**Examen Machine Learning Part 1 :**

**Théorie et pratiques**

Aucune question ne pourra être posée durant l'examen.

En cas de doute concernant le sujet, vous poursuivrez votre réponse en expliquant vos hypothèses.

Durée : 1h30 (15h15- > 16h45)

Épreuve du 05/12/2024

# Modalités du travail

* Durée : 1h30 ;
* Aucun document autorisé, calculatrice non autorisée ;
* Ecrire vos réponses sur la copie, dans les cases réservées à cet effet ;
* Mettre vos noms et prénoms sur chaque feuille ;
* ***Toute réponse donnée sans explications sera considérée comme incorrecte***
* ***Tout texte indéchiffrable sera considéré comme une absence de réponse***

**Part 1 : Questions générales (9 points)**

1. Quels sont les principaux facteurs ayant favorisé la popularité de l’apprentissage automatique ? **(0.25 pt)**

|  |
| --- |
|  |

1. Quelle est la différence entre l'apprentissage automatique (ML) et l'apprentissage profond (DL) ? **(0.25 pt)**

|  |
| --- |
|  |

1. Pourquoi l'apprentissage profond est-il souvent préféré à l'apprentissage automatique classique dans certaines applications ? **(0.5 pt)**

|  |
| --- |
|  |

1. Quel est le rôle du taux d'apprentissage dans la descente de gradient, et comment un taux d'apprentissage inapproprié peut-il impacter l'entraînement du modèle ? **(0.5 pt)**

|  |
| --- |
|  |

1. Définissez le surapprentissage (overfitting) et le sous-apprentissage (underfitting). Donnez un exemple de stratégie pour traiter chacun.. **(0.5 pt)**

|  |
| --- |
|  |

1. Quelles sont les différences principales entre la descente de gradient et la rétropropagation ? **(0.25 pt)**

|  |
| --- |
|  |

1. Expliquez la différence entre des fonctions convexes et non convexes. **(0.25 pt)**

|  |
| --- |
|  |

1. La régression linéaire est l’algorithme classique qu’on utilise pour les problèmes de régression **(2pts)**
   1. Quelle est la différence entre simple linear régression et multiple linear régression ?
   2. Donnez l’équation d’une régression linéaire simple et multiple ?
   3. Comment estimer les valeurs des paramètres d’une régression linéaire (donnez juste les noms des deux méthodes dans le cours)
   4. Donnez le morceau de code l’algorithme gradient descent (la descente du gradient)

|  |
| --- |
|  |

1. Qu’est-ce qu’un hyperparamètre en apprentissage automatique, et en quoi diffère-t-il d’un paramètre de modèle ? Donnez un exemple pour chacun? **(0.5 pt)**

|  |
| --- |
|  |

1. Expliquez l’objectif de la normalisation des caractéristiques (feature scaling) et citez deux méthodes utilisées pour cela? **(0.5 pt)**

|  |
| --- |
|  |

1. Décrivez la différence entre l’apprentissage supervisé et non supervisé. Donnez un exemple d'algorithme pour chaque cas **(0.25 pt)**

|  |
| --- |
|  |

1. Quelle est la fonction de l'activation dans les réseaux neuronaux ? Donnez un exemple d’une fonction d’activation couramment utilisée. ? **(0.25 pt)**

|  |
| --- |
|  |

1. Laquelle des métriques suivantes n’est PAS utilisée pour évaluer les modèles de classification ? **(0.25 pt)**

A. Précision (Accuracy)

B. Précision (Precision)

C. Erreur quadratique moyenne (Mean Squared Error)

D. F1-Score

|  |
| --- |
|  |

1. Quelle fonction de coût est généralement utilisée dans la régression logistique ? **(0.25 pt)**

A. Erreur quadratique moyenne (Mean Squared Error)

B. Perte d’entropie croisée (Cross-Entropy Loss)

C. Perte de Hinge (Hinge Loss)

D. Perte L1 (L1 Loss)

|  |
| --- |
|  |

1. Quelle affirmation concernant la validation croisée en k-fold (k-fold cross-validation) est VRAIE ?

A. Elle utilise un seul ensemble de test pour évaluer le modèle.

B. Elle moyenne les résultats de plusieurs séparations entre entraînement et test.

C. Elle conduit toujours au surapprentissage.

D. Elle ne fonctionne que pour les tâches de régression. **(0.25 pt)**

|  |
| --- |
|  |

1. Laquelle des méthodes suivantes est un exemple de méthode d’ensemble (ensemble method) ?

A. Forêt aléatoire (Random Forest)  
B. Decision Tree  
C. Linear Regression  
D. Descente de gradient (Gradient Descent) **(0.5 pt)**

|  |
| --- |
|  |

1. Qu’est-ce que le "dropout" en deep learning et pourquoi l’utilise-t-on ? **(0.25 pt)**

|  |
| --- |
|  |

1. Donnez les principaux composants d’un réseau de neurones convolutif (CNN) classique et expliquez l'importance de chacun. **(0.5 pt)**

|  |
| --- |
|  |

1. Quelle méthode peut être utilisée pour trouver les meilleurs hyperparamètres d’un modèle d’apprentissage automatique ? **(0.25 pt)**

A. Descente de gradient

B. Grid Search

C. Validation croisée (Cross-Validation)

D. Encodage one-hot

|  |
| --- |
|  |

1. Quel type d’apprentissage est utilisé lorsqu’il n’y a pas de variable cible dans les données ? **(0.25 pt)**  
   A. Apprentissage supervisé (Supervised Learning)

