UNIVERSITY OF INFORMATION TECHNOLOGY – VIETNAM NATIONAL UNIVERSITY HCMC

MATHEMATICS FOR COMPUTER SCIENCE

KERNEL METHODS

- LÊ ĐOÀN PHÚC MINH
- LÊ THẾ VIỆT
- NGUYỄN DUY ĐẠT
- TRẦN NGỌC THÀNH

- <u>Giới thiệu về các giải pháp Kernel</u>
- <u>Dẫn chứng định lý</u>
- <u>Về Kernel</u>:
  - Kernel thường gặp
  - Kernel tịnh tiến bất biến trên $[0, 1]$
  - Kernel tịnh tiến bất biến trên R^d
  - Một số ứng dụng

GIỚI THIỆU VỀ CÁC GIẢI PHÁP KERNEL

Kernels hay các giải pháp kernel là một tập hợp các thuật toán khác nhau được sử dụng nhằm mục đích phân tích mẫu dữ liệu. Chúng được sử dụng để xử lý những bài toán phi tuyến tính bằng cách sử dụng công cụ phân loại tuyến tính.

- **Ý TƯỞNG CƠ BẢN CỦA KERNEL:**

Với dữ liệu ban đầu ở trạng thái không phân biệt tuyến tính, ta sẽ tìm một phép biến đổi sao cho dữ liệu khi được biến đổi qua không gian mới này trở nên phân biệt tuyến tính.

- **TẠI SAO KERNEL QUAN TRỌNG:**
  - Để xử lý được những mô hình tuyến tính trong không gian đa chiều cực lớn, những công cụ cần thiết cho việc phân tích không gian có số chiều vô hạn là điều cần phải có.
  - Các giải pháp Kernel cung cấp cho chúng ta các thuật toán đơn giản và ổn định với việc đảm bảo tính lý thuyết và tính tương thích của hàm số được hướng đến.
  - Có thể dễ dàng áp dụng vào những đối tượng đầu vào không phải là Vector
  - Hữu dụng trong việc xử lý các mô hình khác như Neural Networks

DẪN CHỨNG ĐỊNH LÝ

Xét bài toán cần tối ưu hóa từ máy học với mô hình tuyến tính và dữ liệu trong mô hình có tọa độ $(x, y) \in \mathrm{X} \times \mathrm{Y}, \mathrm{i} = 1, \ldots, \mathrm{n}$ cùng với công thức sau:

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

Trong đó:

