

Examen Machine Learning Part 2 :

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 : 2h

Épreuve du 08/03/2024

Modalités du travail

- ⊘ Durée : 2h ;
- ⊘ 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 (12 points)

1. Quels sont les principaux moteurs (pourquoi il est devenu si populaire) de l'apprentissage automatique ? (0.5p)

2. La différence entre Artificial Intelligence (AI), Machine Learning (ML), et Deep Learning (DL) ? (0.5p)

3. La différence entre la classification et la régression ? Un exemple ? (0.5p)

4. La régression linéaire est l'algorithme classique qu'on utilise pour les problèmes de régression (2p)

- Quelle est la différence entre simple linear régression et multiple linear régression ?
- Donnez l'équation d'une régression linéaire simple et multiple ?
- Comment estimer les valeurs des paramètres d'une régression linéaire (donnez juste les noms des deux méthodes dans le cours)
- Donnez le morceau de code l'algorithme gradient descent (la descente du gradient)

5. On utilise la régression logistique pour quel type de problème (classification, régression ou clustering) ? Pourquoi ? Donnez la formule de la fonction Sigmoid. (1.5p)

6. Donnez un exemple de fonction de cout (loss function) pour la classification et la régression

7. La différence entre un arbre de décision et foret aléatoire.

**8. Pour un arbre de décision pour la classification, quels sont les critères de répartition.
Expliquez d'une manière générale le process de développer un arbre de décision.**

9. Les étapes pour développer un KNN pour la classification et la régression.

**10. Quelles sont les différentes métriques d'évaluation pour les problèmes de classification ?
(Donnez quatre métriques avec les formules)**

**11. Quelles sont les différentes métriques d'évaluation pour les problèmes de régression ?
(Donnez trois métriques avec les formules)**

12. Donnez une petite définition de True Positives, True Negatives, False Positives, et False Negatives

13. Qu'est-ce qu'une Confusion Matrix ?

14. Comment on utilise le ROC-AUC graphe ? Expliquez-les étapes de construire ce graphe ?

15. Quelles sont les étapes pour développer un K-Means ?

16. Expliquez le concept de surapprentissage (overfitting) dans le contexte de l'apprentissage automatique. Comment peut-on le prévenir ?

17. C'est quoi le Bias-Variance tradeoff, et quelle relation avec le surapprentissage (overfitting) et le sousapprentissage (underfitting)

18. Décrivez les étapes pour entrainer un réseau de neurones artificiels (MLP/ANN)

19. Dropout en deep learning? Et pourquoi on l'utilise ?

20. Donner les composants (l'architecture) d'un CNN classique, et l'importance de chaque composant ?

Part 2 : 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 x_1 , x_2 et y dans le tableau précédent ? (0.5 p)
2. Exprimer $h_\theta(x)$ (le modèle linéaire) en fonction des θ_i , des x_i , pour le tableau précédent. (0.5 p)
3. Si on vous donne deux propositions de $\theta = [\theta_0 = 0, \theta_1 = 1 \ \theta_2 = 0.5]$ et $\theta = [\theta_0 = 0, \theta_1 = 1 \ \theta_2 = 1]$ laquelle des deux permet de prédire le mieux la note pour l'étudiant E3 ? Justifier la réponse. (0.5 p)
4. Pour le choix de $\theta = [0, 1, 0.5]$, exprimer l'erreur $J(\theta)$ 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 $\alpha = 0.1$ et de $\theta = [0, 1, 0.5]$, et la formule de mise à jour :

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

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

Part 3 : 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 module 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 $Note \geq 10$ 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 x_1 , x_2 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 $\theta = [-10, 6/10, 4/6]$, la règle est la suivante : si $\sigma(\dots) \geq 0.5$ 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)