Correction du Partiel de Classification et Reconnaissance des Formes

du lundi 25 janvier 2010, 8h00-10h00 January 25, 2010

1) Règle de décision Bayésienne

- Pour $M = 1, P = 1, c_1 = 1, N = 2$ et $\mathbf{h} = (h_0, h_1)^T = (1, 0.5)^T$, on a

$$\begin{aligned}\mathbf{X}(n) &= x(n) \\ &= \widehat{x}(n) + e(n) \\ &= h_0 u(n) + h_1 u(n-1) + e(n) \\ &= u(n) + \frac{1}{2}u(n-1) + e(n)\end{aligned}$$

  – Dans le cas où $e(n) = 0$ (pas de bruit), on a

$$\mathbf{X}(n) = u(n) + \frac{1}{2}u(n-1)$$

  Conditionnellement à la classe C_1 définie par $u(n) = 1$, on a

$$\begin{aligned}\mathbf{X}(n) &= 1 + \frac{1}{2}u(n-1) \\ &= \begin{cases} \frac{3}{2} \text{ avec probabilité } \frac{1}{2} \text{ (correspond à } u(n-1) = 1) \\ \frac{1}{2} \text{ avec probabilité } \frac{1}{2} \text{ (correspond à } u(n-1) = -1) \end{cases}\end{aligned}$$

  Conditionnellement à la classe C_2 définie par $u(n) = -1$, on a

$$\begin{aligned}\mathbf{X}(n) &= -1 + \frac{1}{2}u(n-1) \\ &= \begin{cases} -\frac{3}{2} \text{ avec probabilité } \frac{1}{2} \text{ (correspond à } u(n-1) = -1) \\ -\frac{1}{2} \text{ avec probabilité } \frac{1}{2} \text{ (correspond à } u(n-1) = 1) \end{cases}\end{aligned}$$

  – En présence de bruit, on

$$\begin{aligned}\mathbf{X}(n)|\, C_1 &= 1 + \frac{1}{2}u(n-1) + e(n) \\ &= \begin{cases} \frac{3}{2} + e(n) \text{ avec probabilité } \frac{1}{2} \text{ (correspond à } u(n-1) = 1) \\ \frac{1}{2} + e(n) \text{ avec probabilité } \frac{1}{2} \text{ (correspond à } u(n-1) = -1) \end{cases}\end{aligned}$$

  d'où

$$f\left[\mathbf{X}(n)|\, C_1\right] = \frac{1}{2}\mathcal{N}\left(\frac{3}{2}, \sigma^2\right) + \frac{1}{2}\mathcal{N}\left(\frac{1}{2}, \sigma^2\right)$$

  et

$$\begin{aligned}\mathbf{X}(n)|\, C_2 &= -1 + \frac{1}{2}u(n-1) + e(n) \\ &= \begin{cases} -\frac{1}{2} + e(n) \text{ avec probabilité } \frac{1}{2} \text{ (correspond à } u(n-1) = 1) \\ -\frac{3}{2} + e(n) \text{ avec probabilité } \frac{1}{2} \text{ (correspond à } u(n-1) = -1) \end{cases}\end{aligned}$$

d'où

$$f\left[\mathbf{X}(n)|\, C_2\right] = \frac{1}{2}\mathcal{N}\left(-\frac{3}{2},\sigma^2\right) + \frac{1}{2}\mathcal{N}\left(-\frac{1}{2},\sigma^2\right)$$

Le classifieur Bayésien est défini par

$$d^*\left[\mathbf{X}(n)\right] = C_1 \Longleftrightarrow P\left(C_1|\, \mathbf{X}(n)\right) = \frac{f\left[\mathbf{X}(n)|\, C_1\right] P\left(C_1\right)}{f\left[\mathbf{X}(n)\right]} \geq P\left(C_2|\, \mathbf{X}(n)\right) = \frac{f\left[\mathbf{X}(n)|\, C_2\right] P\left(C_2\right)}{f\left[\mathbf{X}(n)\right]}$$

avec $P\left(C_1\right) = P\left(C_2\right) = \frac{1}{2}$. On a donc $d^*\left[\mathbf{X}(n)\right] = C_1$ si

$$\exp\left(-\frac{\left[\mathbf{X}(n)-\frac{3}{2}\right]^2}{2\sigma^2}\right) + \exp\left(-\frac{\left[\mathbf{X}(n)-\frac{1}{2}\right]^2}{2\sigma^2}\right) \geq \exp\left(-\frac{\left[\mathbf{X}(n)+\frac{3}{2}\right]^2}{2\sigma^2}\right) + \exp\left(-\frac{\left[\mathbf{X}(n)+\frac{1}{2}\right]^2}{2\sigma^2}\right)$$

