

SY09 Printemps 2009

TP 3

Théorie de la décision

Exercice 1. Classifieur euclidien

On veut étudier les performances du classifieur euclidien sur des données de deux classes de \mathbb{R}^2 avec des formes géométriques différentes.

1. Pour cela, générer un jeu de données D1App de taille $n_{1App} = 300$ correspondant à la classe ω_1 et un autre D2App de taille $n_{2App} = 300$ correspondant à la classe ω_2 suivant deux lois normales $\mathcal{N}(\mu_1, \Sigma_1 = \mathbf{I})$ et $\mathcal{N}(\mu_2, \Sigma_2 = a^2 \mathbf{I})$ dont vous changerez les paramètres.
2. Pour chaque cas, générer deux ensembles de test D1Test et D2Test de même taille $n_{Test} = 300$, et estimer la probabilité d'erreur du classifieur euclidien. Interpréter intuitivement les résultats.
Vous devrez écrire votre programme dans un fichier MainEuclidien.r qui utilise deux fonctions :
 - La fonction `erreurEstimee = fonction(D1Test,D2Test,Regle,Theta)` qui retourne la probabilité d'erreur estimée. Utiliser dans cette fonction les commandes `apply` de la manière suivante : `classement1=apply(D1Test,1,Regle,Theta=Theta)` et `classement2=apply(D2Test,1,Regle,Theta=Theta)` où `Theta` est une liste composée de μ_1 et μ_2 .
 - La fonction `regleEuclidienne = fonction(x,Theta)` qui retourne 1 ou 2 suivant l'action à prendre pour l'observation `x` à classer.
3. Refaire 10 fois la question 2 en sauvegardant les résultats et calculer la moyenne et la variance de votre résultat.

Exercice 2. Règles de Neyman-Pearson et Bayes

On considère un problème de détection de cibles dans lequel la classe ω_1 correspond aux missiles, et la classe ω_2 correspond aux avions. Chaque cible est décrite par deux variables X_1 et X_2 issues de deux capteurs différents. Chaque variable suit, dans chaque classe, une loi normale avec les paramètres suivants :

$$f_1(x_1) \sim \mathcal{N}(-1, 1), \quad f_2(x_1) \sim \mathcal{N}(1, 1),$$

$$f_1(x_2) = f_2(x_2) \sim \mathcal{N}(0, 1).$$

On suppose l'indépendance conditionnelle de X_1 et X_2 . Les densités conditionnelles du vecteur $\mathbf{X} = (X_1, X_2)'$ sont donc $f_1(\mathbf{x}) = f_1(x_1)f_1(x_2)$ dans la classe ω_1 et $f_2(\mathbf{x}) = f_2(x_1)f_2(x_2)$ dans la classe ω_2 .

1. Générer suivant la loi f_1 , des vecteurs $Vect1X_1$ et $Vect1X_2$ de taille $n = 300$. Générer, de même, suivant la loi f_2 , des vecteurs $Vect2X_1$ et $Vect2X_2$.
2. A partir des vecteurs générés, et pour différentes valeurs de n , calculer des estimations pour les différents paramètres de f_1 et f_2 . Interpréter les résultats.
3. Donner l'expression de $f_1(\mathbf{x})$ et $f_2(\mathbf{x})$. Montrer que les courbes iso-densité sont des cercles dont on précisera les rayons.
4. Soit $\mathcal{A} = \{a_1, a_2\}$ les actions d'affectation aux classes ω_1 et ω_2 . Pour une règle de décision δ , on définit les probabilités d'erreur $\alpha = P(\delta(\mathbf{X}) = a_2 | \omega_1)$ et $\beta = P(\delta(\mathbf{X}) = a_1 | \omega_2)$. On rappelle que la règle de Neyman-Pearson minimise β sous la contrainte $\alpha \leq \alpha^*$ pour une valeur α^* fixée appelée niveau de signification.
 - (a) Montrer que la règle de Neyman-Pearson pour ce problème s'exprime en fonction d'une seule variable. Interpréter ce résultat.
 - (b) Donner l'expression de cette règle en fonction de α^* .
 - (c) Construire une fonction, qui en fonction de α^* , dessine la frontière de décision correspondante dans le plan (X_1, X_2) . Application : $\alpha^* = 0.05$ et $\alpha^* = 0.1$.
 - (d) A partir des données générées précédemment, donner une estimation de α et β .
 - (e) Donner l'expression de la courbe COR $1 - \beta = g(\alpha^*)$, et tracer cette courbe avec R.
5. Soient π_1 et π_2 les probabilités a priori des deux classes, et c_{lk} le coût associé au choix de l'action a_ℓ lorsque la vraie classe est ω_k . On suppose $c_{11} = c_{22} = 0$. L'ensemble \mathcal{A} des actions est le même que dans la question précédente.

- (a) Donner l'expression de la règle de Bayes δ^* pour ce problème.
- (b) Tracer avec R les frontières de décision correspondantes dans le plan (X_1, X_2) dans les cas suivants :
- i. $c_{12} = c_{21} = 1, \pi_1 = \pi_2$;
 - ii. $c_{12} = 10c_{21} = 1, \pi_1 = \pi_2$;
 - iii. $c_{12} = c_{21} = 1, \pi_2 = 10\pi_1$.
- (c) Pour ces différents cas, et à partir des données générées précédemment, donner une estimation de $\alpha = P(\delta^*(\mathbf{X}) = a_2 | \omega_1)$ et $\beta = P(\delta^*(\mathbf{X}) = a_1 | \omega_2)$. Commenter ces résultats.