①

Leçon n° 5 : Méthodes à noyaux 1

Idée : Beaucoup de problèmes peuvent être résolus de manière linéaire en augmentant la dimension (voire en passant par de la dimension infinie).

I. Motivation et théorème de représentation

Ridge en dimension infinie :

- H : Espace de Hilbert
- $$\hat{\theta} \in \underset{\theta \in H}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, \langle \varphi(x_i), \theta \rangle_H) + \frac{\lambda}{2} \|\theta\|_H^2$$

Proposition (Théorème de représentation) :

Le problème $\hat{\theta} \in \underset{\theta}{\operatorname{argmin}}\, \Psi(\langle \varphi(x_1), \theta \rangle_H, \dots, \langle \varphi(x_n), \theta \rangle_H, \|\theta\|_H^2)$

où Ψ est strictement croissante admet une solution

$$\hat{\theta} \in \operatorname{Vect}\{\varphi(x_1), \dots, \varphi(x_n)\}$$

<u>preuve</u> : Notons $H_D = \mathrm{Vect}\{\varphi(x_1), \dots, \varphi(x_n)\}$.

Soit $\theta \in H$. Notons $\theta = \theta_D + \theta_D^\perp$ où $\theta_D \in H_D$ et $\theta_D^\perp \in H_D^\perp$

Alors $\Psi(\langle \varphi(x_1), \theta\rangle_H, \dots, \langle \varphi(x_n), \theta\rangle_H, \|\theta\|_H^2)$

$$= \Psi(\langle \varphi(x_1), \theta_D\rangle_H, \dots, \langle \varphi(x_n), \theta_D\rangle_H, \|\theta_D\|_H^2 + \|\theta_D^\perp\|_H^2)$$

$$\geq \Psi(\langle \varphi(x_1), \theta_D\rangle_H, \dots, \langle \varphi(x_n), \theta_D\rangle_H, \|\theta_D\|^2)$$

d'où : $\inf_H \Psi(\dots) \geq \inf_{H_D} \Psi(\dots)$ $\square$

Donc, pour résoudre le problème, il suffit de rechercher la solution dans un espace de dimension finie.

<u>Reformulation du problème :</u>

$\theta = \alpha_1 \varphi(x_1) + \dots + \alpha_n \varphi(x_n)$ (changement de variables).

Alors

$$\min_{\theta \in H} \frac{1}{n} \sum_{i=1}^n \ell(y_i, \langle \varphi(x_i), \theta\rangle) + \frac{\lambda}{2} \|\theta\|^2$$

$$= \min_{\alpha \in H} \frac{1}{n} \sum_{i=1}^n \ell(y_i, \langle \varphi(x_i), \sum_{j=1}^n \alpha_j \varphi(x_j)\rangle) + \frac{\lambda}{2} \sum_j \alpha_j^2$$

$$= \min_{\alpha} \frac{1}{n} \sum_{i=1}^n \ell(y_i, \sum_{j=1}^n \alpha_j \langle \varphi(x_i), \varphi(x_j)\rangle) + \frac{\lambda}{2} \sum_i \alpha_i^2$$

Donc : Pour résoudre le problème, il suffit de pouvoir calculer

$$\left(\langle \varphi(x_i), \varphi(x_j)\rangle\right)_{i,j}$$

II. Noyaux (Définis) Positifs (symétriques)

Définition : Soit $K : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$. On dit que K est symétrique ~~défini~~ positif si K est symétrique et

- $\forall n \geq 1$, $\forall x_1, \dots, x_n \in \mathcal{X}$, $(K(x_i, x_j))_{i,j}$ est défini positive
- (de manière équivalente)
  $\forall n \geq 1$, $\forall x_1, \dots, x_n \in \mathcal{X}$, $\forall \lambda_1, \dots, \lambda_n \in \mathbb{R}$,

$$\sum_{i,j} \lambda_i \lambda_j K(x_i, x_j) \geq 0 .$$

Proposition : Si H est un espace de Hilbert et $\varphi : \mathcal{X} \to H$,
Alors $K(x,y) = \langle \varphi(x), \varphi(y)\rangle_H$ est symétrique défini positif.

