

QCM
Examen du 04/04/2023

Instructions :

- Le sujet comprend 3 exercices pour 23 questions au total. Les questions faisant apparaître le symbole ♣ peuvent présenter plusieurs bonnes réponses. Les autres ont une unique bonne réponse. Des points négatifs pourront être affectés à de *mauvaises* réponses.
- Seul le questionnaire à la 4ème feuille (page 7) est à rendre. Vous commencerez par renseigner votre nom et prénom dans la case prévue ainsi que votre **numéro étudiant Rennes 2** (1 case à colorier par colonne).
- Il faut **colorier** les cases correspondants aux bonnes réponses (sur la page 7), mettre une croix dans la case n'est **pas suffisant**. Les cases devront être **coloriées avec un stylo noir** (pas de crayon papier, de stabilo...).
- Tous les sujets sont différents, si vous commettez une erreur en coloriant la case, vous pouvez demander un autre sujet complet, il faudra dans ce cas recommencer le devoir au début.
- Le barème sera effectué de la façon suivante :
 - Aucune case coloriée entraînera une note de 0 sur la question.
 - Pour les questions à une seule bonne réponse (sans le symbole ♣), un nombre de points sera affecté (par exemple +2) si la bonne case est cochée. Un nombre de points sera retranché (par exemple -1) si une mauvaise case est coloriée ou si plusieurs cases sont coloriées.
 - Pour les questions avec plusieurs bonnes réponses (avec le symbole ♣), un nombre de points (par exemple +0.5) sera affecté pour chaque bonne réponse coloriée et pour chaque mauvaise réponse non coloriée. Un nombre de points (par exemple -0.5) sera retranché pour chaque mauvaise réponse coloriée et pour chaque bonne réponse non coloriée.
 - Les nombres de points ajoutés et retranchés ne sont pas forcément fixes et pourront dépendre de la réponse.
- La correction étant automatique, un non respect des consignes aura forcément un impact sur la note finale.

Durée : 1 heure 20 minutes.

Exercice 1. Les questions de cet exercice portent sur des "bases" du cours.

Question 1 On souhaite faire une classification ascendante hiérarchique (CAH) sur des variables qui ne sont pas à la même échelle. Est-il préférable de réduire les variables avant l'analyse ?

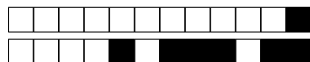
- ☐ A Non (il est préférable de faire l'analyse sur les données brutes) ☐ B Oui

Question 2 On souhaite identifier des chiffres à partir d'images. Pour ce faire, on dispose de données composées d'images sur lesquelles on a identifié le chiffre associé à l'image. Le problème est de prédire automatiquement le chiffre à une nouvelle image. Il s'agit d'un problème d'apprentissage

- ☐ A non supervisé ☐ B supervisé ☐ C Aucune de ces réponses n'est correcte.

Question 3 Le machine learning est un champ d'étude de l'intelligence artificielle, c'est-à-dire qu'il fait partie de l'intelligence artificielle.

- ☐ A Faux ☐ B Vrai



Question 4 La classification ascendante hiérarchique (CAH) permet d'expliquer (et prédire) une variable Y par des variables V_1, \dots, V_p .

- ☐ A Faux ☐ B Vrai

Question 5 Un algorithme d'apprentissage non supervisé permet d'expliquer (et prédire) une variable Y par des variables V_1, \dots, V_p .

- ☐ A Faux ☐ B Vrai

Question 6 ♣ Soit $x = (x_1, \dots, x_p)$ et $x' = (x'_1, \dots, x'_p)$ deux individus. La distance euclidienne entre x et x' est égale à

- | | |
|---|---|
| <input type="checkbox"/> A $\sqrt{\sum_{j=1}^p x_j - x'_j ^2}$ | <input type="checkbox"/> D $\sqrt{\sum_{j=1}^p (x_j - x'_j)}$ |
| <input type="checkbox"/> B $\sqrt{\sum_{j=1}^p x'_j - x_j }$ | <input type="checkbox"/> E $\sqrt{\sum_{j=1}^p (x_j - x'_j)^2}$ |
| <input type="checkbox"/> C $\sqrt{\sum_{j=1}^p (x'_j - x_j)^2}$ | <input type="checkbox"/> F Aucune de ces réponses n'est correcte. |

Exercice 2. Les questions suivantes portent sur l'algorithme des k plus proches voisins.

