

PARCOURS / ETAPE : Master Informatique

Code UE : 4TIN703EX

Epreuve : Analyse, Classification, Indexation des données.

Date : 28/12/2016 Heure : 9h Durée : 1h30

Documents : non autorisés

Epreuve de Mr : Domenger

Question de cours (8 points)

- Donner une définition de l'apprentissage supervisé.
- Expliquer le principe de l'algorithme des k-plus proches voisins.
- Expliquer les différences entre l'approche en composante principale (ACP) et en composante discriminante (ACI). Vous pourrez illustrer votre discours avec une figure.

« Classifieurs Linéaires » (13 points)

Une classe d'apprentissage C1 est représentée par les éléments suivants :

$$C1 = \{ (1,2) ; (2,3) \}$$

Une classe d'apprentissage C2 est représentée par les éléments suivants :

$$C2 = \{ (1,1) ; (3,2) \}$$

Perceptron (7 points)

- Donner une définition d'un « classifieur linéaire », quelles sont ses propriétés.
- Expliquer et formaliser la méthode **du perceptron** sur les données issues des classes C1 et C2.
 - Formalisation du problème,
 - Expliquer l'augmentation en dimension et la modification des données en fonction de l'appartenance à la classe
 - Exprimer le critère à minimiser et l'expliquer.
 - Donner le principe de l'algorithme
- En initialisation, le vecteur W à [+1 ; -2 ; +2], faire tourner l'algorithme jusqu'à convergence en explicitant l'ensemble des « mals classés » ainsi que l'évolution du vecteur W représentant la droite de séparation.
- Si on considère le vecteur W calculé précédemment, à quelle classe appartiendrait le point C = (4,4).
- Comment se comporte l'algorithme dans le cas où le problème n'est pas linéairement séparable. Quelle solution pourriez-vous mettre en œuvre pour que l'algorithme finisse par converger.

Moindre Carré (6 points).

Une autre manière de définir un « classifieur linéaire » est d'utiliser les moindres carrés, la fonction à minimiser s'exprime alors par

$$J(W) = \|YW - b\|^2$$

où W représente la droite de séparation, Y est la matrice représentant l'échantillon augmenté en dimension dont les données de la seconde classe sont modifiées et b un vecteur.

On rappelle la dérivée $\frac{\partial J(W)}{\partial W} = 2Y^t(YW - b) = 2(Y^tYW - Y^tb)$

- Justifier les propriétés que doit vérifier le vecteur b pour obtenir un séparateur linéaire.
- En reprenant les classes C1 et C2 définies précédemment, donner l'expression de Y, proposer un choix pour le vecteur b.

- Expliquer les deux méthodes permettant de calculer le vecteur W .

La fonction à minimiser peut également s'exprimer de la manière suivante :

$$J(W, b) = \| YW - b \|^2$$

On rappelle les dérivées partielles de la fonction $J(W, b)$

$$\frac{\partial J(W, b)}{\partial W} = 2Y^t(YW - b) = 2(Y^tYW - Y^tb) \quad \text{et} \quad \frac{\partial J(W, b)}{\partial b} = -2(YW - b)$$

- Expliquez la différence avec la formulation précédente. Quelle propriété doit être vérifiée par les éléments du vecteur b .
- Proposer un algorithme permettant le calcul du couple (W, b) .

FIN