

PARCOURS / ETAPE : Master Informatique

Code UE : 4TIN703EX

Epreuve : Analyse, Classification, Indexation des données.

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

Documents : non autorisés

Epreuve de Mr : Domenger

Approche Bayésienne (7 points)

On note $P(x \in C_i)$ la probabilité qu'un échantillon x appartienne à la classe C_i . On note $P(x = f | C_i)$ la probabilité que l'échantillon x soit égal à f sachant que x appartient à la classe C_i , on note $P(x \in C_i | x = f)$ la probabilité que l'échantillon x appartienne à la classe sachant que x est égal à f .

- Donner les principales lignes de l'approche Bayésienne, quelles sont les connaissances ou hypothèses nécessaires à sa mise en œuvre ?
- Détailler la règle majoritaire (maximum a priori), la règle du maximum de vraisemblance, la règle de maximum a posteriori ?
- Comment peut-on tenir du compte du coût d'une mauvaise affectation ?
- Dans le cas à 2 classes, comment s'exprime l'erreur de classification ?
- Dans le cas à n classes comment s'exprime la fonction de classification ?

On suppose que $P(x = f | C_1) = \mathcal{N}(\mu_1, \Sigma)$ et que $P(x = f | C_2) = \mathcal{N}(\mu_2, \Sigma)$ où $\mathcal{N}(\mu, \Sigma)$ représente une loi normale multivariée de moyenne μ et de matrice de covariance Σ .

- On suppose que $\Sigma = \sigma^2 I$ avec I la matrice identité et que $P(x \in C_1) = P(x \in C_2)$. Caractériser la frontière de séparation entre les classes C_1 et C_2 (type, point remarquable situé sur la frontière, direction...).
- On suppose maintenant $P(x \in C_1) = P(x \in C_2)$ et que les matrices de covariance sont identiques pour les distributions des 2 classes. Caractériser la frontière de séparation entre les classes C_1 et C_2 (type, points remarquables situés sur la frontière, direction...) qu'elle est la distance qui remplace la distance euclidienne dans ce cas. Donner son expression.

Perceptron (7 points)

Une classe d'apprentissage C_1 est représentée par les éléments suivants :

$$\{ C_{11} = (1,5) ; C_{12} = (4,3) ; C_{13} = (7,5) ; \}$$

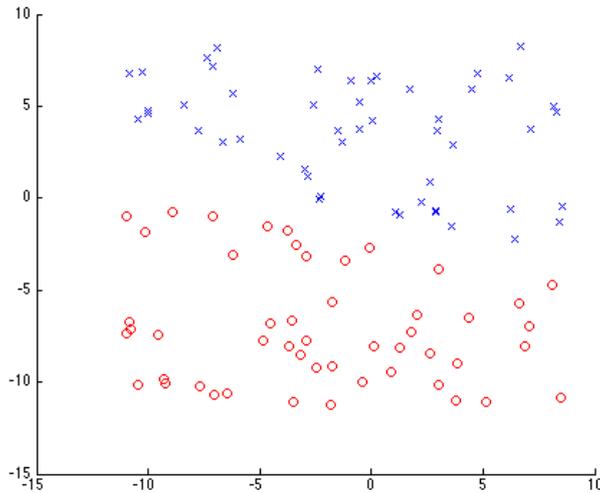
Une classe d'apprentissage C_2 est représentée par les éléments suivants :

$$\{ C_{21} = (2,1) ; C_{22} = (5,3) ; C_{23} = (6,3) \}$$

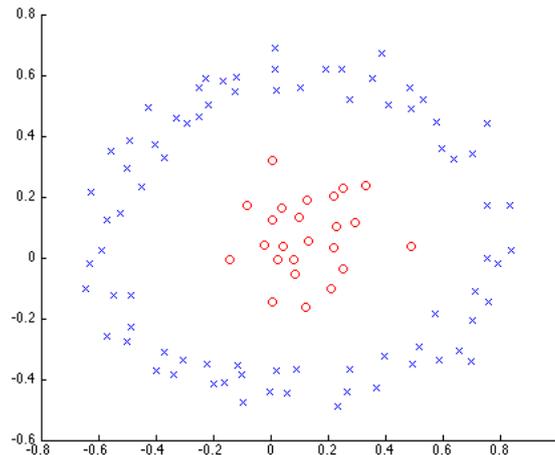
- Expliquer et formaliser la méthode du perceptron sur ces données (équation, propriété de la solution dans le cas d'un problème linéairement séparable, augmentation en dimension des données et modification en fonction de l'appartenance à la classe).
- Faire tourner l'algorithme jusqu'à convergence.
- A quelle classe appartiendra le point $C = (3,2)$
- On considère que le point $C_{14} = (3,5)$ appartient maintenant à la classe C_1 . Est-ce que ce point est bien classé par la solution obtenue précédemment. Le problème est-il encore linéairement séparable après l'ajout du point C_{14} ? Dans ce cas, quel est le résultat produit par l'algorithme du Perceptron. Proposez une adaptation de l'algorithme afin d'obtenir une solution.

Cours (6 points)

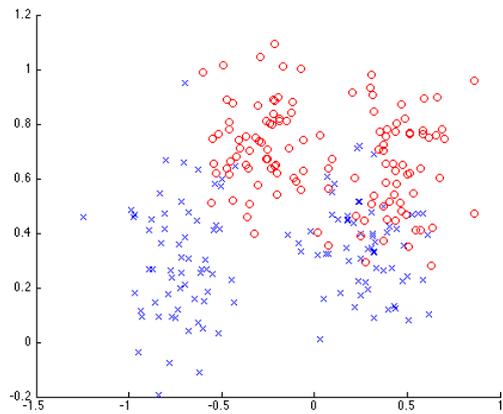
1. Dans le cadre d'un séparateur linéaire, expliquer la différence entre ces deux fonctions à minimiser $J(W) = \|YW - b\|^2$ et $J(W,b) = \|YW - b\|^2$
2. Donner pour chacune des deux fonctions précédentes un schéma de calcul permettant de résoudre ce problème de minimisation.
3. Pour le SVM, donnez la fonction J à optimiser. Expliquer comment le SVM permet de traiter des problèmes non linéairement séparable dans l'espace initial des paramètres. Dans ce cas, quelle est la propriété que doit vérifier le noyau d'un SVM ?.
4. Pour chacune des images suivantes, expliquer et justifier les techniques de classification qui sont utilisables sur les échantillons présentés et aussi celles qui ne le sont pas.



ECHANTILLON 1



ECHANTILLON 2



ECHANTILLON 3

FIN