Examen "Méthodes à noyaux"
DEA M.V.A., ENS Cachan
Durée: 2h

Jean-Philippe Vert

23 mars 2004

Exercice 1 : Noyau pour évènements

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité. Montrer que la fonction

$$K : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$$

définie par :

$$\forall\, (A, B) \in \mathcal{A}^2, \quad K\,(A, B) = P\,(A \cap B) - P(A)P(B)$$

est un noyau défini positif.

Exercice 2 : Splines

Soit $H = C_2\,([0,1])$ l'ensemble des fonctions $f : [0,1] \to \mathbb{R}$ deux fois continuement dérivables. Soit $H_1 \subset H$ l'ensemble des fonctions $f \in H$ telles que :

$$f(0) = f'(0) = 0.$$

2.1. Montrer qu'on peut munir H_1 d'une structure d'espace de Hilbert à noyau reproduisant (rkhs) ayant pour norme :

$$\| f \|_{H_1}^2 = \int_0^1 f''(t)^2 dt,$$

et expliciter le noyau reproduisant K_1 correspondant.

2.2. Soit H_2 l'ensemble des fonctions affines de $[0,1] \to \mathbb{R}$ (les fonctions de la forme $f(x) = ax + b$, avec $a, b \in \mathbb{R}$). Montrer que H_2 peut être muni d'une structure de rkhs avec la norme :

$$\| f \|_{H_2}^2 = f(0)^2 + f'(0)^2$$

et expliciter le noyau correspondant K_2.

2.3. En déduire que H muni de la norme :

$$\| f \|_H^2 = \int_0^1 f''(t)^2 dt + f(0)^2 + f'(0)^2$$

est un rkhs et déterminer son noyau reproduisant K.

2.4. Soient $0 < x_1 < \ldots < x_n < 1$ et $(y_1, \ldots, y_n) \in \mathbb{R}^n$. Pour construire une fonction de régression, on considère le problème suivant :

$$\min_{f \in H} \frac{1}{n} \sum_{i=1}^{n} \left(f\left(x_i\right) - y_i\right)^2 + \lambda \int_0^1 f''(t)^2 dt. \tag{1}$$

Montrer que toute solution de (1) peut s'écrire sous la forme :

$$\hat{f}(x) = \sum_{i=1}^{n} \alpha_i K_1(x_i, x) + \beta_1 x + \beta_2,$$

avec $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)' \in \mathbb{R}^n$ et $\boldsymbol{\beta} = (\beta_0, \beta_1)' \in \mathbb{R}^2$.

2.5. Soient I est la matrice identité $n \times n$, M la matrice carré $n \times n$ définie par :

$$M_{i,j} = \begin{cases} K_1(x_i, x_j) & \text{si } i \neq j, \\ K_1(x_i, x_j) + n\lambda & \text{si } i = j, \end{cases}$$

T la matrice $n \times 2$:

$$T = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix},$$

et $\mathbf{y} = (y_1, \ldots, y_n)'$.

Montrer que $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ satisfont :

$$\begin{cases} T'\boldsymbol{\alpha} = 0, \\ M\boldsymbol{\alpha} + T\boldsymbol{\beta} = \mathbf{y}. \end{cases}$$

2.6. En déduire que $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ peuvent s'écrire :

$$\begin{cases} \boldsymbol{\alpha} = M^{-1}\left(I - T\left(T'M^{-1}T\right)^{-1} T'M^{-1}\right)\mathbf{y}, \\ \boldsymbol{\beta} = \left(T'M^{-1}T\right)^{-1} T'M^{-1}\mathbf{y}. \end{cases}$$

2.7. Montrer que

- $\hat{f} \in C_2\left([0,1]\right)$;
- $\hat{f}$ est un polynome de degré 3 sur chaque intervalle $[x_i, x_{i+1}]$ pour $i = 1, \ldots, n-1$;
- $\hat{f}$ est une fonction affine sur les deux intervalles $[0, x_1]$ et $[x_n, 1]$.

$\hat{f}$ est appelée un *spline*.

Exercice 3 : ν-SVM

Soit $\mathcal{X}$ un ensemble, K un noyau défini positif sur $\mathcal{X} \times \mathcal{X}$, et $(\mathcal{H}, \| \, . \, \|)$ le rkhs associé. Soient $(x_1, \ldots, x_n) \in \mathcal{X}^n$ un ensemble de points, et $(y_1, \ldots, y_n) \in \{-1, +1\}^n$ des labels binaires associés.

Pour tout $\rho \geq 0$, on défini la fonction de perte (pour $x \in \mathcal{X}$, $y \in \{-1, 1\}$ et $f : \mathcal{X} \to \mathbb{R}$) :

$$V_\rho\left(f(x), y\right) = \begin{cases} 0 & \text{si } yf(x) \geq \rho, \\ \rho - yf(x) & \text{sinon.} \end{cases}$$

On définit la méthode ν-SVM, pour tout $\nu > 0$, par la solution du probleme suivant :

$$\hat{f} = \arg\min_{f \in \mathcal{H}} \left\{ \min_{\rho \geq 0} \frac{1}{n} \sum_{i=1}^{n} V_\rho\left(f(x_i), y_i\right) - \nu\rho + \| \, f \, \|^2 \right\}. \tag{2}$$

3.1. Ecrire le Lagrangien du problème et le problème dual.

3.2. Soit a le nombre de points x_i vérifiant $V_\rho\left(\hat{f}(x_i), y_i\right) > 0$, et b le nombre de points x_i vérifiant $V_\rho\left(\hat{f}(x_i), y_i\right) \geq 0$ (b est donc le nombre de vecteurs support). Montrer que si (2) est minimum pour une valeur $\hat{\rho} > 0$, alors :

$$\frac{a}{n} \leq \nu \leq \frac{b}{n}.$$

3.3. On rappelle que les SVM étudiées dans le cours calculent la solution de :

$$\hat{f} = \arg\min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} V_1\left(f(x_i), y_i\right) + \lambda \| \, f \, \|^2. \tag{3}$$

Montrez que si la solution de la ν-SVM (2) est atteinte pour une valeur $\hat{\rho} > 0$, alors la même fonction de décision[1] peut être obtenue avec une SVM (3) pour une valeur λ à déterminer.

[1] Rappel : la fonction de décision associée a une fonction $f : \mathcal{X} \to \mathbb{R}$ vaut $+1$ si $f(x) > 0$, -1 sinon.