Espaces de Hilbert à noyeaux reproduisants (RKHS)

1 Motivations

1.1 Régularisation de Tikhonov

On considère un problème d'estimation : $f(\beta, x) = y$.

Estimation linéaire : $(x_i, y_i)_{i=1}^n \in \mathbb{R}^p \times \mathbb{R}$, avec p la dimension des données.

$X \in \mathbb{R}^{n\times p} = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}$, $Y \in \mathbb{R}^n = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$.

Thikonov : $\min\limits_{\beta\in\mathbb{R}^p} \frac{1}{2}\|X\beta - Y\|_2^2 + \frac{\lambda}{2}\|\beta\|_2^2$.v avec $\lambda > 0$.

On a la CN1 (condition nécessaire d'ordre 1) : $\nabla f(\beta) = 0$.
$\Leftrightarrow X^T(X\beta - Y) + \lambda\beta = 0$.
$\Leftrightarrow X^TX\beta + \lambda\beta = X^TY$.
$\Leftrightarrow (X^TX + \lambda I_p)\beta = X^TY$.
$\Leftrightarrow \hat{\beta} = (X^TX + \lambda I_p)^{-1}X^TY$.

Donc la prédiction sera : $\hat{Y} = X\hat{\beta} = X(X^TX + \lambda I_p)^{-1}X^TY$.

Complexité dépend de X^TX ou XX^T donc $\mathcal{O}(p^2n)$ ou $\mathcal{O}(n^2p)$.

Modèle quadratique : $x^T\theta x + x\beta = \sum_{k=0}^p x_k\beta_k + \sum_{k,l} x_kx_l\theta_{kl}$.

2 inconnues : $\beta \in \mathbb{R}^p$ et $\theta \in \mathbb{R}^{p\times p}$.

Soit $\tilde{\beta} \in \mathbb{R}^{p+p^2} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \\ \theta_{11} \\ \vdots \\ \theta_{pp} \end{pmatrix}$, $\tilde{x_i} = \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^p \\ x_i^1x_i^1 \\ \vdots \\ x_i^px_i^p \end{pmatrix}$, $\tilde{X} = \begin{pmatrix} \tilde{x_1}^T \\ \vdots \\ \tilde{x_n}^T \end{pmatrix}$.

Complexité : $\tilde{X}^T\tilde{X}$ ou $\tilde{X}\tilde{X}^T$ en $\mathcal{O}(np^4)$ ou $\mathcal{O}(n^2p^2)$.

1. Est-ce qu'il existe une fonction qui permet de rester sur une résolution linéaire tout en augmentant la dimension du modèle ?
   $\rightarrow$ Oui, mais (quasi) jamais explicitée.

2. Peut-on s'épargner la complexité du modèle en gardant la complexité des données initiales ?
   $\rightarrow$ Oui, avec un noyau.

Ici, $\begin{array}{rccl} \phi_i : & \mathbb{R}^p & \longrightarrow & \mathbb{R}^{p+p^2} \\ & x_i & \longmapsto & \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^p \\ x_i^1 x_i^1 \\ \vdots \\ x_i^p x_i^p \end{pmatrix} = \tilde{x}_i \end{array}$

Si on se ramène à $\tilde{x_i}^T \tilde{x}_j = (1 + x_i^T x_j)^2$, on n'a plus que des produits de $(x_i^T x_j)$, donc une complexité en $\mathcal{O}(np^2)$.

On définit une application $k : \mathbb{R}^p \times \mathbb{R}^p \longrightarrow \mathbb{R}$ tel que $k(x_i, x_j) = (1 + x_i^T x_j)^2$.

On cherche à prédire $y \in \mathbb{R}$ avec un prédicteur $x \in \mathbb{R}^p$.

- Afin d'appliquer l'approche de régularisation de Tikhonov, on va chercher $\phi : \mathbb{R}^p \longrightarrow \mathbb{R}^d$ avec d à préciser ($d \in \{\mathbb{N}, \infty\}$).
- On va appliquer la régularisatiion aux données transformées $(\phi(x_i))_{i \in [\![1,n]\!]}$ comme prédicteur.
- On va être amené à construire un noyau $K \in \mathcal{M}_p(\mathbb{R})$ avec $K_{ij} = \langle \phi(x_i), \phi(x_j) \rangle$.
  $\rightarrow$ On va chercher $\begin{array}{rccl} k : & \mathbb{R}^p \times \mathbb{R}^p & \longrightarrow & \mathbb{R} \\ & (x_i, x_j) & \longmapsto & \langle \phi(x_i), \phi(x_j) \rangle \end{array}$

2 Espaces de Hilbert à noyeau reproduisant

Soit $\mathcal{H}$ un espace de Hilbert (muni du produit scalaire $\langle \cdot, \cdot \rangle_{\mathcal{H}}$) et $X \subset \mathcal{H}$.

