Partiel de Classification et Reconnaissance des Formes

Lundi 9 décembre 2013, 8h00-9h45.

Partiel avec documents autorisés
Barème indicatif : Ex. 1 (11pts), Ex. 2 (11pts)

Exercice 1 : Questions portant sur le cours

1) **Analyse linéaire discriminante** (3pts) : On considère un problème de classification à deux classes ω_1 et ω_2 avec la base d'apprentissage

$$\omega_1 \quad : \quad \mathbf{x}_1 = (1,1)^T, \mathbf{x}_2 = (1+\varepsilon, 1+\varepsilon)^T, \mathbf{x}_3 = (1+\varepsilon, 1-\varepsilon)^T, \mathbf{x}_4 = (1-\varepsilon, 1+\varepsilon)^T, \mathbf{x}_5 = (1-\varepsilon, 1-\varepsilon)^T$$

$$\omega_2 \quad : \quad \begin{cases} \mathbf{x}_6 = (-1,-1)^T, \mathbf{x}_7 = (-1+\varepsilon, -1+\varepsilon)^T, \mathbf{x}_8 = (-1+\varepsilon, -1-\varepsilon)^T, \mathbf{x}_9 = (-1-\varepsilon, -1+\varepsilon)^T \\ \mathbf{x}_{10} = (-1-\varepsilon, -1-\varepsilon)^T \end{cases}$$

où $\varepsilon > 0$ est un réel positif de petite valeur. On note $\mathbf{S}_1$ et $\mathbf{S}_2$ les matrices de covariance intra-classe des données et $\mathbf{B}$ la matrice de covariance inter-classes..

- Déterminer les matrices $\mathbf{S}_1$,$\mathbf{S}_2$ et $\mathbf{B}$.
- Que peut-on dire des valeurs propres de la matrice $\mathbf{S}^{-1}\mathbf{B}$ avec $\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2$?
- Sans faire de calcul, indiquer l'axe le plus discriminant issu de la maximisation du critère de Fisher.

2) **Classifieur Bayésien** (2pts) : On considère un problème de classification à deux classes équiprobables définies comme suit

$$f(x\,|\omega_1) = \mathcal{N}(0, \sigma^2) \text{ et } f(x\,|\omega_2) = \mathcal{N}(1, \sigma^2).$$

Déterminer et représenter graphiquement $P(\omega_1\,|x)$ et $P(\omega_2\,|x)$. En déduire la règle de décision du classifieur Bayésien pour ce problème.
3) **Classifieur Bayésien** (1pt) : On considère un problème de classification à trois classes équiprobables définies comme suit

$$f(\mathbf{x}\,|\omega_1) = \mathcal{N}(\mathbf{m}_1, \sigma^2\mathbf{I}), f(\mathbf{x}\,|\omega_2) = \mathcal{N}(\mathbf{m}_2, \sigma^2\mathbf{I}_2) \text{ et } f(\mathbf{x}\,|\omega_3) = \mathcal{N}(\mathbf{m}_3, \sigma^2\mathbf{I}_2)$$

où $\mathbf{m}_1 = (1,1)^T, \mathbf{m}_2 = (1,-1)^T$ et $\mathbf{m}_3 = (-1,1)^T$ et où $\mathbf{I}_2$ est la matrice identité de taille 2×2. Sans faire de calcul, représenter les régions du plan associées aux décisions du classifieur Bayésien $d^*(\mathbf{x}) = \omega_1$ (on affecte le vecteur $\mathbf{x}$ à la classe ω_1), $d^*(\mathbf{x}) = \omega_2$ (on affecte le vecteur $\mathbf{x}$ à la classe ω_2) et $d^*(\mathbf{x}) = \omega_3$ (on affecte le vecteur $\mathbf{x}$ à la classe ω_3).
4) **Algorithme K-means** (2pts) : On considère un problème de classification avec la base d'apprentissage

$$\mathbf{x}_1 = (1,1)^T, \mathbf{x}_2 = (1,0)^T, \mathbf{x}_3 = (1,-1)^T, \mathbf{x}_4 = (-1,-1)^T, \mathbf{x}_5 = (-1,0)^T, \mathbf{x}_6 = (-1,1)^T.$$

