Examen partiel 2 - Vendredi 4 mai 2018 - Durée : 60 min
Aucun document, pas de téléphone, pas de calculatrice.

Nom : SKETCH OF Prénom : The Solution Signature :

Exercice 1 :	Exercice 2 :	Total /20 :

Exercise 1 (Classifieur Naïf de Bayes, ≃ 10 pts)
On considère un problème de classification binaire où l'étiquette de classe est notée $Y \in \{0,1\}$ et chaque exemple d'entraînement X contient 2 attributs binaires $X_1, X_2 \in \{0,1\}$. Dans ce problème, on admet que X_1 et X_2 sont conditionnellement indépendants étant donné Y, c'est-à-dire $\Pr_V(X_1, X_2|Y) = \Pr_V(X_1|Y)\Pr_V(X_2|Y)$, que les probabilités a priori sont $\Pr_V(Y=0) = \Pr_V(Y=1) = 0.5$, et que les probabilités conditionnelles sont :

$\Pr_V(X_1\|Y)$	$X_1 = 0$	$X_1 = 1$
$Y = 0$	0.7	0.3
$Y = 1$	0.2	0.8

$\Pr_V(X_2\|Y)$	$X_2 = 0$	$X_2 = 1$
$Y = 0$	0.9	0.1
$Y = 1$	0.5	0.5

La notation $\Pr_V(\cdot)$ indique qu'il s'agit de la vraie distribution des variables aléatoires. La distribution des variables aléaoires modélisée par le test naïf de Bayes sera notée $\Pr_{NB}(\cdot)$.
Le taux d'erreur d'un classifieur est la probabilité que le classifieur fournisse une prédiction incorrecte pour une observation donnée Z : si Y est la vraie étiquette et $Z = Z(X_1, X_2)$ est l'observation utilisée pour prendre une décision ($Z = X_1$, $Z = X_2$ ou $Z = (X_1, X_2)$ dans la suite de l'exercice) et si $\hat{Y}(Z)$ correspond à la classe prédite, alors le taux d'erreur est

$$\Pr_V(Y \neq \hat{Y}(Z)) = \sum_{x_1=0}^{1}\sum_{x_2=0}^{1} \Pr_V\left(X_1 = x_1, X_2 = x_2, Y = 1 - \hat{Y}(Z(x_1, x_2))\right).$$

1. Rappeler la règle de décision du Maximum A Posteriori (MAP) du test naïf de Bayes qui exploite uniquement l'attribut X_1. Montrer que cette règle de décision nécessite uniquement de connaître la distribution conjointe $\Pr_{NB}(X_1, Y)$.

MAP Rule

Decide 0 if $\Pr_{NB}(Y=0|X_1=x_1) \geq \Pr_{NB}(Y=1|X_1=x_1)$

But $\Pr_{NB}(Y=i|X_1=x_1) = \dfrac{\Pr_{NB}(X_1=x_1|Y=i)\Pr_{NB}(Y=i)}{\Pr_{NB}(X_1=x_1)}$

So the MAP rule is equivalent to

Decide 0 if $\boxed{\Pr_{NB}(X_1=x_1|Y=0)\Pr_{NB}(Y=0)} = \Pr_{NB}(X_1=x_1, Y=0)$

$\geq \underbrace{\Pr_{NB}(X_1=x_1|Y=1)\Pr_{NB}(Y=1)}_{\Pr_{NB}(X_1=x_1, Y=1)}$

2. Écrire le test naïf de Bayes qui exploite uniquement l'attribut X_1 pour les 2 valeurs possibles de X_1. Écrire les réponses dans le tableau suivant :

X_1	$\Pr_{NB}(X_1, Y=0)$	$\Pr_{NB}(X_1, Y=1)$	$\hat{Y}(X_1)$
0	$0.7/2 = 0.35$	0.1	0
1	$0.3/2 = 0.15$	0.4	1

