Partiel de Classification et Reconnaissance des Formes

Mardi 17 janvier 2012, 8h00-10h00

Partiel avec documents autorisés

On considère un vecteur formé à partir d'une image contenant le visage d'un individu que l'on cherche à reconnaître. On dispose d'une base d'apprentissage constituée d'images de k individus (chaque individu est associé à une classe notée ω_i pour $i = 1, ..., k$). On suppose que la base d"apprentissage est constituée de n_i images associées à l'individu i et on notera $n = \sum_{i=1}^{k} n_i$ le nombre total d'images de cette base d'apprentissage. On suppose que la loi du vecteur $\mathbf{x}$ conditionnellement à la classe ω_i (formé à partir des images du $i^{ème}$ individu) est définie comme suit

$$f(\mathbf{x} | \omega_i) = \mathcal{N}(\mathbf{m}_i, \mathbf{\Sigma}_i)$$

avec

$$\mathbf{m}_i = \left(\mathbf{0}_{n_1}^T, ..., \mathbf{0}_{n_{i-1}}^T, \boldsymbol{\mu}_i, \mathbf{0}_{n_{i+1}}^T, ..., \mathbf{0}_{n_k}^T\right)^T$$

où $\mathbf{0}_m = (0, ..., 0)^T$ est le vecteur nul de $\mathbb{R}^m$ et $\boldsymbol{\mu}_i = (\mu_1, ..., \mu_{n_i})^T$ est un vecteur de $\mathbb{R}^{n_i}$. Pour simplifier, on supposera que $\mathbf{\Sigma}_i = \sigma^2 I_n$, où I_n est la matrice identité de dimension $n \times n$. On suppose que les classes $\omega_1, ..., \omega_k$ ont des probabilités a priori

$$p_i = \frac{1}{k}, \quad \forall i = 1, ..., k.$$

1) **Règle de décision Bayésienne :** montrer que la règle de décision Bayésienne associée au problème énoncé ci-dessus s'exprime sous la forme

$$d^*(\mathbf{x}) = \omega_i \Leftrightarrow \mathbf{a}_{ij}^T \mathbf{x} \leq b_{ij} \qquad \forall j \tag{1}$$

où $\mathbf{a}_{ij}$ est un vecteur dépendant de $\mathbf{m}_i$ et $\mathbf{m}_j$ et b_{ij} est un réel dépendant également de $\boldsymbol{\mu}_i$ et $\boldsymbol{\mu}_j$ Dans le cas particulier de deux classes ($k = 2$) avec $n_1 = n_2 = 1$, représenter dans le plan (x_1, x_2) les zones d'acceptation des classes ω_1 et ω_2 (on pourra prendra $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 1$).

2) **Probabilité d'erreur :** on rappelle que si $\mathbf{x}$ est un vecteur Gaussien de vecteur moyenne $\mathbf{m}$ et de matrice de covariance $\mathbf{\Sigma}$ (ce que l'on note $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$), et que $\mathbf{a}$ est un vecteur non nul, alors

$$\mathbf{a}^T \mathbf{x} \sim \mathcal{N}\left(\mathbf{a}^T \mathbf{m}, \mathbf{a}^T \mathbf{\Sigma} \mathbf{a}\right).$$

De plus, on note F la fonction de répartition d'une loi normale $\mathcal{N}(0, 1)$ définie par

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

Pour simplifier, on suppose qu'on a un problème à deux classes ($k = 2$) dans toute la suite de ce problème. Exprimer la probabilité d'erreur du classifieur en fonction de F, σ, μ_1 et μ_2. Expliquer le comportement de cette probabilité d'erreur lorsque σ diminue ou lorsque la norme des vecteurs $\boldsymbol{\mu}_1$ et $\boldsymbol{\mu}_2$ augmente.

3) **Estimation paramétrique :** dans les applications pratiques, on ne connaît pas $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ et $\boldsymbol{\sigma}^2$. Expliquer comment on peut estimer les deux vecteurs moyennes $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ et la variance σ^2 à l'aide d'une base d'apprentissage constituée d'éléments des deux classes ω_1 et ω_2 (on pourra noter $\mathbf{t}^{(1)}, ..., \mathbf{t}^{(n_1)}$ les éléments de la classe ω_1 et $\mathbf{u}^{(1)}, ..., \mathbf{u}^{(n_2)}$ les éléments de la classe ω_2).

4) **Questions de cours**

- 4.1) Expliquer le principe de l'analyse en composantes principales. Comment peut-on choisir le nombre d'axes principaux ?
- 4.2) Soit $\mathbf{T}$ la matrice de covariance d'un ensemble de données expertisées et $\mathbf{B}, \mathbf{S}$ les matrices de covariances inter-classe et intra-classe. Expliquer pourquoi les vecteurs propres de $\mathbf{T}^{-1}\mathbf{B}$ et de $\mathbf{S}^{-1}\mathbf{B}$ sont les mêmes. Dans les applications pratiques, est-il plus simple de rechercher les vecteurs propres et valeurs propres de $\mathbf{T}^{-1}\mathbf{B}$ ou de $\mathbf{S}^{-1}\mathbf{B}$? Pourquoi ?
- 4.3) On désire construire une fonction discriminante linéaire pour un problème de classification à deux classes. Quelle fonction de coût $J(\mathbf{w})$ utilise-t-on pour obtenir l'algorithme LMS ? Pourquoi ne peut-on pas utiliser la fonction de coût du filtre de Wiener-Hopf dans la plupart des applications pratiques ?
- 4.4) Dans la théorie des machines à vecteurs supports, qu'est ce que la marge de la base d'apprentissage et qu'appelle-t-on vecteur support ?
- 4.5) Pourquoi est-il utile d'utiliser un prétraitement non-linéaire avant d'appliquer la théorie des machines à vecteurs supports ? Expliquer ce qu'est un noyau de Mercer et son intérêt pour les machines à vecteurs supports.
- 4.6) Quel algorithme utilise-t-on pour mettre à jour les poids d'un réseau de neurones ? Qu'est ce que le théorème d'approximation universelle de Cybenko ?

5) **Questions portant sur l'article**

- 5.1) Dans l'application proposée dans l'article, comment est constitué le vecteur y ?
- 5.2) Quelle est la motivation pour résoudre le problème $\left(l^0\right)$ défini par l'équation (5) ?
- 5.3) Pourquoi préfère-t-on résoudre le problème $\left(l^1\right)$ défini par l'équation (6) au problème $\left(l^0\right)$ défini par l'équation (5) ?
- 5.4) Pourquoi doit-on en pratique remplacer le problème $\left(l^1\right)$ défini par l'équation (6) par le problème $\left(l^1_s\right)$ défini par l'équation (10) ?
- 5.5)Expliquer avec soin la règle de classification (12) (développer la réponse)
- 5.6) Comment voit-on sur la figure 3 que la méthode proposée marche bien pour l'image test ? Inversement, comment voit-on sur la figure 4 que la méthode basée sur une minimisation l_2 ne marche pas pour l'image test ?
- 5.7) A quoi sert la règle de décision (15) ?
- 5.8) En présence d'occlusion, que représente le vecteur $\mathbf{e}_0$ dans (19) ?
- 5.9) A quelle condition sur $\mathbf{e}_0$ peut on estimer $\mathbf{w}_0 = \left[\mathbf{x}_0^T, \mathbf{e}_0^T\right]^T$ sans erreur ?