Méthodes à Noyaux

Lorsque on est capable de mesurer la similarité des données, les noyaux sont une manière de généraliser des algorithmes linéaires pour delp. des méthodes non-linéaires.

I. Similarités

Si (x, y) est dans le jeu de données et on nous montre un autre x' tel que x' "ressemble beaucoup" à x, on aimerait prédire un y' qui "ressemble beaucoup" à y. Les noyaux permettent de faire ça tout en conservant implicitement une structure linéaire.

Notations : $\mathcal{X}$: ensemble dans lequel vivent les xs

Similarité : Soit une fonction $s : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, on dit que s est une fonction de similarité si s respecte l'heuristique suivante : plus $s(x, y)$ est grande, plus x et y se ressemblent

↳ dépend de la tâche.

ex : • Si $\mathcal{X} \subset \mathbb{R}^d$: $s(x, y) = -\|x - y\|$
• Si $\mathcal{X} \subset \mathbb{R}^d$: $s(x, y) = x^T y$

II Noyaux Symmétriques Positifs

__Définition__ : Soit $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, on dit que k est un noyau symmétique positif si :

- $\forall x_1, x_2 \in \mathcal{X}$, $k(x_1, x_2) = k(x_2, x_1)$ (symmétrie)
- $\forall n \geq 1, \forall x_1, \dots, x_n \in \mathcal{X}, \forall \alpha_1, \dots, \alpha_n \in \mathbb{R}$,

$$\sum_{i,j} \alpha_i \alpha_j k(x_i, x_j) \geq 0 \quad \text{(positivité)}$$

__Remarque__ : La contrainte de positivité est équivalente à :

$$\forall n \geq 1, \forall x_1, \dots, x_n \in \mathcal{X}, \quad (k(x_i, x_j))_{i,j} \in S_n^+(\mathbb{R})$$

__Théorème__ : (Aronszajn)

Si $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ est un noyau symmétrique positif, alors il existe un __espace de Hilbert__ H_k et une application $\phi_k : \mathcal{X} \to H_k$ tq

$$\forall x, y \in \mathcal{X}, \quad k(x, y) = \langle \phi_k(x), \phi_k(y) \rangle_{H_k}$$

__démonstration__ : Voir site web.

Parmis les fonctions de similarité, les noyaux symmétriques positifs sont très interessants car ils exhibent de manière __implicite__ une structure de Hilbert (donc __linéaire__).

Exemple : Si H est un espace de Hilbert et $\phi : \mathcal{X} \to H$ est une application, alors $x, y \mapsto \underbrace{\langle \phi(x), \phi(y) \rangle_H}_{:= k(x,y)}$ est un noyau symétrique positif.

démonstration : Soit $n \geq 1$, $x_1, \dots, x_n \in \mathcal{X}$, $\alpha_1, \dots, \alpha_n \in \mathbb{R}$,

$$\sum_{i,j} \alpha_i \alpha_j k(x_i, y_j) = \sum_{i,j} \alpha_i \alpha_j \langle \phi(x_i), \phi(x_j) \rangle_H$$

$$= \left\langle \sum_i \alpha_i \phi(x_i), \sum_j \alpha_j \phi(x_j) \right\rangle_H$$

$$= \left\| \sum_i \alpha_i \phi(x_i) \right\|_H^2$$

$$\geq 0 \qquad \square$$

Théorème (de représentation)

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

↳ exemple $\hat{\theta} \in \operatorname{argmin} \frac{1}{n} \sum_{i=1}^{n} \ell_{logistic}\left(y_i \langle \varphi(x_i), \theta \rangle_H\right) + \|\theta\|_H^2$

admet une solution $\hat{\theta} \in \operatorname{Vect}\{\varphi(x_1), \dots, \varphi(x_n)\}$ si Ψ est strictement croissante par rapport à sa dernière composante.

Remarque : Ce théorème permet de chercher des solutions dans un espace de dimension finie et permet d'implémenter les méthodes théoriques.

démonstration :

Notons $H_D = \mathrm{Vect}\{\varphi(x_1), \dots, \varphi(x_n)\}$.

Soit $\theta \in H$, $\theta = \theta_D + \theta_D^\perp$ avec $\theta_D \in H_D$ et $\theta_D^\perp \in H_D^\perp$.

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

$$= \Psi(\quad '' \quad , \|\theta_D\|^2 + \|\theta_D^\perp\|^2)$$

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

$$\geq \Psi(\quad '' \quad , \|\theta_D\|^2)$$

donc $\inf_H \Psi(\dots) \geq \inf_{H_D} \Psi(\dots)$ □

III. Construction de noyaux symétriques positifs.

Nous avons déjà vu que les noyaux symétriques positifs sont exactement les produits-scalaires dans des espaces de Hilbert. Nous allons voir dans cette section des règles de construction.

Théorème : L'ensemble des noyaux symétriques positifs est stable par :

- Multiplication par des réels positifs
- Additions
- Multiplications
- Passage à la limite simple si bien définie.

<u>démonstration:</u> Nous allons regarder le cas du produit, les autres cas étant immédiats.