Justifier vos calculs ci-dessous :

$$\Pr_{NB}(X_1 = 0, Y=0) = \Pr_{NB}(X_1 = 0 \mid Y=0)\Pr_{NB}(Y=0)$$
$$= \Pr_V(X_1=0 \mid Y=0)\Pr_V(Y=0)$$
$$= 0.7 \times 1/2 = 0.35$$

3. Calculer le taux d'erreur p_1 du test naïf de Bayes qui exploite seulement X_1.

$$p_1 = \Pr_V(X_1=0, Y=1) + \Pr_V(X_1=1, Y=0)$$
$$= 0.1 + 0.15 = 0.25$$

4. Écrire le test naïf de Bayes qui exploite uniquement l'attribut X_2 pour les 2 valeurs possibles de X_2. Écrire les réponses dans le tableau suivant :

X_2	$\Pr_{NB}(X_2, Y=0)$	$\Pr_{NB}(X_2, Y=1)$	$\hat{Y}(X_2)$
0	0.45	0.25	0
1	0.05	0.25	1

Justifier vos calculs ci-dessous :

$$\Pr_{NB}(X_2 = 0, Y=0) = \Pr_V(X_2=0 \mid Y=0)\Pr_V(Y=0)$$
$$= 0.9 \times \frac{1}{2} = 0.45$$

5. Calculer le taux d'erreur p_2 du test naïf de Bayes qui exploite seulement X_2.

$P_2 = 0.25 + 0.05 = 0.3.$

6. Écrire le test naïf de Bayes pour les 4 configurations possibles de X_1, X_2. Écrire les réponses dans le tableau suivant :

X_1	X_2	$\Pr_{NB}(X_1, X_2, Y=0)$	$\Pr_{NB}(X_1, X_2, Y=1)$	$\hat{Y}(X_1, X_2)$
0	0	0.315	$0.2 \times 0.5 \times 0.5 = 0.05$	0
0	1	$0.7 \times 0.1 \times 0.5 = 0.035$	$0.2 \times 0.5 \times 0.5 = 0.05$	1
1	0	$0.3 \times 0.9 \times 0.5 = 0.135$	$0.8 \times 0.5 \times 0.5 = 0.2$	1
1	1	$0.3 \times 0.1 \times 0.5 = 0.015$	$0.8 \times 0.5 \times 0.5 = 0.2$	1

Justifier vos calculs ci-dessous :

$$\Pr_{NB}(X_1=0, X_2=0, Y=0) = \Pr_{NB}(X_1=0, X_2=0 \mid Y=0)\Pr_{NB}(Y=0)$$
$$= \Pr_{NB}(X_1=0 \mid Y=0)\Pr_{NB}(X_2=0 \mid Y=0)\Pr_{NB}(Y=0)$$
$$= \Pr_V(X_1=0 \mid Y=0)\Pr_V(X_2=0 \mid Y=0)\Pr_V(Y=0)$$
$$= 0.7 \times 0.9 \times \frac{1}{2} = \frac{0.63}{2} = 0{,}315$$

7. Calculer le taux d'erreur $p_{1,2}$ du test naïf de Bayes qui exploite X_1 et X_2.

$$P_{1,2} = 0.05 + 0.035 + 0{,}135 + 0.015$$
$$= 0.235$$

8. On suppose maintenant créer un nouveau attribut X_3 qui est une copie déterministe de X_2. Écrire le test naïf de Bayes pour les 8 configurations possibles de X_1, X_2 et X_3.

