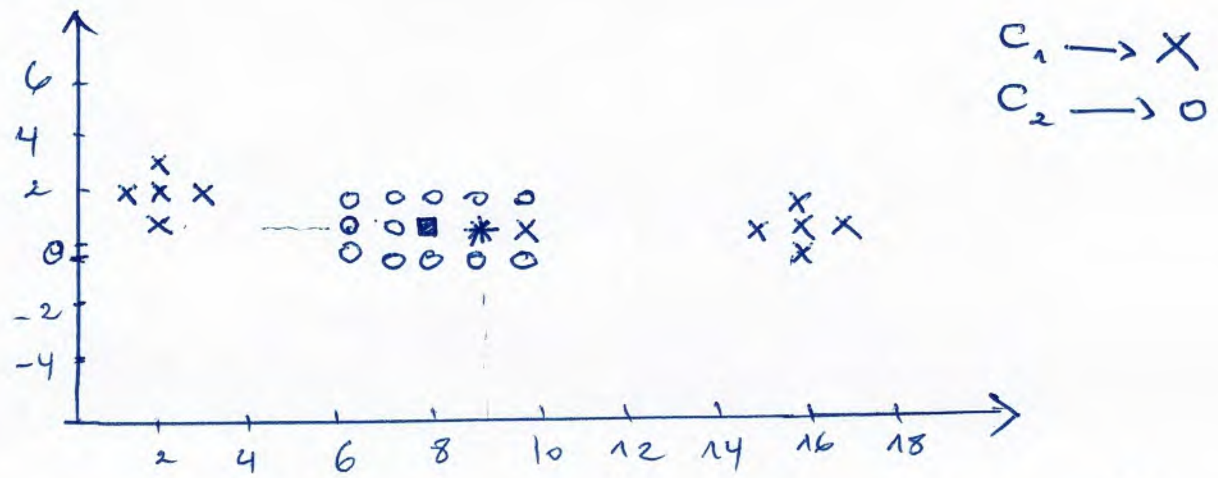
Solution TD n° 4

Exo1 :



1) - La moyenne de classe : $\mu_1 = (\underbrace{9{,}09, 2}_{*})^T$, $\mu_2 = (\underbrace{7{,}75, 2}_{\blacksquare})^t$

*) Le point $x = (9{,}6\ , 2)$ est classé dans C_1 parcequ'il est plus proche à sa moyenne c.a.d. $\|x - \mu_1\| < \|x - \mu_2\|$ il est clair que cette approche n'est pas la meilleure !

2/ Si $K = 1$, le plus proche voisin à x dans l'ensemble des données d'apprentissage appartient à C_1 donc $x \in C_1$.

3/ Si $K = 3$, les 3 ppV à x sont ceux qui ont l'abssice $x = 10$, la majorité de ces points $\in C_2 \Rightarrow x \in C_2$

Exo 2 :

1) L'entropie de 2 évenement equiprobables,

$X = \{\frac{1}{2}, \frac{1}{2}\}$, $H(x) = -\left[\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log\left(\frac{1}{2}\right)\right] = 1$

2/ $X = \left\{\frac{1}{8}, \frac{3}{4}, \frac{1}{16}, \frac{1}{16}\right\}$.

$H(x) = -\left[\frac{1}{8}\log_2\left(\frac{1}{8}\right) + \frac{3}{4}\log_2\left(\frac{3}{4}\right) + \frac{1}{16}\log_2\left(\frac{1}{16}\right) + \frac{1}{16}\log_e\left(\frac{1}{16}\right)\right]$