Question 7 ♣ On considère la méthode de validation hold out présentée dans le cours. On rappelle que cette méthode consiste à séparer les données en une base d'apprentissage \mathcal{D}_{app} et une base test $\mathcal{D}_{\text{test}}$. Cochez la (ou les) assertion(s) vraie(s).

- | | |
|--|---|
| <input type="checkbox"/> A Les valeurs prédites sont confrontées aux valeurs observées sur la base d'apprentissage | <input type="checkbox"/> E La valeur de k renvoyée par la méthode est toujours entre $n/2$ et n |
| <input type="checkbox"/> B Cette méthode permet d'éviter le sur-apprentissage | <input type="checkbox"/> F La base test est généralement plus grande que celle d'apprentissage |
| <input type="checkbox"/> C Cette méthode est généralement utilisée pour choisir la meilleure valeur de k | <input type="checkbox"/> G La valeur de k renvoyée par la méthode est toujours entre 1 et 10 |
| <input type="checkbox"/> D On entraîne l'algorithme des k plus proches voisins sur la base d'apprentissage | <input type="checkbox"/> H Aucune de ces réponses n'est correcte. |

Question 8 On dispose de n observations et on effectue l'algorithme des k plus proches voisins avec $k = 1$. Cochez la (ou les) assertion(s) vraie(s).

- | | |
|--|---|
| <input type="checkbox"/> A L'algorithme prédit toujours la même valeur | <input type="checkbox"/> B L'algorithme sur-apprend |
| <input type="checkbox"/> C Aucune de ces réponses n'est correcte. | |

Question 9 On dispose de n observations et on effectue l'algorithme des k plus proches voisins avec $k = n$. Cochez la (ou les) assertion(s) vraie(s).

- | | |
|--|---|
| <input type="checkbox"/> A L'algorithme prédit toujours la même valeur | <input type="checkbox"/> B L'algorithme sur-apprend |
| <input type="checkbox"/> C Aucune de ces réponses n'est correcte. | |



Question 10 Quelles que soient les données,

- ☐ A la valeur de k n'a pas d'importance, l'algorithme fonctionne toujours bien
- ☐ B il faut toujours choisir k le plus grand possible pour que l'algorithme soit efficace
- ☐ C il faut toujours choisir k le plus petit possible pour que l'algorithme soit efficace
- ☐ D Aucune de ces réponses n'est correcte.

Dans les 6 questions suivantes (questions 11 à 16), on considère les données suivantes

V_1	V_2	Y
1	2	mauvais
3	1	bon
2	3	bon
4	5	mauvais
1	3	mauvais

Le problème est d'expliquer/prédire la variable Y par V_1 et V_2 . On note $x = (3, 2)$ un nouvel individu pour lequel on a observé V_1 (qui vaut 3) et V_2 (qui vaut 2). On souhaite appliquer la méthode des k plus proches voisins avec la distance euclidienne pour prédire la valeur de Y de ce nouvel individu. Les variables étant à la même échelle, il n'est pas nécessaire de réduire les données.

Question 11 Les variables explicatives sont

- ☐ A quantitatives ☐ B qualitatives

Question 12 La variable à expliquer est

- ☐ A qualitative ☐ B quantitative

Question 13 Le groupe prédit par l'algorithme du 1 plus proche voisin est

- ☐ A On ne peut pas savoir ☐ B mauvais ☐ C bon

Question 14 Le groupe prédit par l'algorithme du 3 plus proches voisins est

- ☐ A bon ☐ B mauvais ☐ C On ne peut pas savoir

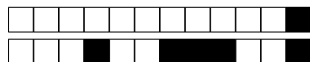
Question 15 On s'intéresse toujours au nouvel individu x mais on veut maintenant prédire la probabilité qu'il soit dans le groupe bon. En utilisant la règle des 3 plus proches voisins cette prévision est égale à

- ☐ A 1/2 ☐ B 2/3 ☐ C 1 ☐ D 0 ☐ E 3/4 ☐ F 1/4 ☐ G $\sqrt{2}$
- ☐ H Aucune de ces réponses n'est correcte.