X_1	X_2	X_3	$\Pr_{NB}(X_1, X_2, X_3, Y = 0)$	$\Pr_{NB}(X_1, X_2, X_3, Y = 1)$	$\hat{Y}(X_1, X_2, X_3)$
0	0	0	0.7×0.9×0.9×0.5	0.2×0.5×0.5×0.5	0
0	0	1	0.7×0.9×0.1×0.5	0.2×0.5×0.5×0.5	0
0	1	0	0.7×0.1×0.9×0.5	0.2×0.5×0.5×0.5	0
0	1	1	0.7×0.1×0.1×0.5	0.2×0.5×0.5×0.5	1
1	0	0	0.3×0.9×0.9×0.5	0.8×0.5×0.5×0.5	0
1	0	1	0.3×0.9×0.1×0.5	0.8×0.5×0.5×0.5	1
1	1	0	0.3×0.1×0.9×0.5	0.8×0.5×0.5×0.5	1
1	1	1	0.3×0.1×0.1×0.5	0.8×0.5×0.5×0.5	1

Justifier (brièvement) vos calculs ci-dessous :

Just assume X_2 and X_3 are independent and do the same calculation as before

9. Calculer le taux d'erreur $p_{1,2,3}$ du test naïf de Bayes qui exploite X_1, X_2 et X_3.

$p_{1,2,3} = 0.2\times0.5\times0.5 + 0.7\times0.1\times0.5$
$+ 0.8\times0.5\times0.5 + 0.3\times0.1\times0.5$
$= 0.5\times(0.1 + 0.07 + 0.4 + 0.03) = 0.3$

10. Pouvez-vous expliquer la différence entre le taux d'erreur $p_{1,2}$ basé sur (X_1, X_2) et le taux d'erreur $p_{1,2,3}$ basé sur (X_1, X_2, X_3) ?

The independence assumption is wrong
Hence the Bayes test with $X_1, X_2 = X_3$ is worst than the Bayes test with just X_1 and X_2.
Some cases do not occur (exp: $X_1=0, X_2=0, X_3=1$).

Exercise 2 (Régression Logistique, ≃ 10 pts)
106 patients atteints par la grippe ont été hospitalisés. On note Y_k la variable indicatrice valant 1 si le patient k est décédé et 0 sinon. Un facteur de risque X_k a été relevé : $X_k = 1$ si le patient k a plus de 65 ans et $X_k = 0$ si la personne a 65 ans ou moins.
Le tableau de contingence résumant les données observées est donné ci-dessous :

	$Y=0$	$Y=1$	
$X=0$	82	14	96
$X=1$	3	7	10
	85	21	106

On note n_{ij} le nombre de patients tels que $X = i$ et $Y = j$ pour $i = 0,1$ et $j = 0,1$, soit $n_{00} = 82$, $n_{01} = 14$, $n_{10} = 3$ et $n_{11} = 7$.
Nous voulons utiliser un modèle de régression logistique pour modéliser le lien entre la variable Y et le facteur de risque X.

1. Notons $P(X) = \Pr(Y = 1|X)$ la probabilité de décès connaissant le facteur de risque X. Décrire le modèle mathématique entre $P(X)$ et X sous la forme $\text{logit}(P(X)) = \cdots$ où vous définirez la fonction “logit”. Le coefficient associé à X sera noté a et le terme constant sera noté b.

$\text{logit}\, P(X) = aX + b$

$\text{logit}\,(p) = \ln \frac{p}{1-p}$

2. Quelle est l'inverse de la fonction “logit(u)” pour $u \in]0,1[$? Prouver le résultat annoncé.

Let $u \in (0,1)$ and $z \in \mathbb{R}$

$z = \text{logit}\,(u) = \ln \frac{u}{1-u}$

$\iff e^z = \frac{u}{1-u} \iff (1-u)\, e^z = u$

$\iff e^z = u\,(1 + e^z) \iff u = \frac{1}{1+e^{-z}} = \sigma(z)$

since $\underline{1+e^z > 0}$

3. On note $p_0 = P(0)$ et $p_1 = P(1)$. Exprimer p_0 et p_1 en fonction de a et b.

We have $P(X) = \sigma(aX + b)$

$p_0 = P(0) = \sigma(b) = \frac{1}{1+e^{-b}}$

$p_1 = P(1) = \sigma(a+b) = \frac{1}{1+e^{-a-b}}$

4. Montrer que la log-vraisemblance $L(a,b)$ du modèle logistique s'exprime sous la forme

$$L(a,b) = -n_{00}\ln\left(1+e^{b}\right) - n_{01}\ln\left(1+e^{-b}\right) - n_{10}\ln\left(1+e^{a+b}\right) - n_{11}\ln\left(1+e^{-a-b}\right).$$