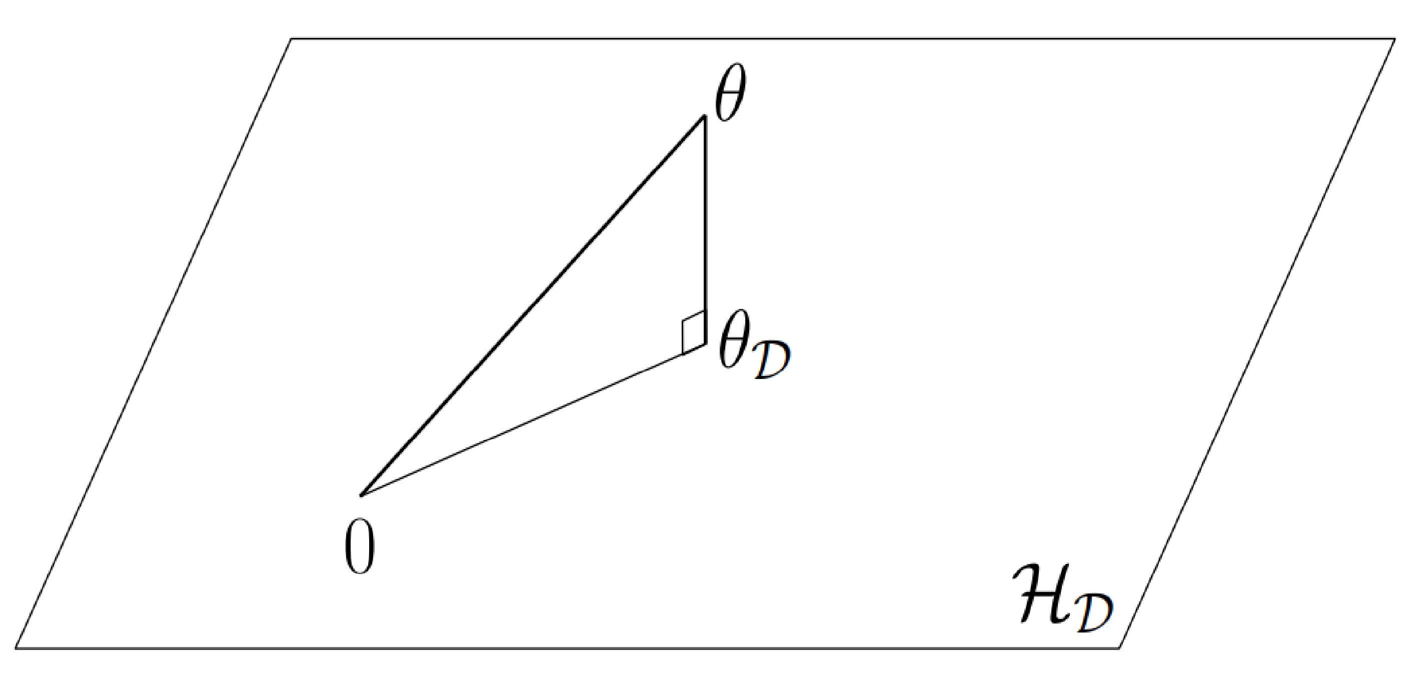
- H là không gian Hilbert
- $\langle \varphi(x_i), \theta \rangle$ là tích vô hướng của $\varphi(x_i)$ và θ.
- ℓ là hàm mất mát
- $||\theta||$ là norm Hilbert

<u>Định lý:</u> Gọi $\varphi\colon\ X \to H$, $(\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{n}) \in \mathrm{X}^\mathrm{n}$. Biết rằng với hàm số $\psi(psi)\colon\ \mathrm{R}^{\mathrm{n}+1} \to \mathrm{R}$ tịnh tiến lần lượt cho đến giá trị cuối cùng. Khi đó ta có thể lấy được cận dưới của $\psi(\langle\theta, \varphi(x_1)\rangle, \dots, \langle\theta, \varphi(x_\mathrm{n})\rangle, ||\theta||^2)$ bằng cách giới hạn đến vector θ dạng:

$$\theta = \sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_\mathrm{i} \varphi(\mathrm{x}_\mathrm{i}) \ \ (\alpha \in \mathrm{R}^\mathrm{n})$$

Chứng minh định lý: Giả sử $\theta \in \mathrm{H}$ và $\mathrm{H_D} = \{\sum_{i=1}^{n} \alpha_i \varphi(x_i), \alpha \in \mathrm{R^n}\} \subset \mathrm{H}$ là không gian con tuyến tính của các Vector. Đồng thời, giả sử $\theta \in \mathrm{H_D}$ và $\theta_\perp \in \mathrm{H_D^\perp}$ mà $\theta = \theta_D + \theta_\perp$ là một khai triển của không gian Hilbert.

Khi đó, $\forall i \in \{1, \dots, n\}, \langle \theta, \varphi(x_i) \rangle = \langle \theta_D, \varphi(x_i) \rangle + \langle \theta_\perp, \varphi(x_i) \rangle$ với $\langle \theta_\perp, \varphi(x_i) \rangle = 0$.



Áp dụng định lý Pythagoreas, ta có công thức sau: $\|\theta\|^2 = \|\theta_D\|^2 + \|\theta_\perp\|^2$.