- Pour $M = 2, P = 1, c_1 = 1, N = 2$ et $\mathbf{h} = (h_0, h_1) = (1, 0.5)$, on a

$$\begin{aligned}\mathbf{X}(n) &= \begin{bmatrix} x(n) \\ x(n-1) \end{bmatrix} \\ &= \begin{bmatrix} u(n) + \frac{1}{2}u(n-1) + e(n) \\ u(n-1) + \frac{1}{2}u(n-2) + e(n-1) \end{bmatrix}\end{aligned}$$

  - Dans le cas où $e(n) = 0$ (pas de bruit), on a

$$\mathbf{X}(n) = \begin{bmatrix} u(n) + \frac{1}{2}u(n-1) \\ u(n-1) + \frac{1}{2}u(n-2) \end{bmatrix}$$

Conditionnellement à la classe C_1 définie par $u(n) = 1$, on a

$$\begin{aligned}\mathbf{X}(n) &= \begin{bmatrix} 1 + \frac{1}{2}u(n-1) \\ u(n-1) + \frac{1}{2}u(n-2) \end{bmatrix} \\ &= \begin{cases} \begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = 1 \text{ et } u(n-2) = 1) \\ \begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = 1 \text{ et } u(n-2) = -1) \\ \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = -1 \text{ et } u(n-2) = 1) \\ \begin{bmatrix} \frac{1}{2} \\ -\frac{3}{2} \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = -1 \text{ et } u(n-2) = -1) \end{cases}\end{aligned}$$

Conditionnellement à la classe C_2 définie par $u(n) = -1$, on a

$$\begin{aligned}\mathbf{X}(n) &= \begin{bmatrix} -1 + \frac{1}{2}u(n-1) \\ u(n-1) + \frac{1}{2}u(n-2) \end{bmatrix} \\ &= \begin{cases} \begin{bmatrix} -\frac{1}{2} \\ \frac{3}{2} \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = 1 \text{ et } u(n-2) = 1) \\ \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = 1 \text{ et } u(n-2) = -1) \\ \begin{bmatrix} -\frac{3}{2} \\ -\frac{1}{2} \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = -1 \text{ et } u(n-2) = 1) \\ \begin{bmatrix} -\frac{3}{2} \\ -\frac{3}{2} \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = -1 \text{ et } u(n-2) = -1) \end{cases}\end{aligned}$$

– En présence de bruit, on en déduit

$$f\left[\mathbf{X}(n)|\,C_1\right] = \frac{1}{4}\mathcal{N}\left(m_{11}, \sigma^2 I_2\right) + \mathcal{N}\left(m_{12}, \sigma^2 I_2\right) + \frac{1}{2}\mathcal{N}\left(m_{13}, \sigma^2 I_2\right) + \frac{1}{2}\mathcal{N}\left(m_{14}, \sigma^2 I_2\right)$$

On obtient un mélange de quatre Gaussiennes de moyennes

$$m_{11} = \begin{pmatrix} \frac{3}{2} \\ \frac{3}{2} \end{pmatrix}, m_{12} = \begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \end{pmatrix}, m_{13} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, m_{14} = \begin{pmatrix} \frac{1}{2} \\ -\frac{3}{2} \end{pmatrix}$$

De même

$$f\left[\mathbf{X}(n)|\,C_2\right] = \frac{1}{4}\mathcal{N}\left(m_{21}, \sigma^2 I_2\right) + \mathcal{N}\left(m_{22}, \sigma^2 I_2\right) + \frac{1}{2}\mathcal{N}\left(m_{23}, \sigma^2 I_2\right) + \frac{1}{2}\mathcal{N}\left(m_{24}, \sigma^2 I_2\right)$$

avec

$$m_{21} = \begin{pmatrix} -\frac{1}{2} \\ \frac{3}{2} \end{pmatrix}, m_{22} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, m_{23} = \begin{pmatrix} -\frac{3}{2} \\ -\frac{1}{2} \end{pmatrix}, m_{24} = \begin{pmatrix} -\frac{3}{2} \\ -\frac{3}{2} \end{pmatrix}$$

Le classifieur Bayésien est défini par

$$d^*\left[\mathbf{X}(n)\right] = C_1 \Longleftrightarrow P\left(C_1|\,\mathbf{X}(n)\right) = P\left(C_2|\,\mathbf{X}(n)\right)$$

soit

$$d^*\left[\mathbf{X}(n)\right] = C_1 \Longleftrightarrow f\left[\mathbf{X}(n)|\,C_1\right] \geq f\left[\mathbf{X}(n)|\,C_2\right]$$

