Examen "Méthodes à noyaux"
M2 Probabilité et Applications, Paris 6
M2 Modélisation aléatoire, Paris 7

Jean-Philippe Vert

A rendre avant le **21 mars 2006**

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 : Complexité de Rademacher

On appelle variable de Rademacher une variable aléatoire σ valant 1 avec probabilité $1/2$ et -1 avec probabilité $1/2$.

3.1. Soient $(u_1, u_2, \ldots, u_N)$ des vecteurs d'un espace de Hilbert muni d'un produit scalaire $< .,. >$. Montrer que si $\sigma_1, \sigma_2, \ldots, \sigma_N$ sont N variables de Rademacher indépendantes, alors :

$$\mathbb{E} \sum_{i,j=1}^{N} \sigma_i \sigma_j < u_i, u_j >= \sum_{i=1}^{N} \| u_i \|^2.$$

3.2. Soit K un noyau défini positif sur un espace $\mathcal{X}$, $\mathcal{H}_K$ l'espace de Hilbert à noyau reproduisant associé, et $B_R = \{f \in \mathcal{H}_K, \| f \|_{\mathcal{H}_R} \leq R\}$. Pour tout ensemble de points $\mathcal{S} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ avec $\mathbf{x}_i \in \mathcal{X}$ $(i = 1, \ldots, N)$, et $\sigma_1, \sigma_2, \ldots, \sigma_N$ un ensemble de N variables de Rademacher indépendantes, montrer que

$$\mathbb{E} \sup_{f \in B_R} \left| \sum_{i=1}^{N} \sigma_i f(\mathbf{x}_i) \right| \leq R \sqrt{\sum_{i=1}^{N} K(\mathbf{x}_i, \mathbf{x}_i)}.$$