$$\mathcal{L}(a,b) = \ln \Pr(Y=0 \mid X=0)^{n_{00}} + \ln \Pr(Y=0 \mid X=1)^{n_{10}} + \ln \Pr(Y=1 \mid X=0)^{n_{01}} + \ln \Pr(Y=1 \mid X=1)^{n_{11}}$$

$$\Pr(Y=0 \mid X=0) = 1 - \Pr(Y=1 \mid X=0) = 1 - p_0 = 1 - \frac{1}{1+e^{-b}} = \frac{1}{1+e^{b}}$$

$$\Pr(Y=0 \mid X=1) = 1 - p_1 = \frac{1}{1+e^{a+b}}$$

Hence

$$\mathcal{L}(a,b) = -n_{00}\ln(1+e^{b}) - n_{01}\ln(1+e^{-b}) - n_{10}\ln(1+e^{a+b}) - n_{11}\ln e^{-a-b}$$

5. Calculer le gradient $\nabla L(a,b)$ de $L(a,b)$ par rapport au couple de variables (a,b).

$$\frac{\partial \mathcal{L}(a,b)}{\partial a} = \boxed{-n_{10}\frac{e^{a+b}}{1+e^{a+b}} + n_{11}\frac{e^{-a-b}}{1+e^{-a-b}}}$$

Note that it is the same term.

$$\frac{\partial \mathcal{L}(a,b)}{\partial b} = -n_{00}\frac{e^{b}}{1+e^{b}} + n_{01}\frac{e^{-b}}{1+e^{-b}} \boxed{- n_{10}\frac{e^{a+b}}{1+e^{a+b}} + n_{11}\frac{e^{-a-b}}{1+e^{-a-b}}}$$

$$\nabla \mathcal{L}(a,b) = \begin{pmatrix} \dfrac{\partial \mathcal{L}(a,b)}{\partial a} \\[2ex] \dfrac{\partial \mathcal{L}(a,b)}{\partial b} \end{pmatrix}$$

6. Calculer $\hat{a}$ et $\hat{b}$ qui maximise $L(a,b)$ à partir du calcul du gradient de la question précédente.

$\hat a, \hat b$ satisfy $\nabla \mathcal{L}(\hat a, \hat b) = 0$

(1) $$-n_{10}\frac{e^{\hat a+\hat b}}{1+e^{\hat a+\hat b}} + n_{11}\frac{e^{-\hat a-\hat b}}{1+e^{-\hat a-\hat b}} = 0$$

$$n_{10}\frac{e^{\hat a+\hat b}}{1+e^{\hat a+\hat b}} = n_{11}\frac{1}{1+e^{\hat a+\hat b}} \Rightarrow e^{\hat a+\hat b} = \frac{n_{11}}{n_{10}}$$

(2) $$-n_{00}\frac{e^{\hat b}}{1+e^{\hat b}} + n_{01}\frac{e^{-\hat b}}{1+e^{-\hat b}} = 0 \Rightarrow -n_{00}e^{\hat b} + n_{01} = 0$$

$$\Rightarrow e^{\hat b} = \frac{n_{01}}{n_{00}}$$

$$\Rightarrow \hat b = \ln\left[\frac{n_{01}}{n_{00}}\right]$$

(3) $$e^{\hat a} = \frac{n_{11}}{n_{10}} \times \frac{n_{00}}{n_{01}} \Rightarrow \hat a = \ln\left[\frac{n_{11}\, n_{00}}{n_{10}\, n_{01}}\right]$$

$$= \ln\left[\frac{n_{11}}{n_{10}}\right] - \hat b$$