preuve : Soit $x_1, \dots, x_n \in \mathcal{X}$, $\lambda_1, \dots, \lambda_n \in \mathbb{R}$.

$$\sum_{i,j} \lambda_i \lambda_j K(x_i, x_j) = \sum_{i,j} \lambda_i \lambda_j \langle \varphi(x_i), \varphi(x_j)\rangle_H$$

$$= \left\langle \sum_i \lambda_i \varphi(x_i), \sum_j \lambda_j \varphi(x_j)\right\rangle_H$$

$$= \left\| \sum_i \lambda_i \varphi(x_i)\right\|_H^2 \geq 0 . \qquad \square$$

Exemple: Le noyau linéaire $K(x,y) = x^T y$ est P.S.D sur $\mathbb{R}^d$

Théorème: $\Phi(x-y)$ est PSD si Φ est la transformée de Fourier d'une mesure de Borel positive

↳ $K(x,y) = \exp(-\alpha\|x-y\|_2^2)$ est P.S.D.

Proposition (Opérations sur les noyaux P.S.D)

Soient $(K_i)_{i \in \mathbb{N}}$ des noyaux P.S.D. sur $\mathcal{X}$. Alors

- $\forall I \subset \mathbb{N}$ finie,

  $\sum_{i \in I} \lambda_i K_i$ est P.S.D pour tous $(\lambda_i)_{i \in I} \geq 0$

- $\forall I \subset \mathbb{N}$ finie,

  $\prod_{i \in I} K_i$ est P.S.D

- La limite ponctuelle $K = \lim_{n \to +\infty} K_n$ est P.S.D

(si elle existe).

preuve:

* La somme et la limite sont laissées en exercice

* Pour le produit, montrons-le pour deux.

  Notons $K_1 = \left(K_1(x_i, x_j)\right)_{i,j}$

  $K_2 = \left(K_2(x_i, x_j)\right)_{i,j}$

$K_1 \succcurlyeq 0$ donc $\exists U_1$ tq $K_1 = U_1^T U_1$

de même, $\exists U_2$ tq $K_2 = U_2^T U_2$.

Notons $K^{(i)}$ la i-ième vecteur colonne de K. Alors

$$K_1(x_i, x_j) K_2(x_i, x_j)$$
$$= U_1^{(i)T} U_1^{(j)} U_2^{(j)T} U_2^{(i)}$$
$$= \mathrm{tr}\left(U_1^{(i)T} U_1^{(j)} U_2^{(j)T} U_2^{(i)}\right)$$
$$= \mathrm{tr}\left(U_2^{(i)} U_1^{(i)T} U_1^{(j)} U_2^{(j)T}\right)$$
$$= \mathrm{tr}\left(\left(U_1^{(i)} U_2^{(i)T}\right)^T U_1^{(j)} U_2^{(j)}\right)$$
$$= \langle \phi(i), \phi(j) \rangle_{Fr} \quad \text{où } \Phi(i) = U_1^{(i)} U_2^{(i)T}$$

□

Exercice : Si K est P.S.D, Alors e^K est P.S.D.

III. Théorème de Aronszajn

Théorème : Si $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ est P.S.D.
Alors $\exists$ H un espace de Hilbert et $\varphi : \mathcal{X} \to H$ tq
$\forall x, y \in \mathcal{X}$, $K(x,y) = \langle \varphi(x), \varphi(y) \rangle_H$.

