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

Nom : Prénom : 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 $\Pr_{NB}(X_1, Y)$.

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			
1			

Justifier vos calculs ci-dessous :

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

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			
1			

Justifier vos calculs ci-dessous :

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

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	1			
1	0			
1	1			

Justifier vos calculs ci-dessous :

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

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	$\mathrm{Pr}_{NB}(X_1, X_2, X_3, Y = 0)$	$\mathrm{Pr}_{NB}(X_1, X_2, X_3, Y = 1)$	$\hat{Y}(X_1, X_2, X_3)$
0	0	0			
0	0	1			
0	1	0			
0	1	1			
1	0	0			
1	0	1			
1	1	0			
1	1	1			

Justifier (brièvement) vos calculs ci-dessous :

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

Exercise 2 (Régression Logistique, $\simeq$ 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.

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

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

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

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

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

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