Question 16 Même question que la précédente mais avec la règle des 2 plus proches voisins

- ☐ A 2/3 ☐ B $\sqrt{2}$ ☐ C 3/4 ☐ D 1/4 ☐ E 0 ☐ F 1 ☐ G 1/2
- ☐ H Aucune de ces réponses n'est correcte.

Exercice 3. On souhaite faire une classification ascendante hiérarchique pour regrouper 8 individus notés A, B, C... Pour ce faire, on a réduit les données et on a calculé la matrice des distances (comme en cours). Elle est donnée ci-après :



	A	B	C	D	E	F	G
B	4.15						
C	4.36	4.07					
D	4.05	3.28	0.80				
E	5.15	4.01	1.00	1.11			
F	3.75	1.03	3.04	2.25	3.04		
G	4.09	0.50	3.59	2.80	3.51	0.58	
H	5.10	4.50	0.74	1.32	0.67	3.50	4.01

Question 17 La distance euclidienne entre C et G est égale à

- ☐ A 0
 ☐ B 4.15
 ☐ C 3.59
 ☐ D 0.58
 ☐ E 0.50
 ☐ F 1.795
☐ G 0.74
☐ H Aucune de ces réponses n'est correcte.

Question 18 La distance du saut minimum entre A et le groupe {E, G} est égale à

- ☐ A 3.51
 ☐ B 5.15
 ☐ C 4.62
 ☐ D 3.58
 ☐ E 4.09
 ☐ F 4.15
☐ G 3.75
☐ H Aucune de ces réponses n'est correcte.

Question 19 La distance du saut minimum entre le groupe {A, B, C, D} et le groupe {E, F, G, H} est égale à

- ☐ A 4.09
 ☐ B 0.50
 ☐ C 3.75
 ☐ D 3.58
 ☐ E 5.15
 ☐ F 3.51
☐ G 4.62
☐ H Aucune de ces réponses n'est correcte.

Question 20 On rappelle qu'à la première étape de la CAH chaque observation forme un groupe. L'étape suivante va consister à assembler les observations

- ☐ A C et E
 ☐ B A et E
 ☐ C F et G
 ☐ D B et G
 ☐ E C et H
☐ F E et H
☐ G E et F
☐ H Aucune de ces réponses n'est correcte.

Question 21 Parmi les 4 dendrogrammes de la figure 1, lequel correspond à celui de la CAH sur les données de cet exercice :

- ☐ A En haut à gauche
 ☐ C En bas à droite
☐ B En haut à droite
☐ D En bas à gauche

Question 22 En vous basant sur le (bon) dendrogramme de la figure 1, combien choisissez-vous de groupes ?

- ☐ A 1
 ☐ B 7
 ☐ C Plus de 8
 ☐ D 2
 ☐ E 4
 ☐ F 0
 ☐ G 8
☐ H 3
☐ I 6
☐ J 5

Question 23 ♣ La figure 2 propose un dendrogramme issue d'une CAH faite sur un autre jeu de données que celui de l'exercice. Cochez la (ou les) assertion(s) vraie(s).

- ☐ A Toutes les observations sont confondues
 ☐ E La distance entre B et E est la même que la distance entre A et D
☐ B Le bon nombre de groupes pour ces données vaut 4
 ☐ F La distance entre B et E est la même que la distance entre A et C
☐ C La distance entre D et C est la même que la distance entre A et C
 ☐ G Toutes les observations sont à la même distance les unes des autres
☐ D Le bon nombre de groupes pour ces données vaut 1 ou 5
 ☐ H Aucune de ces réponses n'est correcte.