- Pour $M = 2, P = 3, \mathbf{c} = (c_1, c_2, c_3)^T = (1, 0, -0.9)^T, N = 2$ et $\mathbf{h} = (h_0, h_1)^T = (1, 0.5)^T$, on a

$$\begin{aligned} \mathbf{X}(n) &= \begin{bmatrix} \widehat{x}(n) + e(n) \\ \widehat{x}(n-1) + e(n-1) \end{bmatrix} \\ &= \begin{bmatrix} \widetilde{x}(n) - 0.9\widetilde{x}^3(n) + e(n) \\ \widetilde{x}(n-1) - 0.9\widetilde{x}^3(n-1) + e(n-1) \end{bmatrix} \end{aligned}$$

avec

$$\begin{aligned} \widetilde{x}(n) &= u(n) + \frac{1}{2}u(n-1) \\ \widetilde{x}(n-1) &= u(n-1) + \frac{1}{2}u(n-2) \end{aligned}$$

Donc par rapport au cas précédent, on doit remplacer $\widetilde{x}(n)$ par $\widetilde{x}(n) - 0.9\widetilde{x}^3(n)$. On en déduit

– Pour $e(n) = 0$ et $u(n) = 1$

$$\mathbf{X}(n) = \begin{cases} \begin{bmatrix} -1.54 \\ -1.54 \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = 1 \text{ et } u(n-2) = 1) \\ \begin{bmatrix} -1.54 \\ 0.39 \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = 1 \text{ et } u(n-2) = -1) \\ \begin{bmatrix} 0.39 \\ -0.39 \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = -1 \text{ et } u(n-2) = 1) \\ \begin{bmatrix} 0.39 \\ 1.54 \end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1) = -1 \text{ et } u(n-2) = -1) \end{cases}$$

pour $e(n)=0$ et $u(n)=-1$

$$\mathbf{X}(n)=\begin{cases}\begin{bmatrix}-0.39\\-1.54\end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1)=1 \text{ et } u(n-2)=1)\\ \begin{bmatrix}-0.39\\0.39\end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1)=1 \text{ et } u(n-2)=-1)\\ \begin{bmatrix}1.54\\-0.39\end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1)=-1 \text{ et } u(n-2)=1)\\ \begin{bmatrix}1.54\\1.54\end{bmatrix} & \text{avec probabilité } \frac{1}{4} \text{ (correspond à } u(n-1)=-1 \text{ et } u(n-2)=-1)\end{cases}$$

On retrouve les points noirs et les croix noires de la figure 3 de l'article page 3221.

2) Apprentissage

- On désire estimer σ^2 à l'aide de $\mathbf{X}(n)$ associé à une suite de bits $u(n)$. On a

$$\mathbf{X}(n)=\begin{pmatrix}x(n)\\x(n-1)\\\vdots\\x(n-M+1)\end{pmatrix}=\begin{pmatrix}\widehat{x}(n)+e(n)\\\widehat{x}(n-1)+e(n-1)\\\vdots\\\widehat{x}(n-M+1)+e(n-M+1)\end{pmatrix}$$

Lorsqu'on envoie une suite de bits connus, on peut en déduire les valeurs de $\widehat{x}(n),\widehat{x}(n-1),...,\widehat{x}(n-M+1)$ qui sont donc connues. Puisque les variables aléatoires $e(n),e(n-1),...,e(n-M+1)$ sont indépendantes, la vraisemblance s'écrit

$$\begin{aligned} f\left[\mathbf{X}(n);\sigma^2\right] &= \prod_{m=0}^{M-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\left[x(m)-\widehat{x}(m)\right]^2\right)\\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{M/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{m=1}^{M-1}\left[x(m)-\widehat{x}(m)\right]^2\right)\end{aligned}$$

- En dérivant la log-vraisemblance par rapport à σ^2, on obtient

$$-\frac{M}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{m=0}^{M-1}\left[x(m)-\widehat{x}(m)\right]^2=0$$

d'où

$$\widehat{\sigma^2}_{MV}=\frac{1}{M}\sum_{m=0}^{M-1}\left[x(m)-\widehat{x}(m)\right]^2$$