Từ đó ta có:

$$\begin{aligned}\psi(\langle\theta,\varphi(x_1)\rangle,\dots,\langle\theta,\varphi(x_n)\rangle,\|\theta\|^2) &= \psi(\langle\theta_D,\varphi(x_1)\rangle,\dots,\langle\theta_D,\varphi(x_n)\rangle,\|\theta_D\|^2+\|\theta_\perp\|^2)\\ &\geq \psi(\langle\theta_D,\varphi(x_1)\rangle,\dots,\langle\theta_D,\varphi(x_n)\rangle,\|\theta_D\|^2).\end{aligned}$$

Do đó:

$$\inf_{\theta\in H}\psi(\langle\theta,\varphi(x_1)\rangle,\dots,\langle\theta,\varphi(x_n)\rangle,\|\theta\|^2) = \inf_{\theta\in H_D}\psi(\langle\theta,\varphi(x_1)\rangle,\dots,\langle\theta,\varphi(x_n)\rangle,\|\theta\|^2),$$

chính là kết quả cần tìm.

Hệ quả: Với $\lambda > 0$ thì:

$$\inf_{\theta\in H}\frac{1}{n}\sum \ell(y_i,\langle\theta,\varphi(x_i)\rangle) + \frac{\lambda}{2}\|\theta\|^2 = \inf_{\alpha\in R^n}\frac{1}{n}\sum \ell(y_i,\langle\theta,\varphi(x_i)\rangle + \frac{\lambda}{2}\|\theta\|^2$$

$$\Rightarrow \theta = \sum_{i=1}^{n}\alpha_i\varphi(x_i).$$

Lưu ý là không có giả thiết cho hàm mất mát ℓ

Với hệ quả trên, ta có thể viết lại công thức cho bài toán máy học với hàm số Kernel k và tích vô hướng là:

$$k(x,x') = \langle\varphi(x),\varphi(x')\rangle$$

Ta có: $\forall j \in \{1, \dots, n\}$, $\langle \theta, \varphi(x_j) \rangle = \sum_{i=1}^{n} a_i k(x_i, x_j) = (K\alpha)_j$

Khi $K \in R^{n \times n}$ là ma trận Kernel => $K_{ij} = \langle \varphi(x_i), \varphi(x_j) \rangle = k(x_i, x_j)$ và

$$\|\theta\|^2 = \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} \langle \varphi(x_i), \varphi(x_j) \rangle = \sum_{\mathrm{i}=1}^{\mathrm{n}} \sum_{\mathrm{j}=1}^{\mathrm{n}} \alpha_{\mathrm{i}} \alpha_{\mathrm{j}} \mathrm{K}_{\mathrm{ij}} = \alpha^{\mathrm{T}} \mathrm{K} \alpha.$$

Ta có thể viết:

$$\inf_{\theta \in H} \frac{1}{n} \sum \ell(y_i, \langle \theta, \varphi(x_i) \rangle) + \frac{\lambda}{2} \|\theta\|^2 = \inf_{\alpha \in R^n} \frac{1}{n} \sum \ell(y_i, (\mathrm{K}\alpha)_{\mathrm{i}}) + \frac{\lambda}{2} \alpha^{\mathrm{T}} \mathrm{K} \alpha$$

Với một điểm thử nghiệm $x \in X$ nhất định, chúng ta có: $f(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i)$

Đây được gọi là Kernel Trick

Lợi ích của Kernel Trick:

- Thay thế không gian Hilbert thành R^n
- Phân tách bài toán, các thuật toán và phân tích → Nhiều loại Kernel có thể được sử dụng với nhiều kiểu dữ liệu khác nhau

VỀ KERNEL

- **KERNEL THƯỜNG GẶP**

Kernel tuyến tính:

Công thức:

$$k(x, x') = x^T x'$$

(Tương ứng với công thức $f_\theta(x) = \theta^T x$ và hàm "phạt" L2 $||\theta||_2^2$)

• KERNEL THƯỜNG GẶP

<u>Kernel đa thức:</u> Với r là một số nguyên dương, chúng ta có thể mở rộng Kernel (có định lý nhị thức) $k(x, x') = x^T x'$ thành:

$$k(x, x') = \left(\sum_{i=1}^{d} x_i x_i'\right)^r = \sum_{\alpha_1+\cdots+\alpha_d=r} \binom{r}{\alpha_1, \dots, \alpha_d} \underbrace{(x_1 x_1')^{\alpha_1} \dots (x_d x_d')^{\alpha_d}}_{(\mathrm{x}_1^{\alpha_1} \dots \mathrm{x}_\mathrm{d}^{\alpha_\mathrm{d}})\left((\mathrm{x}_1')^{\alpha_1} \dots (\mathrm{x}_\mathrm{d}')^{\alpha_\mathrm{d}}\right)}$$