Soient k_1, k_2 deux noyaux symétrique positifs.

Soit $n \geq 1$, $x_1, \dots, x_n \in \mathcal{X}$. Notons $K_1 = (k_1(x_i, x_j))_{i,j}$

$K_2 = (k_2(x_i, x_j))_{i,j}$

K_1 (resp. K_2) $\in S_n^+(\mathbb{R})$, donc il existe U_1 (resp. U_2) tq

$K_1 = U_1^T U_1$ (resp $U_2^T U_2$).

$$k_1(x_i, x_j) k_2(x_i, x_j) = (\underbrace{U_1^{(i)T}}_{\text{i-ème vecteur colone.}} U_1^{(j)})(U_2^{(i)T} U_2^{(j)})$$

$$= (U_1^{(i)T} U_1^{(j)}) \, tr(U_2^{(i)T} U_2^{(j)})$$

$$= (U_1^{(i)T} U_1^{(j)}) \, tr(U_2^{(j)T} U_2^{(i)})$$

$$= tr((U_1^{(i)T} U_1^{(j)}) U_2^{(j)T} U_2^{(i)})$$

$$= tr(U_2^{(i)} (U_1^{(i)T} U_1^{(j)}) U_2^{(j)T})$$

$$= tr((U_2^{(i)} U_1^{(i)T}) U_1^{(j)} U_2^{(j)T})$$

$$= tr((U_1^{(i)} U_2^{(i)T})^T U_1^{(j)} U_2^{(j)T})$$

$$= \langle U_1^{(i)} U_2^{(i)T}, U_1^{(j)} U_2^{(j)T} \rangle_{Fr}$$

donc $(k_1(x_i, x_j) k_2(x_i, x_j))_{i,j}$ est une matrice de Gram donc $\in S_n^+(\mathbb{R})$.

Exercice :

- Montrer que $\forall q$, $x, y \mapsto (x^T y)^q$ est symmétrique positif (noyau polynomial)
- Montrer que si k est symmétrique positif, alors $\exp(k)$ l'est également.

Théorème (de Bochner) :

$k : x, y \mapsto q(x-y)$ est symmétrique positif ssi q est la transformée de Fourier inverse d'une mesure Borélienne positive.

démonstration ($\Leftarrow$) :

$$k(x,y) = q(x-y) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i(x-y)^T\omega} \underbrace{d\mu(\omega)}_{\text{positive.}}$$

Soit $n \geq 1$, $x_1, \dots, x_n \in \mathcal{X}$, $\alpha_1, \dots, \alpha_n \in \mathbb{R}$,

$$\sum_{i,j} \alpha_i \alpha_j k(x_i, x_j) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\sum_{i,j} \alpha_i \alpha_j e^{i(x_i - x_j)^T\omega} \right) d\mu(\omega)$$

$$= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\sum_{i,j} \left(\alpha_i e^{i x_i^T \omega}\right) \overline{\left(\alpha_j e^{i x_j^T \omega}\right)} \right) d\mu(\omega)$$

$$= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \underbrace{\langle \sum_i \alpha_i e^{i x_i^T\omega}, \sum_j \alpha_j e^{i x_j^T \omega} \rangle_{\mathbb{C}}}_{= \underbrace{\left| \sum_i \alpha_i e^{i x_i^T \omega} \right|^2}_{\geq 0}} \underbrace{d\mu(\omega)}_{\geq 0}$$

≥ 0. $\square$

IV SVM à Noyaux

SVM classique $\hat{\theta} \in \underset{\theta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left(1 - y_i \theta^T \varphi(x_i)\right)_+ + \frac{\lambda}{2}\|\theta\|^2$

SVM à noyaux :

Soit k un noyau symétrique positif, H l'espace de Hilbert donné par le théorème d'Aronszajn, et ϕ la fonction associée.

on cherche $\hat{\theta} \in \underset{\theta \in H}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left(1 - y_i \langle \theta, \phi(x_i) \rangle\right)_+ + \frac{\lambda}{2}\|\theta\|^2$ (SPM)

⚠ Ce problème est maintenant un problème de dimension infinie, on ne peut plus le résoudre par SGD.

Avec le théorème de représentation:

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

donc (SPM) est équivalent à

$$\hat{\alpha} \in \underset{\alpha \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left(1 - y_i \left\langle \sum_j \alpha_j \phi(x_j), \phi(x_i) \right\rangle\right)_+ + \frac{\lambda}{2} \left\| \sum_j \alpha_j \phi(x_j) \right\|^2$$

$$= \underset{\alpha \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left(1 - y_i \sum_{j=1}^{n} \alpha_j \langle \phi(x_j), \phi(x_i) \rangle\right)_+ + \frac{\lambda}{2} \sum_i \sum_j \alpha_i \alpha_j \langle \phi(x_i, x_j) \rangle$$

⑧

et en notant $K = (k(x_i, x_j))_{i,j}$, ce problème est équivalent à

$$\hat{\alpha} \in \underset{\alpha}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left(1 - y_i(\alpha^T K^{(i)})\right)_+ + \frac{\lambda}{2} \alpha^T K \alpha$$

Résolution de ce problème en TP.