Contrôle de Connaissances - Brique RDF – Module ADES

autorisés : polycopié Ades uniquement et calculette
durée 1h30

mars 2005

Exercice 1

Soit un problème de classification à 2 classes ω1 et ω2 dont les distributions mono-dimensionnelles p(x|ωi), i=1,2 sont supposées connues et égales à :

$p(x|\omega 1) \sim \mathcal{N}(m_1,\sigma)$

$p(x|\omega 2) \sim \mathcal{N}(m_2,\sigma)$

on supposera sans perte de généralité que m1<m2

1. Dans cette question, on applique la règle de Bayes avec coûts. On introduit la matrice de coûts :

$\lambda(\omega_i|\ \omega_j)$ i, j =1,2 $i \neq j$ coût de classer un objet de la classe ω_j dans la classe ω_i

$\lambda(\omega_i|\ \omega_i)=0$ i=1,2

1.a) exprimer la règle de Bayes (coût minimum) et calculer le point frontière x_b* entre les 2 classes en fonction de m_1, m_2, σ, des coûts λ, et des probabilités a priori notées P_1 et P_2.

1.b) application : m_1=1, m_2=3, σ =1, P1=1/3, $\lambda(\omega_2|\ \omega_1)$=4, $\lambda(\omega_1|\ \omega_2)$=1

2. Dans cette question on introduit un nouveau critère de classification (critère de Neyman-Pearson) : on impose que l'erreur de type 1 soit faible et égale à ε_1.

2.a) Qu'est-ce qu'une erreur de type 1 ?

2.b) Exprimer la relation que vérifie le point frontière x_n en fonction de m_1, σ, P_1, ε_1 et à l'aide de la fonction *erfc* définie par :

$$erfc(x)=\frac{2}{\sqrt{\pi}}\int_x^{+\infty} e^{-t^2}\,dt$$

Exercice 2

Soient les trois prototypes suivants d'une base d'apprentissage :

x	y	classe
1	1	+
-1	-2	–
-1	0	–

1) rappeler la règle séquentielle du perceptron
2) appliquer cette règle à l'ensemble d'apprentissage pour trouver un hyperplan séparateur à partir de l'hyperplan initial y=0

Classification automatique

On considère une image consituée de pixels ne prenant que 3 valeurs : 0, 255 et une troisième valeur X comprise entre 1 et 254. On appelle A la catégorie des pixels de valeur 0, B la catégorie des pixels de valeur X et C la catégorie des pixels de valeur 255. Il y a N pixels de catégorie A, N pixels de catégorie C et M pixels de catégorie B. On supposera $M \geq N$.

On recherche une classification en deux classes : K1 et K2, selon l'algorithme des k-moyennes. Eventuellement, dans les réponses aux questions, on détaillera les différentes étapes de cet algorithme menant à sa convergence.

Question 1 : En initialisant la classe K1 par un élément de la catégorie A et la classe K2 par un élément de la catégorie B, montrer que l'algorithme des K-moyennes converge vers une solution que l'on établira et qui dépend éventuellement de X, M et N.

Question 2 : En initialisant la classe K1 par un élément de la catégorie B et la classe K2 par un élément de la catégorie C, montrer que l'algorithme des K-moyennes converge vers une solution que l'on établira et qui dépend éventuellement de X, M et N.

Question 3 : Comparer les résultats précédents. Comment pouvez vous choisir entre deux solutions différentes ?

Question 4 : Calculer la dispersion intra-classe et la dispersion inter-classe dans les deux cas précédents. Définir alors la solution minimisant globalement l'erreur quadratique moyenne. *Remarque* : dans cette question, on recherchera la méthode la plus élégante et simple en prenant en compte les acquis du cours.