Trong đó:

- Tổng ở đây là tổng các Vector nguyên không âm $(\alpha_1, \dots, \alpha_\mathrm{d})$
- Tập hợp các hàm của đa thức thuần nhất trên R^d có không gian $\binom{d+r-1}{r}$

- **KERNEL TỊNH TIẾN BẤT BIẾN TRÊN [0, 1]**

Xét $X = [0, 1]$ và Kernel dưới dạng $k(x, x') = q(x - x')$ với hàm $q: [0,1] \to R$ (Tức tuần hoàn đơn). Có thể thấy, ta có thể mở rộng bình phương hàm có thể khả tích chu kỳ đơn trong chuỗi Fourier.

$\Rightarrow q(x) = \sum_{m \in Z} e^{2im\pi x} \hat{q}_m$ với $\hat{q}_m = \int_0^1 q(x) e^{-2im\pi x}\, dx\ (m \in Z)$.

- **KERNEL TỊNH TIẾN BẤT BIẾN TRÊN [0, 1]**

Cho rằng một hàm tuần hoàn đơn f đã được khai triển thành chuỗi Fourier $f(x) = \sum_{m\in Z} e^{2im\pi x} \hat{f}_m$, ta có hiệu chuẩn là $\sum_{m\in Z} \left|\hat{f}_m\right|^2$ với $c \in R_+^Z$. Hàm này tương đương với Vector đặc trưng $\varphi(x)_m = \frac{e^{2im\pi x}}{\sqrt{c_m}}$ và $\theta \in C^Z$.

Mà $\theta_m = \hat{f}_m\sqrt{c_m}$ nên suy ra $f(x) = \langle\theta, \varphi(x)\rangle$ và $\sum_{m\in Z}|\theta_m|^2$ tương đương với norm của $\sum_{m\in Z} c_m \left|\hat{f}_m\right|^2$.

- **KERNEL TỊNH TIẾN BẤT BIẾN TRÊN [0, 1]**

Do đó, Kernel liên kết là:

$$k(x,x') = \sum_{m\in Z} \varphi(x)_m \varphi(x)_m^* = \sum_{m\in Z} \frac{e^{2im\pi x}}{\sqrt{c_m}} \frac{e^{-2im\pi x'}}{\sqrt{c_m}} = \sum_{m\in Z} \frac{1}{c_m} e^{2im\pi(x-y)} = q(x-x')$$

Bất cứ mọi hiệu chuẩn dưới dạng $\sum_{m\in Z} c_m \left|\hat{f}_m\right|^2$ xác định được norm của RKHS* khi c_m dương và $\sum_{m\in Z} \frac{1}{c_m}$ là hữu hạn.

*RKHS: Reproducing Kernel Hilbert space (Kernel không gian Hilbert tái lập)

- **KERNEL TỊNH TIẾN BẤT BIẾN TRÊN [0, 1]**

Hàm "phạt" (Penalization) của các đạo hàm: Với từng "phạt" dựa trên c, có một liên kết tự nhiên giữa các "phạt" trên đạo hàm, nếu f có số lần làm mượt (smoothness) khác nhau với bình phương tích phân đạo hàm, ta có:

$$f^{(s)}(x) = \sum_{m \in Z} (2im\pi)^s \, e^{2im\pi x} \hat{f}_m$$

Và từ định lý Parseval, ta có:

$$\int_0^1 \left|f^{(s)}(x)\right|^2 dx = (2\pi)^{2s} \sum_{m \in Z} m^{2s} \left|\hat{f}_m\right|^2$$

