = \frac{22}{7} \simeq 3{,}1 > 3$

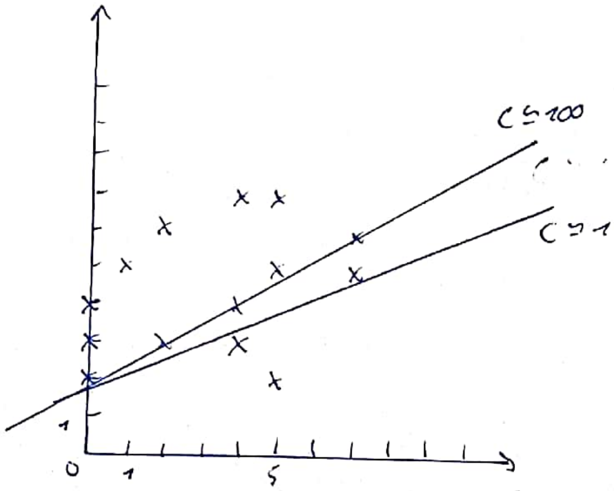
$\frac{4}{7} \times 4 + 2 = \frac{30}{7} \simeq 4{,}29 > 4$

le point intérieur (5, 5) se situe dans ← $\frac{4}{7} \times 5 + 2 = \frac{20}{7} + 2 = \frac{34}{7} \simeq 4{,}86 < 5$

l'enveloppe convexe des points extérieurs

↳ Non séparable par SVM linéaire.

d)



e) Une gaussienne centrée sur (4,3) pourrait fonctionner.

$$f(x_1, x_2) = e^{-\frac{1}{2}\left((x_1-4)^2+(x_2-3)^2\right)}$$

Exercice 3 :

a) Test de χ_2

b) Il faudrait savoir avec quelle famille de séparateurs on travaille. Partir de 0, et monter de 1 en 1 jusqu'à ce qu'on arrive plus à pulvériser.

c) Je prendrai un exemple où l'on connait "les évènements et leur probabilités" et un autre où l'on connait "les évènements sans leurs probabilités".