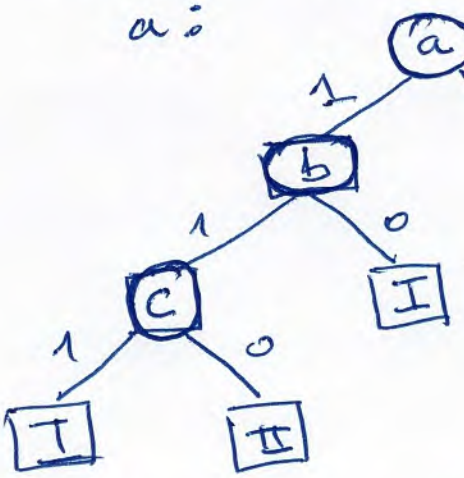
$H(x) = 1{,}18625$

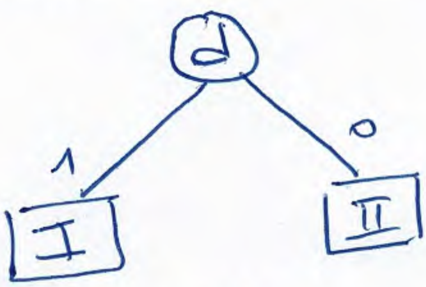
3/

a	b	c	d	class
1	0	1	1	I
1	1	0	0	II
0	1	1	0	II
1	1	1	1	I

- construction de l'arbre de decision à partir de la caract

a :





figure(*)
(reponse 4)

4/ Construction de l'arbre à partir de mesure du contenu d'information pour chaque caract. a, b, c.

$$I(C,a) = P(C=I, a=1)\log_2 \frac{P(C=I, a=1)}{P(C=I)\,P(a=1)} +$$

$$P(C=I, a=0)\log_2 \frac{P(C=I, a=0)}{P(C=I)\,P(a=0)} + P(C=II, a=1)\log_2 \frac{P(C=II, a=1)}{P(C=II)\,P(a=1)}$$

$$+ P(C=II, a=0)\log_2 \frac{P(C=II, a=0)}{P(C=II)\,P(a=0)} =$$

$$= 0{,}5\log_2 \frac{0{,}5}{0{,}5\times 0{,}75} + 0 + 0{,}25\log_2 \frac{0{,}25}{0{,}5\times 0{,}75} + 0{,}25\log_2 \frac{0{,}25}{0{,}5\times 0{,}25}$$

$$= 0{,}3113$$

$$I(C,b) = 0{,}25\log_2\left(\frac{0{,}25}{0{,}5\times 0{,}75}\right) + 0{,}25\log_2\left(\frac{0{,}25}{0{,}5\times 0{,}25}\right) +$$

$$0{,}5\log_2\left(\frac{0{,}5}{0{,}5\times 0{,}75}\right) + 0 = 0{,}3113$$

$$I(C,c) = 0{,}5\log_2\left(\frac{0{,}5}{0{,}5\times 0{,}75}\right) + 0 + 0{,}25\log_2\left(\frac{0{,}25}{0{,}5\times 0{,}75}\right) + 0$$

$$= 0{,}3113$$

$$I(c,d) = 0{,}5 \log_2\left(\frac{0{,}5}{0{,}5\times 0{,}5}\right) + 0 + 0 + 0{,}5 \log_2\left(\frac{0{,}5}{0{,}5\times 0{,}5}\right) = 1$$

La caract. d qui a un contenu d'info 1 donne plus d'info. pour déterminer la classe. figure (*)

Exo 3 :

Dans un problème de décision binaire, le résultat de la classification sont soit positifs ou négatifs. La matrice de confusion est :

résultat	réel Positive	réel négative
positive	TP	FP
negative	FN	TN

TP = True positives
FP false 〃
TN True negatives
FN false 〃

Nous pouvons caractériser les erreurs de classification en calculant :