Propriété

Soit $\phi : X \longrightarrow \mathcal{H}$.
On définit $\begin{array}{rccl} k : & X \times X & \longrightarrow & \mathbb{R} \\ & (x, x') & \longmapsto & \langle \phi(x), \phi(x') \rangle_{\mathcal{H}} \end{array}$

Soit K la matrice définie par :
$\forall i, j \in [\![1, n]\!], \quad K_{ij} = k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$.

Alors K est symétrique positive.

- Soit $\alpha \in \mathbb{R}^n$, $(x_i) \in X^n$.

$$
\begin{aligned}
\text{On a : } \alpha^T K \alpha &= \sum_{i,j} \alpha_i \alpha_j k(x_i, x_j) \\
&= \sum_{i,j} \alpha_i \alpha_j \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}} \\
&= \langle \sum_{i}^{n} \alpha_i \phi(x_i), \sum_{j}^{n} \alpha_j \phi(x_j) \rangle_{\mathcal{H}} \\
&= ||\mathcal{Z}||_{\mathcal{H}}^2 \geq 0 \qquad \text{avec } \mathcal{Z} = \sum_{i}^{n} \alpha_i \phi(x_i)
\end{aligned}
$$

Définition

Un *noyau défini positif* (ou *noyau*) est une application symétrique :
$k : X \times X \longrightarrow \mathbb{R}$ telle que $\forall (x_i) \in X^n$, la matrice $K \in \mathcal{M}_n(\mathbb{R})$ définie par $K_{ij} = k(x_i, x_j)$ est symétrique positive.

Propriété

Soit $(k_i)_{i \in J}$ une famille de noyaux.

i) Soit $(\alpha_i, \alpha_j) \in \mathbb{R}_+^2$. Alors $\alpha_i k_i + \alpha_j k_j$ est un noyau.

ii) Si $\exists \lim_{p \to +\infty} k_p(x, x'), \forall (x, x') \in X^2$, alors k la limite simple de $(k_p)_{p \in J}$ est un noyau.

iii) Le produit (terme à terme) $k_i k_j$ est un noyau.
$k(x, x') = k_i(x, x') k_j(x, x')$

Exercice :

1. Noyeau linéaire : $X = \mathbb{R}^p$ et $\forall (x, x') \in X^2, \quad k(x, x') = x^T x'$.
2. Noyeau polynomial : $X = \mathbb{R}^p$ et $\forall (x, x') \in X^2, \quad k(x, x') = (1 + x^T x')^d$.
3. Noyeau exponentiel : $X = \mathbb{R}^p$ et $\forall (x, x') \in X^2, \quad k(x, x') = \exp(-\frac{||x - x'||^2}{2\sigma^2})$. avec $\sigma > 0$.

Montrer que les noyeaux polynomial et exponentiel sont des noyaux.

Correction :

1. Noyeau par définition.
2. Par produit de noyaux.
3. $k(x, x') = \exp(-\frac{||x - x'||^2}{2\sigma^2}) = \exp(-\frac{||x||^2}{2\sigma^2}) \exp(\frac{x^T x'}{\sigma^2}) \exp(-\frac{||x'||^2}{2\sigma^2}) = k_1(x, x') k_2(x, x')$.
   avec $k_1(x, x') = \exp(-\frac{||x||^2}{2\sigma^2}) \exp(-\frac{||x'||^2}{2\sigma^2})$ et $k_2(x, x') = \exp(\frac{x^T x'}{\sigma^2})$.

   $k_1(x, x') = \langle \phi(x), \phi(x') \rangle$ avec $\phi(x) = \exp(-\frac{||x||^2}{2\sigma^2})$. k_1 noyeau (proposition 1).
   $k_2(x, x') = \sum_{i=0}^{+\infty} \frac{1}{i!} \left(\frac{x^T x'}{\sigma^2} \right)^i$. k_2 noyeau (proposition 2).

Théorème de Moore-Aronszajn

Soit k un noyeau, $k : X \times X \longrightarrow \mathbb{R}$.
Alors il existe un espace de Hilbert $\mathcal{H}$ et une application $\phi : X \longrightarrow \mathcal{H}$ tel que $\forall (x, x') \in X^2, \quad k(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$.

► Soit $\mathcal{H} = \{f : X \to \mathbb{R}$ tel que $\exists n \in \mathbb{N}, \exists (\alpha_j)_{j \in [\![1,n]\!]} \in \mathbb{R}^n$, $\exists (x_j)_{j \in [\![1,n]\!]} \in X^n, \quad f(\cdot) = \sum_{j=1}^{n} \alpha_j k(\cdot, x_j)\}$.

Texte manquant.