Partiel de Classification et Reconnaissance des Formes

Vendredi 23 janvier 2009, 8h00-10h00

Partiel avec documents autorisés

1) Règle de décision Bayésienne

- On considère un problème de classification à deux classes C_0 et C_1 de probabilités a priori $\pi_0 = P(C_0)$ et $\pi_1 = P(C_1)$. On suppose que les vecteurs de données à classifier sont des éléments de $\mathbb{R}^n$. On note $x = [x(1), ..., x(n)]^T$ un vecteur à classifier et $z \in \{-1, 1\}$ une étiquette indiquant la classe de ce vecteur x ($z = -1$ si $x \in C_0$ et $z = 1$ si $x \in C_1$). On suppose que les densités de x conditionnellement aux classes C_0 et C_1 sont gaussiennes de moyennes μ_0 et μ_1 (avec $\mu_0 \neq \mu_1$) et de même matrice de covariance Σ.

  – Déterminer la probabilité a posteriori $P(C_1| x)$ et montrer qu'elle peut s'écrire

  $$P(z = 1| x) = P(C_1| x) = \frac{1}{1 + \exp(-x^T\beta - \beta_0)}$$

  où $\beta = (\beta_1, ..., \beta_n)^T$ est un vecteur de $\mathbb{R}^n$ et β_0 un scalaire que l'on donnera en fonction de μ_0, μ_1, π_0 et π_1. Déterminer de la même façon $P(z = -1| x) = P(C_0| x)$. En déduire

  $$P(z = \varepsilon| x) = \frac{1}{1 + \exp[-\varepsilon(x^T\beta + \beta_0)]}, \qquad \varepsilon \in \{-1, 1\}.$$

  On notera $P(\varepsilon| x) = P(z = \varepsilon| x)$ dans ce qui suit.

  – Les auteurs de l'article posent $P_1 = P(z = 1| x)$ (notation que l'on pourrait critiquer car P_1 dépend de x!). Montrer que l'équation (1) de l'article est correcte.

  – En utilisant l'expression de $P(z = \pm 1| x)$ donnée ci-dessus, déterminer la règle de décision Bayésienne associée à ce problème (en fonction de β et β_0) dans le cas où les deux classes C_0 et C_1 sont équiprobables et pour des fonctions de coût $c_{ij} = 1-\delta_{ij}$. Représenter les régions associées aux décisions $d^*(x) = C_0$ et $d^*(x) = C_1$ dans le cas particulier $n = 2$ (i.e. $x = (x_1, x_2)^T$).

- Montrer que la loi de $T(x) = x^T\beta + \beta_0$ conditionnellement à la classe C_0 est une loi normale dont on déterminera les paramètres m_0 et σ^2 en fonction de β, β_0, μ_0 et Σ (i.e. $T(x)| C_0 \sim \mathcal{N}(m_0, \sigma^2)$). Montrer que la loi de $T(x) = x^T\beta+\beta_0$ conditionnellement à la classe C_1 est une loi normale $\mathcal{N}(m_1, \sigma^2)$ où on précisera la valeur de $m_1$en fonction de β, β_0 et μ. Montrer enfin que $m_1 = -m_0$ (on supposera dans la suite $m_1 > 0$ et $m_0 < 0$)

  En déduire que la probabilité d'erreur du classifieur Bayésien s'écrit

  $$P_e = F\left(\frac{m_0}{\sigma}\right) = F\left(-\frac{m_1}{\sigma}\right) \text{ avec } F(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

  Que se passe-t-il lorsque $\sigma \to 0$?

2) Apprentissage

On désire estimer β et β_0 à l'aide d'une base d'apprentissage constituée de p vecteurs $x_1, ..., x_p$ de $\mathbb{R}^n$dont les classes sont connues. Chaque vecteur x_i est muni d'une étiquette $z_i \in \{-1, 1\}$ telle que

$$P\left(z_i | x_i\right) = \frac{1}{1 + \exp\left[-z_i\left(x_i^T \beta + \beta_0\right)\right]}, \qquad z_i \in \{-1, 1\}$$

- En supposant que les étiquettes $z_1, ..., z_p$ sont indépendantes conditionnellement à $x_1, ..., x_p$, déterminer

$$g(\beta, \beta_0) = \ln P\left(z_1, ..., z_p | x_1, ..., x_p\right)$$

- Déterminer le système de $n+1$ équations à $n+1$ inconnues que l'on doit résoudre pour déterminer β et β_0 qui maximisent $g(\beta, \beta_0)$.
- Comment peut-on en pratique déterminer β et β_0 ?

3) Perceptron Multi-couches

Dans la partie 2.5, les auteurs proposent d'utiliser un perceptron multi-couches pour la classification des signaux EEG issus de patients normaux ou épileptiques.

- Quelle est la structure du perceptron proposée dans cet article (on définira avec soin le nombre et la nature des entrées du réseau, le nombre de couches et de noeuds par couche, ainsi que la sortie désirée du réseau pour chaque classe).
- On suppose dans la suite que les non-linéarités du réseau sont des fonctions sigmoidales définies par

$$f(x) = \frac{1}{1 + e^{-\alpha x}}$$

  avec $\alpha > 0$. Donner les règles de mise à jour des poids du réseau issues de la règle du gradient (on notera $u_1, ..., u_m, a$ les poids et le biais de la dernière couche et $w_{i,j}$ et b_j les poids et les biais de la première couche).

4) Questions diverses sur l'article

- Comment les auteurs de l'article motivent-ils l'utilisation de la transformée en ondelettes pour détecter les EEG issus de patients épileptiques (plutôt par exemple que d'utiliser le spectre des signaux EEG) ?
- A quoi correspondent les quatre signaux représentés sur les figures 1 ?
- Comment peut-on déterminer A_i à partir de A_{i+1} et D_{i+1} sur la figure 3 ?
- Comment les auteurs ont-ils procédé pour déterminer le nombre de noeuds du perceptron représenté sur la figure 5 ?
- Á plusieurs reprises, les auteurs parlent de l'algorithme "Backpropagation". De quel algorithme s'agit-il ?
- p. 95, les auteurs parlent de "Receiver operating characteristic (ROC) analysis". De quoi s'agit-il ?