- **KERNEL TỊNH TIẾN BẤT BIẾN TRÊN [0, 1]**

Đa thức Bernoulli: Xét $c_0 = (2\pi)^{-2s}$ và $c_m = |m|^{2s}$ $(m \neq 0)$ mà norm liên kết là $\|f\|_H^2 = \frac{1}{(2\pi)^{2s}}\left(\int_0^1 \left|f^{(s)}(x)\right|^2 dx + \int_0^1 |f(x)|^2 dx\right)$. Kernel hiện tại là $k(x,x')$ có thể được viết thành:

$$k(x,x') = \sum_{m\in Z} c_m^{-1} e^{2im\pi(x-x')} = (2\pi)^{2s} + \sum_{m\geq 1} \frac{2cos[2\pi m(x-x')]}{m^{2s}}$$

Với $s = 1$, ta có $k(x,x) = (2\pi)^2 + 2\sum_{m\geq 1} m^{-2} = (2\pi)^2 + \frac{\pi^2}{3}$. Ngoài ra, bằng cách sử dụng chuỗi Fourier khai triển $\{t\} = \frac{1}{2} - \frac{1}{2\pi}\sum_{m\geq 1}\frac{2sin[2\pi mt]}{m}$, và lấy tích phân nó, ta được:

$$k(x,x') = 2\pi^2\{x-x'\}^2 - 2\pi^2\{x-x'\} + \frac{\pi^2}{3} + (2\pi)^2$$

- **KERNEL TỊNH TIẾN BẤT BIẾN TRÊN [0, 1]**

Với $s \geq 1$, ta có công thức dạng đóng $k(x, x') = (2\pi)^{2s} + (-1)^{s-1}\frac{(2\pi)^{2s}}{(2s)!}B_{2s}(\{x - x'\})$, trong đó B_{2s} là đa thức Bernoulli thứ 2s ($B_2(t) = t^2 - t + \frac{1}{6}$).

Kernel hàm số mũ chu kỳ: Xét $c_m = 1 + a^2|m|^2$, từ đó ta có một công thức dạng đóng với hiệu chuẩn $||\mathrm{f}||_{\mathrm{H}}^2 = \frac{\alpha^2}{(2\pi)^2}\int_0^1 \left|f^{(s)}(x)\right|^2 dx + \int_0^1 |f(x)|^2 dx$

- **KERNEL TỊNH TIẾN BẤT BIẾN TRÊN R^d**

Xét $x = R^d$ và một Kernel dưới dạng $k(x, x') = q(x - x')$ với hàm $q\colon\ R^d \to R$.

<u>Định lý:</u> Kernel k được xác định là dương khi và chỉ khi q là một phép đo Borel không âm của phép biến đổi Fourier.

<u>Hệ quả:</u> Nếu $q \in L^1(dx)$ và phép biến đổi Fourier của nó chỉ có giá trị thực không âm, khi đó k được xác định là dương.

- **KERNEL TỊNH TIẾN BẤT BIẾN TRÊN R^d**

Quá trình xây dựng norm liên kết: Ta cho một biện luận trực giác (không chặt chẽ):
Nếu q nằm trong $L^1(dx)$, khi đó $\hat{q}(\omega)$ tồn tại và ta có một biểu diễn chi tiết với:

$$k(x,x') = \frac{1}{(2\pi)^2}\int_{R^d}\left\langle\sqrt{\hat{q}(\omega)}e^{i\omega^T x}, \sqrt{\hat{q}(\omega)}e^{i\omega^T x\prime}\right\rangle d\omega = \int_{R^d}\langle\varphi(x)_\omega, \varphi(x')_\omega\rangle d\omega$$

Với $\varphi(x)_\omega = \frac{1}{(2\pi)^{\frac{d}{2}}}\sqrt{\hat{q}(\omega)}e^{i\omega^T x}$.

Xét $f(x) = \int_{R^d}(\varphi(x)_\omega\theta_\omega)d\omega = \langle\varphi(x),\theta\rangle$, khi đó $\theta_\omega = \frac{1}{(2\pi)^{\frac{d}{2}}}\hat{f}(\omega)/\sqrt{\hat{q}(\omega)}$ và bình phương norm của θ tương đương với $\frac{1}{(2\pi)^d}\int_{R^d}\frac{\left|\hat{f}(w)\right|^2}{\hat{q}(\omega)}d\omega$.