On suppose que ce problème admet deux classes ω_1 et ω_2 et que ces classes admettent comme représentants $\mathbf{p}_1 = (-0.5, 0)^T$ et $\mathbf{p}_2 = (0.5, 0)^T$. Décrire les itérations de l'algorithme K-means initialisé avec les représentants $\mathbf{p}_1$ et $\mathbf{p}_2$ et donner la répartition finale des six points de la base d'apprentissage dans les deux classes ω_1 et ω_2. Faire de même avec $\mathbf{p}_1 = (0, -0.5)^T$ et $\mathbf{p}_2 = (0, 1)^T$. Que peut-on en conclure ?
5) **Perceptron** (3pts) : on considère un problème de classification à deux classes ω_1 et ω_2 avec la base d'apprentissage constituée des $n = 7$ points suivants

$$\begin{aligned} \omega_1 &: \; x_1 = 1, x_2 = 2, x_3 = 4, x_4 = 5 \\ \omega_2 &: \; x_5 = 8, x_6 = 9, x_7 = 11 \end{aligned}$$

- Quelle est la règle de décision issue de la méthode des machines à vecteurs supports ?
- On construit un perceptron à une couche possédant un seul neurone avec une fonction d'activation linéaire. Si l'entrée de ce neurone est x, alors la sortie de ce neurone est $y = ax + b$. De plus pour toute entrée x_i de la classe ω_1, la sortie désirée est $d_i = -1$ tandis que pour toute entrée x_i de la classe ω_2, la sortie désirée est $d_i = 1$. Quelle est la fonction de coût $E(a, b)$ à considérer pour le filtre de Wiener ? Montrer que lorsqu'on remplace les espérances mathématiques par des moyennes empiriques, cette fonction de coût est minimale pour

$$\begin{aligned} b &= \frac{1}{n}\sum_{i=1}^{n} d_i - \frac{a}{n}\sum_{i=1}^{n} x_i \\ a &= \frac{\frac{1}{n}\sum_{i=1}^{n} d_i x_i - \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} \end{aligned}$$

Après application numérique, on trouve $a \simeq 0.153$ et $b \simeq -1.014$. Quelle est la règle de décision associée ?

Exercice 2 : Questions portant sur l'article

1) (2pts) Dans l'introduction de l'article les auteurs parlent de “standard linear and quadratic discriminant analysis”. Expliquer le principe de ces règles de classification (vues en cours).
2) (1pt) Rappeler l'expression du classifieur Bayésien dans le cas de densités normales avec une fonction de coût $c_{ij} = 1 - \delta_{ij}$. N'y-t-il pas une erreur dans l'équation (3) de l'article.
3) (2pts) Notons $\{\mathbf{x}_{ik}, i = 1, ..., N_k\}$ les vecteurs de la base d'apprentissage associés à la classe ω_k. Démontrer que l'estimateur du maximum de vraisemblance du vecteur moyenne $\boldsymbol{\mu}_k$ dans le cas Gaussien est

$$\widehat{\boldsymbol{\mu}}_k = \frac{1}{N_k}\sum_{i=1}^{N_k} \mathbf{x}_{ik}.$$

Comment devrait-on procéder pour retrouver l'expression de $\widehat{\boldsymbol{\Sigma}}_k$ donnée dans l'article (on demande juste d'expliquer la méthode) ?
4) (1pt) Expliquer l'influence du paramètre h_{kj} de l'équation (5) sur l'estimateur $\widehat{f}_k(\mathbf{x})$.
5) (2pts) Démontrer le résultat de l'équation (6).
6) (2pts) Expliquer en quoi les théorèmes 3.1 et 3.2 permettent d'estimer la matrice $\mathbf{M}$ à partir de données $\{\mathbf{x}_{ik}, i = 1, ..., N_k\}$ (expliquer le principe de l'estimation de $\mathbf{M}$)
7) (1pt) Quelle est la principale motivation pour utiliser la méthode de classification proposée dans la section 4 ?