- La loi *a posteriori* du paramètre σ^2 s'écrit

$$\begin{aligned} f\left[\sigma^2\left|\mathbf{X}(n)\right.\right] &= \frac{f\left[\mathbf{X}(n)\left|\sigma^2\right.\right]f\left(\sigma^2\right)}{\mathbf{X}(n)}\\ &\propto f\left[\mathbf{X}(n)\left|\sigma^2\right.\right]f\left(\sigma^2\right)\end{aligned}$$

On en déduit

$$f\left[\sigma^2\left|\mathbf{X}(n)\right.\right]\propto\frac{1}{\left(\sigma^2\right)^{1+\alpha+M/2}}\exp\left[-\frac{1}{\sigma^2}\left(\beta+\frac{1}{2}\sum_{m=0}^{M-1}\left[x(m)-\widehat{x}(m)\right]^2\right)\right]$$

qui est une loi inverse gamma

$$IG\left(\alpha + \frac{M}{2}, \beta + \frac{1}{2}\sum_{m=0}^{M-1}\left[x(m) - \widehat{x}(m)\right]^2\right)$$

En maximisant le logarithme de cette loi, on obtient

$$\begin{aligned} \widehat{\sigma^2}_{MAP} &= \frac{\beta + \frac{1}{2}\sum_{m=0}^{M-1}\left[x(m) - \widehat{x}(m)\right]^2}{\alpha + \frac{M}{2} + 1} \\ &= \frac{\frac{2\beta}{M} + \widehat{\sigma^2}_{MV}}{\frac{2(\alpha+1)}{M} + 1} \end{aligned}$$

On voit que $\widehat{\sigma^2}_{MAP}$ se comporte comme $\widehat{\sigma^2}_{MV}$ pour M "grand".

3) Fonction discriminante linéaire

- La fonction g est adaptée au problème de classification (1) car sans bruit les points des classes C_1 et C_2 sont linéairement séparables. Donc, lorsque le bruit est faible, on peut penser que ces classes resteront linéairement séparables.

- On a

$$J(\mathbf{w}) = E\left(\left[u(n) - \mathbf{Y}^T(n)\mathbf{w}\right]^2\right)$$

donc

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = -2E\left(\left[u(n) - \mathbf{Y}^T(n)\mathbf{w}\right]\mathbf{Y}(n)\right)$$

En annulant le gradient de $J(\mathbf{w})$, on en déduit

$$E\left[u(n)\mathbf{Y}(n)\right] = \mathbf{w}^T E\left[\mathbf{Y}(n)\mathbf{Y}^T(n)\right]$$

soit

$$\begin{pmatrix} E\left[u(n)x(n)\right] \\ E\left[u(n)x(n-1)\right] \\ E\left[u(n)\right] \end{pmatrix} = \begin{pmatrix} E\left[x^2(n)\right] & E\left[x(n)x(n-1)\right] & m \\ E\left[x(n)x(n-1)\right] & E\left[x^2(n)\right] & m \\ m & m & 1 \end{pmatrix}\mathbf{w}$$

c'est-à-dire

$$\begin{pmatrix} r_{ux}(0) \\ r_{ux}(1) \\ 0 \end{pmatrix} = \begin{pmatrix} r_x(0) & r_x(1) & m \\ r_x(1) & r_x(0) & m \\ m & m & 1 \end{pmatrix}\mathbf{w}$$

On a bien un système d'équations linéaires par rapport à $\mathbf{w}$ dont les coefficients dépendent de m, des autocorrélations $r_x(0)$ et $r_x(1)$ et des intercorrélations $r_{ux}(0) = E\left[u(n)x(n)\right]$ et $r_{ux}(1) = E\left[u(n)x(n-1)\right]$.

- Les règles de mise à jour des poids w_1, w_2 et w_3 obtenues à partir de l'algorithme LMS sont obtenues en dérivant l'erreur instantanée

$$e^2(n) = \left[u(n) - \mathbf{Y}^T(n)\mathbf{w}\right]^2$$

et en utilisant la règle du gradient

$$w_i(n+1) = w_i(n) - \frac{\mu}{2} \left. \frac{\partial e^2(n)}{\partial w_i} \right|_{w_i = w_i(n)}$$

On obtient

$$\begin{aligned} w_1(n+1) &= w_1(n) + \mu e(n) x(n) \\ w_2(n+1) &= w_2(n) + \mu e(n) x(n-1) \\ w_3(n+1) &= w_3(n) + \mu e(n) \end{aligned}$$

avec

$$e(n) = u(n) - \mathbf{Y}^T(n)\mathbf{w}(\mathbf{n})$$

4) Machines à vecteurs supports

On considère dans cette partie le scénario 3 et on considère l'égaliseur SVM présenté sur la figure 2 page 3220 avec $D = 0$.