B. Apprentissage non supervisé (Unsupervised Learning)

C. Apprentissage par renforcement (Reinforcement Learning)

D. Apprentissage semi-supervisé (Semi-Supervised Learning)

|  |
| --- |
|  |

1. Quelle technique est utilisée pour prévenir le surapprentissage (overfitting) dans les réseaux neuronaux profonds ? **(0.25 pt)**

A. Normalisation des caractéristiques (Feature Scaling)

B. Dropout

C. Augmentation des données (Data Augmentation)

D. Toutes les réponses ci-dessus

|  |
| --- |
|  |

**Part 2 : KNN (3 points)**

Supposons que vous disposez d’un petit ensemble de données avec trois caractéristiques et une variable cible continue (Target). Voici les données :

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature 1** | **Feature 2** | **Feature 3** | **Target** |
| 1.2 | 0.8 | 2.5 | 12.5 |
| 2.0 | 1.5 | 2.9 | 15.0 |
| 1.8 | 1.0 | 3.1 | 14.2 |
| 2.5 | 2.2 | 3.5 | 16.8 |
| 3.0 | 2.7 | 3.8 | 18.5 |
| 2.8 | 3.0 | 3.3 | 17.3 |

* **Calculer la distance :** utiliser la distance euclidienne pour calculer la distance entre un nouveau point de données et tous les points de l'ensemble de données. (1 pts)

Nouveau point de données : [Feature 1 = 2.3, Feature 2 = 1.9, Feature 3 = 3.0]

* **Utiliser K = 3** (les 3 voisins les plus proches) pour prédire la valeur cible. (1 pt)
* **Proposez deux métriques d'évaluation** pour mesurer la performance d’un modèle KNN dans une tâche de régression. (1 pt)

|  |
| --- |
|  |

**Part 3 : Régression linéaire (4 points)**

Le tableau suivant décrit l’expérience de 5 étudiants avant l’examen d’un module et la note qu’ils ont obtenu en conséquence du nombre d’heures qu’ils ont passé à étudier et du nombre d’heures qu’ils ont dormi la veille de l’examen. La première colonne contient E1 jusqu’à E5 : qui est l’identifiant de chaque étudiant. La deuxième colonne définit le nombre d’heures total passé par chaque étudiant à étudier le module, la troisième colonne définit le nombre d’heures que chaque étudiant a dormi la veille de l’examen et la dernière colonne définit la note obtenue pour ce module.

|  |  |  |  |
| --- | --- | --- | --- |
| Etudiant | Nombre d’heures d’études | Nombre d’heures de sommeil la veille de l’examen | Note |
| E1 | 1 | 8 | 3 |
| E2 | 20 | 8 | 18 |
| E3 | 5 | 5 | 7 |
| E4 | 15 | 3 | 14 |
| E5 | 25 | 8 | 19 |

Nous voulons définir la fonction qui exprime la note en fonction du nombre d’heures d’études et du nombre d’heures de sommeil la veille de l’examen en utilisant la régression linéaire :  
Note = ***f (***Nombres d’heures d’études, Nombre d’heures de sommeil la veille de l’examen***)***

1. À quelles colonnes correspondent x1, x2 et y dans le tableau précédent ? (0.5 p)
2. Exprimer hθ(x) (le modèle linéaire) en fonction des θi, des xi, pour le tableau précèdent. (0.5 p)
3. Si on vous donne deux propositions de θ = **[θ0 = 0, θ1 = 1 θ2 = 0.5]** et θ = **[θ0 = 0, θ1 = 1 θ2 = 1]** laquelle des deux permetde prédire le mieux la note pour l’étudiant E3 ? Justifier la réponse. (0.5 p)
4. Pour le choix de θ = **[0, 1, 0.5]**, exprimer l’erreur **J(θ)** pour les étudiants sur le tableau en exprimant la formule et les calculs d’une manière claire. (1.5 p)
5. Pour le choix de α = 0.1 et de θ = **[0, 1, 0.5]**, et la formule de mise à jour :

Calculer la nouvelle valeur de θ après une mise à jour, détaillez le résultat. (1 p)

|  |
| --- |
|  |

**Part 4 : Régression logistique (4 points)**

Pour adapter le tableau précédent au problème de **régression logistique** nous proposons un tableau où la note est remplacée par la validation : V. Le but est de trouver un modèle de régression logistique qui doit prédire si un étudiant a validé un modèle en utilisant le nombre d’heure qu’il a passé à étudier ce module et le nombre d’heures qu’il a dormi la veille de l’examen. Le tableau est le même en changeant uniquement la dernière colonne par l’utilisation de la règle suivante. Si Note>=10 alors V=1 sinon V=0.

|  |  |  |  |
| --- | --- | --- | --- |
| Etudiant | Nombre d’heures d’études | Nombre d’heures de sommeil la veille de l’examen | Note |
| E1 | 1 | 8 | 0 |
| E2 | 20 | 8 | 1 |
| E3 | 5 | 5 | 0 |
| E4 | 15 | 3 | 1 |
| E5 | 25 | 8 | 1 |

1. À quelles colonnes correspondent x1, x2 et y dans le tableau précédent. (0.5 p)
2. Un nouvel étudiant a étudié pendant 10 heures et a dormi pendant 6 heures. Selon le modèle, a-t-il validé le module quand θ = **[-10, 6/10, 4/6]**, la règle est la suivante : si σ(...) ⩾ 0.5 alors l’étudiant a validé sinon il n’a pas validé, écrire la formule et les calculs correspondants. (1.5 p)
3. Vérifier si pour la même valeur θ si le modèle prédit bien si un étudiant a validé ou pas dans le cas de l’étudiant à la première ligne et celui à la quatrième ligne. Justifier en appliquant les règles et les calculs correspondants. (2 p)

|  |
| --- |
|  |