

Matricule : Nom et Prénom : Groupe :

Interrogation 2021-2022**Exercice 1 : QCM (5 points)**

Cocher la (les) bonne (s) réponse (s) :

- 1- On peut généraliser le modèle linéaire de régression pour la classification en utilisant :
 - A. Une fonction d'activation non-linéaire.
 - B. Une transformation d'espace des variables d'entrée.
 - C. En utilisant le surapprentissage (overfitting).
 - D. Aucune de ces réponses
- 2- Les limitations de la méthode des moindres carrés pour la classification sont:
 - A. Les classes sont séparables et les données des classes suivent des lois Gaussiennes.
 - B. Elle n'est pas robuste aux données aberrantes.
 - C. C'est méthode très simple à programmer.
 - D. Elle donne toujours la solution la plus optimale.
 - E. Aucune de ces réponses
- 3- Si $(X^T X)^{-1}$ est inversible, on a recours aux solutions suivantes :
 - A. Transformer les données
 - B. Changer la classe d'hypothèses
 - C. Utiliser une approche classique telle que les systèmes experts
 - D. Appliquer une recherche exhaustive
 - E. Aucune de ces réponses
- 4- Afin de résoudre le problème de la régression, les paramètres w peuvent être calculés avec la formule:
 - A. $w = (X^T X)^{-1} X^T y$
 - B. $w = (X^T X)^{-1} y X^T y$
 - C. $w = (X^T X)^{-1} x X^T y$
 - D. $w = (X^T X)^{-1} Y^T x$
 - E. Aucune de ces réponses
- 5- Pour trouver l'hypothèse f la plus optimale pour la prédication,
 - A. On prend w qui minimise la fonction d'erreur.
 - B. On prend w qui maximise la fonction d'erreur.
 - C. On prend w qui minimise la fonction $f(x)$.
 - D. On prend w qui maximise la fonction $f(x)$.
 - E. Aucune de ces réponses

Exercice 2 : Arbre de décision (5 points)

Notre département MI désire utiliser l'apprentissage automatique afin d'améliorer le processus de sélection des d'étudiants pour l'accès à la première année master STIC pour la prochaine année 2022/2023. Pour atteindre cet objectif, le responsable de la spécialité STIC a décidé d'utiliser les résultats des d'étudiants inscrits en M1 STIC pour l'année en cours en se basant sur leurs résultats dans les matières : Algorithmique, Bases de données et Compilation obtenus en Licence. La table suivante résume les données rassemblées pour l'entraînement :

N°	BDD (L2)	Compil (L3)	Admis (2022/2023)
1	Excellent	Moyen	Oui
2	Faible	Excellent	Non
3	Moyen	Moyen	Non
4	Moyen	Excellent	Oui
5	Faible	Faible	Non
6	Excellent	Faible	Non

1) Construire un arbre de décision à partir de la table d'entraînement en utilisant le gain d'information. On donne le tableau suivant représentant les valeurs de la fonction :

$$H(x_1, x_2) = -\frac{x_1}{x_1 + x_2} \log_2 \left(\frac{x_1}{x_1 + x_2} \right) - \frac{x_2}{x_1 + x_2} \log_2 \left(\frac{x_2}{x_1 + x_2} \right)$$

$x_1 \backslash x_2$	1	2	3	4	5	6
6	0,592	0,811	0,918	0,971	0,994	1
5	0,650	0,863	0,954	0,991	1	
4	0,722	0,918	0,985	1		
3	0,811	0,971	1			
2	0,918	1				
1	1					

2) Calculer la décision obtenue par ce modèle pour les résultats de l'étudiant suivant:

(BDD : Faible, Compil : Faible)