



Prénom(s) :

Numéro
Inscription :

Né(e) le :

(Le numéro est celui qui figure sur la convocation ou la feuille d'émargement)

(Remplir cette partie à l'aide de la notice)

Concours / Examen :

Section/Sécialité/Série :

Epreuve : EXAMEN DATA Mining

Matière :

Session :

CONSIGNES

- Remplir soigneusement, sur CHAQUE feuille officielle, la zone d'identification en MAJUSCULES.
- Ne pas signer la composition et ne pas y apporter de signe distinctif pouvant indiquer sa provenance.
- Numéroté chaque PAGE (cadre en bas à droite de la page) et placer les feuilles dans le bon sens et dans l'ordre.
- Rédiger avec un stylo à encre foncée (bleue ou noire) et ne pas utiliser de stylo plume à encre claire.
- N'effectuer aucun collage ou découpage de sujets ou de feuille officielle. Ne joindre aucun brouillon.

Correction Examen 2023 proposé par un élève

* Niveau de confiance de la réponse : * peu confiant de la validité de la réponse ** assez confiant

Exercice 1: *** très confiant de la validité de la réponse

1- On observe des p-value supérieures à 0,05 \Rightarrow on constate que les coefficients ne sont pas significatives au seuil de significativité de 5% mais on peut accepter la constante à un seuil de significativité de 10% et on peut accepter le coefficient de X_2 car il est proche de 0,1

Quant au test Anova, on remarque que la variable X_2 a un effet significatif sur Y , par contre X_1 et X_3 ont une p-value qui est supérieure à 0,05. *

2- On a: $\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$ tq p : la probabilité d'appartenir à la classe 1.

$$\Rightarrow p = \frac{1}{1 + e^{-Z}} \quad \text{tq } Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad \text{or } \beta_0 = 14,49379, \beta_1 = -0,1256, \beta_2 = -0,063 \text{ et } \beta_3 = 1,71$$

d'où un individu ayant $X_1 = 50, X_2 = 126$ et $X_3 = 1$ aura la probabilité 0,179 d'appartenir à la classe 1 donc sa classe d'affectation est 1. ***

3- ***

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
Réel	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
prédit	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

- TP = 3

TP: True positive: Réel 1 et prédit 1

• FP = 1

FP: False positive: Réel 0 et prédit 1

• TN = 13

TN: True négative: Réel 0 et prédit 0

• FN = 3

FN: False négative: Réel 1 et prédit 0

$$\text{Taux de bon choix} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$= \frac{16}{20} = \frac{4}{5} = 0,8 = 80\%$$

.... /

4- $OR_{X_1} = e^{\beta_1} = 0,8819 \Rightarrow$ Quand X_1 augmente d'une unité, la probabilité que l'individu \in à la classe 1 diminue de 12%, la probabilité devient 88,19% de sa valeur précédente. ***

$OR_{X_2} = e^{\beta_2} = 0,9384 \Rightarrow$ Quand X_2 augmente d'une unité, la probabilité que l'individu \in à la classe 1 diminue de 7%. ***

$OR_{X_3} = e^{\beta_3} = 5,823 \Rightarrow$ Etre de la classe 1 signifie 6 fois plus de chance d'appartenir à la classe 1 de X_3 . *

5- modèle - simple $\leftarrow \text{glm}(df\$Y \sim df\$X_1, \text{"binomial"})$

pr. $\text{step}(\text{modèle - simple}, \text{scope} = (\sim df\$X_1 + df\$X_2 + df\$X_3), \text{dir} = \text{"forward"})$ *

Exercice 2: ***

Si 1- On a 5 règles:

1) Si petal length $> 2,45$ et inférieur ou égale à 4,95, et sepal length $> 5,95$, il s'agit de Virginica

2) Si petal length est supérieur à 4,95 et petal width $> 1,75 \rightarrow$ Virginica supérieure

3) Si petal length supérieur à 2,45 et inférieur ou égale à 4,95, et petal width inférieure ou égale à 1,65 \rightarrow Virginica.

4) Si petal length $> 4,95$ et petal width inférieure ou égale à 1,55 \rightarrow Virginica

5) Si petal length $> 4,95$ et petal width $> 1,55$ mais inférieure ou égale à 1,75, et sepal length $> 6,95 \rightarrow$ Virginica.

2- (a) - X

(K) - 5

(b) - Y,

(P) - grid-search

(c) - 0,2

(m) - X-train, Y-train

(d) - 42

(n) - Y-test, Y-pred

(e) - 0,01, 0,05, 0,1

(f) - 4, 5, 7

(g) - 3, 4, 5

(h) - 'gini', 'entropy'

(i) - clf

(j) - params

.... /