Examen de Classification
et de Reconnaissance des Formes
Mardi 18 Décembre 2007

1) Règle de décision Bayésienne

- Afin de préciser la règle de décision Bayésienne évoquée dans la partie 2 (Approach), on considère un exemple simple où les données sont dans $\mathbb{R}^2$ et on note $z = (x, y)$ un vecteur à classifier. On suppose que les densités de z conditionnellement aux classes ω_1 et ω_2 sont gaussiennes de moyennes $\mu_1 = (-a, a)$ et $\mu_2 = (a, -a)$ avec $a > 0$ et de matrice de covariance $\Sigma = \sigma^2 \mathbb{I}_2$, où $\mathbb{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ est la matrice identité de taille 2×2. Déterminer la règle de décision Bayésienne associée à ce problème dans le cas où les deux classes ω_1 et ω_2 sont équiprobables et pour des fonctions de coût $c_{ij} = 1 - \delta_{ij}$. Représenter dans le plan, les régions associées aux décisions $d^*(z) = \omega_1$ et $d^*(z) = \omega_2$.

- Montrer que la probabilité d'erreur associée à la règle ci-dessus s'écrit

$$P_e(a, \sigma) = \int_{-\infty}^{\infty} f(u) \Phi \left(u - \frac{2a}{\sigma} \right) du$$

avec

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{u^2}{2} \right) \text{ et } \Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) dx$$

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

- Conformément à l'article, expliciter la règle de décision Bayésienne lorsque les classes ω_1 et ω_2 sont équiprobables avec les coûts suivants

$$c_{11} = c_{22} = 0 \text{ et } c_{12} = 2c_{21} = 2$$

Représenter dans le plan, les régions associées aux décisions $d^*(z) = \omega_1$ et $d^*(z) = \omega_2$.

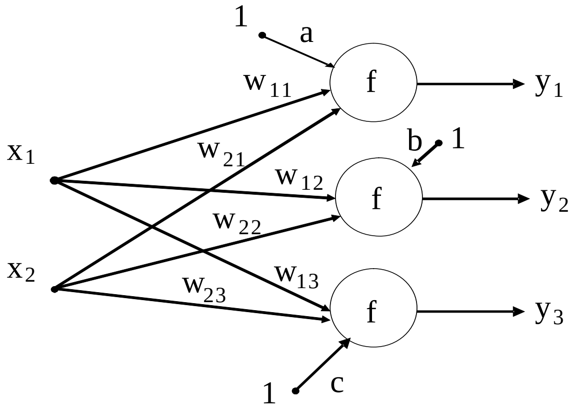
2) Apprentissage Bayésien.

- Dans le cas où les densités conditionnelles $f(z|\omega_1)$ et $f(z|\omega_2)$ sont gaussiennes de moyennes μ_1 et μ_2 inconnues et de matrices de covariance $\Sigma_1 = \sigma_1^2 \mathbb{I}_2$ et $\Sigma_2 = \sigma_2^2 \mathbb{I}_2$ inconnues, on peut estimer μ_1, μ_2, σ_1^2 et σ_2^2 à l'aide d'une base d'apprentissage constituée de n_1 vecteurs $v_1, ..., v_{n_1}$associés à la classe ω_1 et de n_2 vecteurs $w_1, ..., w_{n_2}$ associés à la classe ω_2. Quels sont les estimateurs du maximum de vraisemblance de μ_1, μ_2, σ_1^2 et σ_2^2 notés $\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma_1^2}$ et $\widehat{\sigma_2^2}$. Expliciter la règle de classification obtenue à l'aide de ces estimateurs dans le cas où les deux classes ω_1 et ω_2 sont équiprobables et pour des fonctions de coût $c_{ij} = 1 - \delta_{ij}$.

- Dans les parties 2.1 et 2.2, les auteurs parlent de méthodes d'estimation de densités basées sur des histogrammes dans le cas de données uni-dimensionnelles ("one-feature problem"). Expliquer brièvement quelels méthodes d'estimation de densité s'appliquent au cas de données multi-dimensionnelles. Expliquer également la phrase "It should be noted that the distributions are now represented non-parametrically".

3) Perceptron Multi-couches

Dans la partie 3, les auteurs proposent d'utiliser un perceptron multi-couches pour des données synthétiques issues de trois classes Gaussiennes de moyennes μ_1, μ_2 et μ_3 et de matrices de covariances Σ_1, Σ_2 et Σ_3 (notées σ_1, σ_2 et σ_3 dans l'article). Ce perceptron est représenté ci-dessous :



- On suppose dans un premier temps que f est linéaire telle que $f(t) = t$ et on définit les sorties désirées du réseau à l'instant n notées $d_1(n), d_2(n)$ et $d_3(n)$ par

$$\begin{aligned}
d_1(n) &= 1 \text{ si } x(n) = [x_1(n), x_2(n)] \in \omega_1 \text{ et } d_1(n) = 0 \text{ sinon} \\
d_2(n) &= 1 \text{ si } x(n) = [x_1(n), x_2(n)] \in \omega_2 \text{ et } d_2(n) = 0 \text{ sinon} \\
d_3(n) &= 1 \text{ si } x(n) = [x_1(n), x_2(n)] \in \omega_3 \text{ et } d_3(n) = 0 \text{ sinon}
\end{aligned}$$

  On pose $w_1 = (w_{11,}w_{21}, a)$, $w_2 = (w_{12,}w_{22}, b)$ et $w_3 = (w_{13,}w_{23}, c)$. Déterminer les trois systèmes de trois équations de Wiener-Hopf obtenus en cherchant w_1, w_2 et w_3 qui minimisent les erreurs quadratiques moyennes $E\left[e_i^2(n)\right]$ pour $i = 1, 2, 3$ avec

$$e_i(n) = d_i(n) - y_i(n)$$

- On suppose dans la suite que f est une fonction sigmoidale définie par

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

  avec $\alpha > 0$. Donner les règles de mise à jour des paramètres w_{11} et w_{21} et du biais a issues de la règle du gradient (de manière très similaire, on en déduirait les règles de mise à jour associées aux vecteurs w_2 et w_3 mais on ne demande pas de le faire ici).

4) Règle DCMS

On s'intéresse pour terminer à la règle de décision DCMS proposée dans cet article.

- Expliquer la relation (2) qui exprime le coût moyen d'affecter x à la classe c (comme d'habitude, on choisira $c_{ii} = 0$, $\forall i = 1, ..., c$).
- Déterminer la relation similaire à (2) vérifiée par $E[\nu_c]$, qui est le cout moyen de ne pas prendre la décision "classe c" alors que les données proviennent de la classe c.
- Dans la partie 3.1, les auteurs disent qu'un tiers des données sert pour déterminer les seuils optimaux. Expliquer comment ces seuils sont déterminés.
- Dans la partie 3.2, les auteurs étudient un classifieur basé sur la distance de Mahalanobis et le classifieur KNN. Expliquer somairement le fonctionnement de ces deux classifieurs.