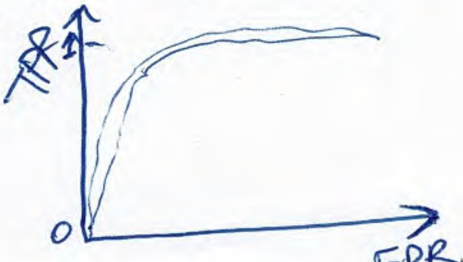
$$\text{Rappel} = \text{Recall} = \frac{TP}{TP+FN}, \quad \underset{R = \text{rate}}{TPR} = \frac{TP}{TP+FN}$$

$$\text{Précision} = \text{Precision} = \frac{TP}{TP+FP} \qquad FPR = \frac{FP}{FP+TN}$$

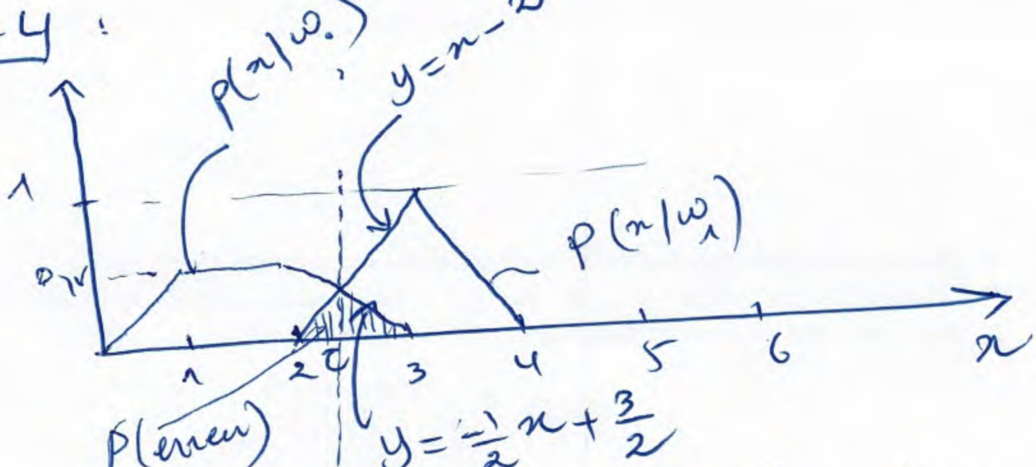
1/ $$\text{precision} = \frac{TP}{TP+FP} = \frac{300}{300+50} = \frac{300}{350} = \frac{6}{7}$$

$$\text{recall} = \text{rappel} = \frac{TP}{TP+FN} = \frac{300}{300+200} = \frac{300}{500} = \frac{3}{5}$$

2/ Pour tracer la courbe ROC, on prend le TPR sur y et le FPR sur X



Exercice 4 :



- La règle de décision à erreur minimale classe x à la classe i pour laquelle $p(\omega_i|x)$ est maximale donc dans la région R_1 $p(\omega_0|x) < p(\omega_1|x)$

il est clair dans la figure que :

$$R_1 = \{x \mid x > \tau\} \text{ ou } 2 < \tau < 3 \quad \text{et } R_0 = R \setminus R_1$$

$$p(\omega_0|x) < p(\omega_1|x)$$

$$\Rightarrow p(\omega_0)p(x|\omega_0) < p(\omega_1)p(x|\omega_1)$$

$$\Rightarrow \frac{3}{5}\left(\frac{-1}{2}x+\frac{3}{2}\right) < \frac{2}{5}(x-2) \Rightarrow x > \frac{17}{7}$$

$$\tau = \frac{17}{7}$$

$$p(\text{erreur}) = \int_{R_0} p(x|\omega_1)p(\omega_1)\,dx + \int_{R_1} p(\omega_0)p(x|\omega_0)\,dx$$

$$= \int_2^{\tau} \frac{2}{5}(x-2)\,dx + \int_{\tau}^{3}\left(-\frac{1}{2}x+\frac{3}{2}\right)dx = \Big[\frac{x^2}{5}-\frac{4}{5}x\Big]_2^{\tau} + \Big[\frac{-3x^2}{20}+\frac{9}{10}x\Big]_{\tau}^{3} = \frac{21}{245}$$