(6) (Preuve de Jean-Philippe Vert)

1) Construction d'un pré-Hilbertien

Considérons $H_0 \overset{\text{déf}}{=} \{ y \mapsto K(x,y) \equiv K_x(y) \, ; \, x \in \mathcal{X} \}$.

muni de $\left\langle \sum_{i=1}^{m} a_i K_{x_i} , \sum_{j=1}^{n} b_j K_{y_j} \right\rangle_{H_0}$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j K(x_i, y_j).$$

Alors H_0 est <u>symétrique</u> et <u>bilinéaire</u> sur H_0 et <u>positive</u> (car K PSD).

Il reste à montrer que $\langle f, f \rangle_{H_0} = 0 \Rightarrow f = 0$.

Remarquons que $\boxed{\forall f \in H_0, \ \forall x \in \mathcal{X}, \ f(x) = \langle f, K_x \rangle_{H_0}}$ (1)

donc, d'après Cauchy-Schwarz (ne nécessite pas le critère défini),

$$\boxed{\forall f \in H_0, \ \forall x \in \mathcal{X}, \ |f(x)| \leq \sqrt{\langle f, f \rangle_{H_0}} \sqrt{K(x,x)}} \quad (2)$$

et $\|f\|_{H_0} = 0 \Rightarrow \left(f(x) = 0 \ \forall x \in \mathcal{X} \right)$

i.e $f = 0$

Donc $\left(H_0 , \langle \cdot , \cdot \rangle_{H_0} \right)$ est pré-Hilbertien.

2) Complétisation → Complétion

Comme souvent, l'idée est de compléter H_0 en un espace de Hilbert en considérant les suites de Cauchy.

Soit (f_n) une suite de Cauchy dans H_0. Alors (2) implique que $\forall x \in \mathcal{X}$, $(f_n(x))$ est une suite de Cauchy dans $\mathbb{R}$ qui admet donc une limite finie.

Considérons $H = \{$ limites ponctuelles de suites de Cauchy de $H_0 \}$.

Montrons que nous pouvons équiper H d'une structure de Hilbert qui généralise $(H_0, \langle \cdot, \cdot \rangle_{H_0})$.

- Définition du produit scalaire.

Soit (f_n), (g_n) deux suites de Cauchy dans H_0.

Nous allons montrer que $\left(\langle f_n, g_n \rangle_{H_0}\right)$ converge en montrant que c'est une suite de Cauchy dans $\mathbb{R}$.

(8) $|\langle f_n, g_n\rangle_{H_0} - \langle f_m, g_m\rangle_{H_0}|$

$$= |\langle f_n, g_n - g_m\rangle_{H_0} + \langle f_n - f_m, g_m\rangle_{H_0}|$$

$$\underset{C.S.}{\leq} \underbrace{\|f_n\|_{H_0}}_{\text{borné}} \underbrace{\|g_n - g_m\|_{H_0}}_{\substack{\text{aussi petit} \\ \text{qu'on veut} \\ (m,n)\to+\infty}} + \underbrace{\|f_n - f_m\|_{H_0}}_{\text{Idem}} \underbrace{\|g_m\|_{H_0}}_{\text{borné}}$$

d'où le résultat.

Pour définir le produit scalaire comme la limite, il reste à vérifier que cette dernière est indep des ~~la suite~~ suites de cauchy choisies.

<u>Lemme</u>: Si (f_n) est une suite de cauchy ds H_0 qui converge ponctuellement vers 0, alors $\|f_n\|_{H_0}$ tend vers 0.

<u>preuve</u>: $\|f_n\|^2_{H_0} = \langle f_n - f_N, f_n\rangle_{H_0} + \langle f_N, f_n\rangle_{H_0}$

$$\leq \underbrace{\|f_n - f_N\|_{H_0}}_{\substack{\leq \frac{\varepsilon}{B} \text{ pour} \\ \min(n,N)\geq N_0}} \underbrace{\|f_n\|_{H_0}}_{\text{borné (par } B)} + \langle f_N, f_n\rangle_{H_0}$$

$$f_{N_0} = \sum_{i=1}^{p} \alpha_i K(x_i, x)$$

$$\leq \varepsilon + \underbrace{\sum_{i=1}^{p} \alpha_i f_n(x_i)}_{\underset{n\to+\infty}{\longrightarrow} 0}$$

$\square$

(9) Soient $(f_n), (f_n')$ deux suites de Cauchy qui convergent ponctuellement vers f
" $(g_n), (g_n')$ " g.

