

Techniques d'apprentissage
IFT 603 - 712Examen Final : XYZ %
XYZ mars 20XY, 10 h 30 à 12 h 20

Professeur : Pierre-Marc Jodoin

Nom :

Prénom :

Signature :

Matricule :

Consignes

1. Cet examen comporte **huit (8) questions** au total.
2. Vous avez droit à des feuilles manuscrites **non photocopiées** pour toute documentation.
3. Les calculatrices sont permises.
4. Répondez sur le questionnaire.

Question 1 (X points)

Dites comment un modèle de classification binaire $y_w(\vec{x})$ peut être utilisé pour classer des données en K classes où $K > 2$.

Question 2 (X points)

Prouvez mathématiquement qu'un réseau de neurones multicouche ayant pour fonction d'activation la $\tanh(\cdot)$ peut vraisemblablement souffrir d'un problème de disparition de gradient. (Suggestion : vous pouvez vous inspirer de la fonction (5.59) de Bishop).

Question 3 (X points)

Expliquez pourquoi, lorsqu'on entraîne une méthode d'apprentissage sur une base de données fortement déséquilibrée on obtient souvent une très haute précision pour une classe mais un rappel très faible pour l'autre classe avec, au total, une très bonne "accuracy".

Question 4 (X points)

Prouvez que si $k_1(\vec{x}, \vec{x}')$ et $k_2(\vec{x}, \vec{x}')$ sont des noyaux valides, alors $k(\vec{x}, \vec{x}') = k_1(\vec{x}, \vec{x}')k_2(\vec{x}, \vec{x}')$ est également un noyau valide.

Question 5 (X points)

Prouvez que $t_i \frac{y(\vec{x}_i)}{\|\vec{W}\|}$ est la plus petite distance signée entre le point \vec{x}_i et l'hyperplan défini par le vecteur de paramètres \vec{W} .

Question 6 (X points)

Expliquez pourquoi les machines à vecteur de support imposent qu'une donnée bien classée doit être à une distance supérieure à 1 de la surface de séparation.

Question 7 (X points)

Dites comment à l'aide de l'algorithme des K-moyennes (*k-means*) on peut compresser l'espace mémoire requis pour stocker les paramètres d'un réseau de neurones sans trop en réduire les performances.

Question 8 (X points)

Soit un réseau de neurones à K classes (ayant donc K neurones en sortie) et la couche $sm(\vec{W}^T \vec{x})$ (*softmax*) placée à la fin du réseau de neurones. Expliquez pourquoi, lorsque le score pour un des neurones de sortie est très élevé, la sortie du softmax possède parfois de "NaN". Donnez une façon de se prémunir contre cet effet pervers.

Question 9 (X points)

Nous avons vu que pour calculer le gradient d'un réseau de neurones avec entropie croisée, il nous faut calculer une matrice jacobienne. Selon vous, si on remplace l'entropie-croisée par une distance L2, sera-t-il toujours nécessaire de calculer cette matrice ?

Question 10 (X points)

Votre employeur vous demande de trouver les "clusters" dans un ensemble de données $D = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N\}$. Donnez un cas de figure pour lequel

1. l'algorithme *K-Means* est le plus approprié
2. l'algorithme *E-M* (*expectation-maximization*) est le plus approprié
3. l'algorithme *Mean-Shift* est préférable à *E-M* et *K-Means*.

Question 11 (X points)

Expliquez en quoi consiste la méthode de classification dites des forêts aléatoires (*random forests*). On dit que cette méthode implique à la fois du bagging et du boosting. Expliquez pourquoi.

Question 12 (X points)

Expliquez pourquoi un modèle peut avoir à la fois une très bonne précision et une très mauvaise justesse (*accuracy*).

Question 13 (X points)

Votre employeur se questionne sur les performances d'une méthode de classification que vous lui proposez. Afin de le convaincre, partant d'un ensemble d'entraînement $D = \{(\vec{x}_1, t_1), (\vec{x}_2, t_2), \dots, (\vec{x}_N, t_N)\}$ où N est élevé, donnez deux algorithmes : un pour calculer le biais et un pour calculer la variance de votre modèle.

Question 14 (X points)

Expliquez en quoi consiste un modèle dit "génératif" et le lien qu'il entretient avec le concept de vraisemblance des données.

Question 15 (X points)

Un médecin souhaite mettre sur pied une base de données provenant de plusieurs hôpitaux afin d'entraîner des algorithmes intelligents pour classifier cinq (5) types de tumeurs cérébrales. Il vous dit qu'une bonne prédiction doit s'appuyer sur 50 critères anatomiques, tous des réels. Certaines de ces tumeurs sont très fréquentes alors que d'autres le sont beaucoup moins. De plus, le médecin aimerait que le système puisse identifier les patients n'ayant aucune tumeur.

(a) Combien de classes au total la base de données contiendra ?

(b) Donnez trois (3) recommandations importantes afin que la base de données permette d'entraîner de bons algorithmes intelligents.

Question 16 (X points)

Lorsqu'un réseau de neurones ne performe pas mieux qu'une méthode aléatoire, on dit de lui qu'il a une entropie élevée. Expliquez mathématiquement pourquoi. (Note : il est question ici d'entropie et non d'entropie croisée).