- Expliquer tout d'abord pourquoi la fonction discriminante g de la question 3) n'est pas adaptée à ce scénario.

  Réponse : comme le montre la figure 3 de l'article page 3221, il n'est pas possible de séparer linéairement les points noirs des croix noires donc une fonction discriminante linéaire ne convient pas.

- Quelle est la règle de classification utilisée par l'égaliseur SVM pour un vecteur d'entrée $\mathbf{X}(n) = [x(n), x(n-1)]^T$?

  Réponse : la règle de classification est

$$\begin{aligned} \widehat{u}(n) &= 1 \text{ si } f(x) > 0 \\ \widehat{u}(n) &= -1 \text{ si } f(x) < 0 \end{aligned}$$

  avec

$$f(x) = \sum_{i \in S} \alpha_i y_i K(\mathbf{X}_i, \mathbf{x}) + b$$

  où y_i est la valeur du bit $u(n)$ associée au $i^{\grave{e}me}$ vecteur $\mathbf{X}_i$ de la base d'apprentissage, $K(.,.)$ est la fonction noyau utilisée pour le classifieur SVM (noyau polynômial dans l'article), S est l'ensemble des indices des vecteurs supports (correspondant à $\alpha_i > 0$). La fonction f dépend aussi des coefficients α_i, $i \in S$ et b dont la détermination fait l'objet des deux questions suivantes.

- Comment détermine-t-on les coefficients α_i intervenant dans l'équation (1) de l'article p. 3218 ?

  Réponse : la détermination des coefficients α_i se fait par résolution d'un problème d'optimisation sous contraintes défini par les équations (2), (2') et (2") de l'article. Ce problème est usuellement appelé problème de programmation quadratique (QP) pour lequel il existe de nombreux algorithmes dans la littérature (voir par exemple la fonction quadprog.m de Matlab).

- Comment calcule-t-on l'offset b intervenant dans l'équation (1) de l'article p. 3218 ? Justifier brièvement ce calcul.

Réponse : l'offset b intervenant dans l'équation (1) de l'article p. 3218 se calcule à l'aide de l'équation (4) de l'article p. 3219. En effet, comme il est précisé dans l'article, pour un vecteur support d'étiquette y_j, on a

$$\sum_{i \in S} \alpha_i y_i K(\mathbf{X}_i, \mathbf{x}) + b = \frac{1}{y_j}$$

Mais puisque $y_j \in \{-1, 1\}$, on a $\frac{1}{y_j} = y_j$, d'où

$$b = y_j - \sum_{i \in S} \alpha_i y_i K(\mathbf{X}_i, \mathbf{x})$$

Comme ce résultat est vérifié pour tout vecteur support, afin de ne privilégier aucun de ces vecteurs, il semble raisonnable d'estimer b en moyennant les valeurs de

$$y_j - \sum_{i \in S} \alpha_i y_i K(\mathbf{X}_i, \mathbf{x})$$

obtenues pour tous les vecteurs supports. On obtient alors l'équation (4) p. 3219.

- Retrouver la valeur de ξ donnée après l'équation (8) p. 3222.

  Réponse : pour déterminer la valeur de ξ, il suffit d'utiliser la relation

$$E\left[e(n)e(n-1)\right] = \sigma_e^2 \rho$$

avec

$$e(n) = \frac{\sigma_e}{\sqrt{1+\xi^2}} w(n) + \frac{\sigma_e \xi}{\sqrt{1+\xi^2}} w(n-1).$$

On a alors

$$\begin{aligned} & E\left[e(n)e(n-1)\right] \\ = \;& E\left[\left\{\frac{\sigma_e}{\sqrt{1+\xi^2}} w(n) + \frac{\sigma_e \xi}{\sqrt{1+\xi^2}} w(n-1)\right\}\left\{\frac{\sigma_e}{\sqrt{1+\xi^2}} w(n-1) + \frac{\sigma_e \xi}{\sqrt{1+\xi^2}} w(n-2)\right\}\right] \\ = \;& \frac{\sigma_e^2 \xi}{1+\xi^2} \end{aligned}$$

d'où

$$\sigma_e^2 \rho = \frac{\sigma_e^2 \xi}{1+\xi^2} \iff \rho \xi^2 - \xi + \rho = 0$$

Les solutions de cette équation du second degré sont

$$\xi = \frac{1}{2\rho} \pm \sqrt{\frac{1}{4\rho^2} - 1}$$

ce qui est en accord avec l'équation (8) p. 3222.