Alors,

$$|\langle f_n, g_n\rangle_{H_0} - \langle f_n', g_n'\rangle_{H_0}| = |\langle f_n, g_n - g_n'\rangle_{H_0} + \langle f_n - f_n', g_n'\rangle_{H_0}|$$

$$\leq \underbrace{\|f_n\|_{H_0}}_{\text{borné}} \underbrace{\|g_n - g_n'\|_{H_0}}_{\to 0 \text{ d'après lemme}} + \underbrace{\|g_n'\|_{H_0}}_{\text{borné}} \underbrace{\|f_n - f_n'\|_{H_0}}_{\to 0 \text{ d'après lemme}}$$

$$\xrightarrow[n \to +\infty]{} 0$$

Donc la limite ne dépend pas du représentant.

Donc, $(H, \langle\, ,\, \rangle)$ est bien défini. Il reste à prouver qu'il s'agit d'un espace de Hilbert.

- $\langle \cdot, \cdot \rangle$ est symétrique bilinéaire et positive (évident).

- Soit f tq $\|f\|_H = 0 = \lim \|f_n\|_{H_0}$.

Alors $(1) \Rightarrow f(x) = 0 \ \forall x \in X$
(i.e. $f = 0$).

~~Il reste à prouver la complétude.~~

Soit $f \in H$, $\exists (f_n)$ suite de Cauchy de H_0 qui converge simplement vers f ~~dès~~.

Alors $(f_p - f_n)_p$ est une suite de Cauchy de H_0 qui CV simplement vers $f - f_n$. Donc

$$\lim_{n \to +\infty} \|f - f_n\|_H = \lim_{n \to +\infty} \lim_{p \to +\infty} \|f_p - f_n\|_{H_0}$$

$$= 0 \quad (\text{car } (f_n) \text{ de Cauchy ds } H_0).$$

d'où : $\boxed{H_0 \text{ est dense ds } H}$

• Prouvons la complétude :

Soit (f_n) une suite de Cauchy de H.

Par complétude, $\exists (f_n')$ suite de H_0 tq $\lim_n \|f_n - f_n'\|_H = 0$.

Mq (f_n') est de Cauchy ds H_0

$$\|f_n' - f_m'\|_{H_0} = \|f_n' - f_m'\|_H$$

$$= \|f_n' - f_n\|_H + \|f_n - f_m\|_H + \|f_m - f_m'\|_H$$

$$\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}$$

par (m, n) assez grand.

et donc (f_n') converge ponctuellement vers f.

Nous avons $\lim_{n \to +\infty} \|f - f_n'\|_H = \lim_{n \to +\infty} \lim_{p \to +\infty} \|f_p' - f_n'\|_{H_0}$

$= 0$.

donc $\lim_{n \to +\infty} \|f - f_n\|_H \leq \lim_{n \to +\infty} \|f - f_n'\|_H + \lim_{n \to +\infty} \|f_n' - f_n\|_H$

$= 0$.

3) Propriétés Caractéristiques

- $\forall x,\ K(x, \cdot) \in H$
- $\forall x, \forall f \in H,\ f(x) = \langle f, K_x \rangle_H$

car si (f_n) est une suite de Cauchy de H_0 qui tend simplement vers f,

$$f(x) = \lim_{n \to +\infty} f_n(x)$$

$$= \lim_{n \to +\infty} \langle f_n, K_x \rangle_{H_0}$$

$$= \langle f, K_x \rangle_H.$$

- $\forall x, y,\ K(x, y) = \langle K_x, K_y \rangle_H$.