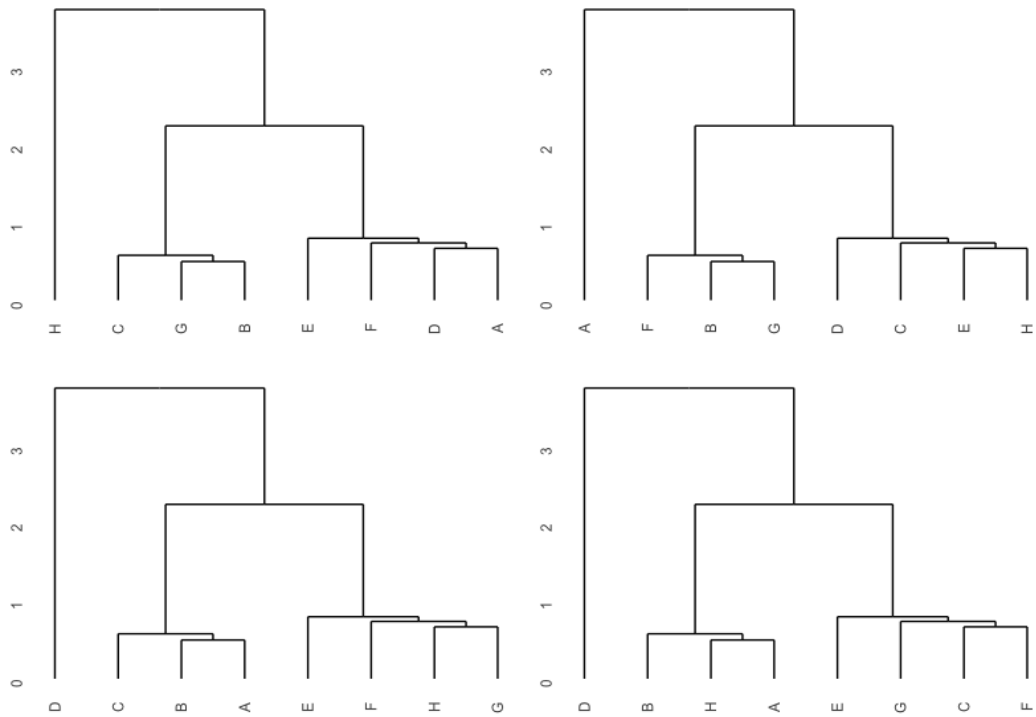


Figure 1: 4 exemples de dendrogrammes - questions 21 et 22

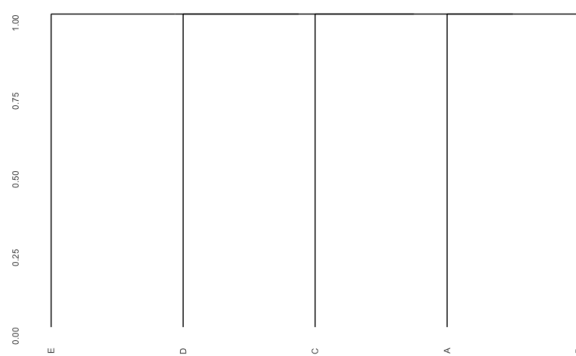
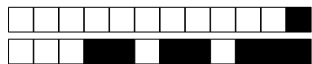


Figure 2: Dendrogramme de la question 23



+1/6/55+



Feuille de réponses :

0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9

← codez votre numéro d'étudiant ci-contre, et inscrivez votre nom et prénom ci-dessous.

Nom et prénom :

.....

.....

Les réponses aux questions sont à donner exclusivement sur cette feuille : les réponses données sur les feuilles précédentes ne seront pas prises en compte.

QUESTION 1 : ☐ A ☐ B

QUESTION 2 : ☐ A ☐ B ☐ C

QUESTION 3 : ☐ A ☐ B

QUESTION 4 : ☐ A ☐ B

QUESTION 5 : ☐ A ☐ B

QUESTION 6 : ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F

QUESTION 7 : ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

QUESTION 8 : ☐ A ☐ B ☐ C

QUESTION 9 : ☐ A ☐ B ☐ C

QUESTION 10 : ☐ A ☐ B ☐ C ☐ D

QUESTION 11 : ☐ A ☐ B

QUESTION 12 : ☐ A ☐ B

QUESTION 13 : ☐ A ☐ B ☐ C

QUESTION 14 : ☐ A ☐ B ☐ C

QUESTION 15 : ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

QUESTION 16 : ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

QUESTION 17 : ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

QUESTION 18 : ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

QUESTION 19 : ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

QUESTION 20 : ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H

QUESTION 21 : ☐ A ☐ B ☐ C ☐ D

QUESTION 22 : ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H ☐ I ☐ J

QUESTION 23 : ☐ A ☐ B ☐ C ☐ D ☐ E ☐ F ☐ G ☐ H



+1/8/53+