Vậy norm của hàm $f \in H$ là: $\|f\|_H^2 = \frac{1}{(2\pi)^d}\int_{R^d}\frac{\left|\hat{f}(w)\right|^2}{\hat{q}(\omega)}d\omega$.

• KERNEL TỊNH TIẾN BẤT BIẾN TRÊN R^d

<u>Liên kết với đạo hàm:</u> Khi f có đạo hàm riêng, thì phép biến đổi Fourier của $\frac{\partial f}{\partial x_j}$ tương đương với $i\omega_j$ lần phép biến đổi Fourier của f. Điều này dẫn đến việc áp dụng định lý Parseval $\left(\frac{1}{(2\pi)^d}\int_{R^d}\left|\omega_j\right|^2\left|\hat{f}(w)\right|^2 d\omega = \int_{R^d}\left|\frac{\partial f}{\partial x_j}(x)\right|^2 dx\right)$ có thể kéo dài đến thứ tự đạo hàm cao hơn:

$$\frac{1}{(2\pi)^d}\int_{R^d}\left|\omega_1^{\alpha_1}\ldots\omega_d^{\alpha_d}\right|^2\left|\hat{f}(w)\right|^2 d\omega = \int_{R^d}\left|\frac{\partial^{\alpha} f}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}}(x)\right|^2 dx$$

- **KERNEL TỊNH TIẾN BẤT BIẾN TRÊN R^d**

Kernel hàm số mũ: Đây là kernel có dạng $q(x - x') = exp(-\alpha\|x - x'\|_2)$ với phần biến đổi Fourier có thể được tính là $\hat{q}(\omega) = 2^d \pi^{\frac{d-1}{2}} \gamma \frac{d+1}{2} \frac{\alpha}{(\alpha^2 + \|\omega\|_2^2)^{\frac{d+1}{2}}}$.

Do đó, $\hat{q}(\omega)^{-1}$ là tổng của các đơn thức và nếu ta để ý thứ tự của các đơn thức này, ta thấy rằng norm các RKHS tương ứng đang "phạt" các đạo hàm tới toàn bộ thứ tự $(d + 1)/2$, mà với mọi $\alpha \in N^d$ nên $\alpha_1 + \cdots + \alpha_{\mathrm{d}} = \frac{\mathrm{d}+1}{2}$ (Đây là không gian Sobolev).

- **KERNEL TỊNH TIẾN BẤT BIẾN TRÊN R^d**

Đặc biệt, với $d = 1$, ta có $\hat{q}(\omega) = \frac{2\alpha}{\alpha^2+\omega^2}$. Do đó:

$$\|f\|_H^2 = \frac{1}{2\pi}\int_R \frac{\left|\hat{f}(w)\right|^2}{\hat{q}(\omega)} d\omega = \frac{\alpha}{2}\frac{1}{2\pi}\int_R \left|\hat{f}(w)\right|^2 d\omega + \frac{1}{2\alpha}\frac{1}{2\pi}\int_R \left|\omega\hat{f}(w)\right|^2 d\omega$$

$$= \frac{\alpha}{2}\int_R |f(x)|^2 dx + \frac{1}{2\alpha}\int_R |f'(x)|^2 dx$$

Lúc này ta phục hồi không gian Sobolev của hàm với các đạo hàm bình phương khả tích được.

• KERNEL TỊNH TIẾN BẤT BIẾN TRÊN R^d

Kernel Gaussian RBF: Đây là kernel có dạng $q(x - x') = exp(-\alpha\|x - x'\|^2)$ với phần biến đổi Fourier có thể được tính là $\hat{q}(\omega) = \left(\frac{\pi}{\alpha}\right)^{\frac{d}{2}} \sum_{s=0}^{\infty}(-1)^s \frac{\|\omega\|_2^{2s}}{(4\alpha)^s s!}$, tương ứng với norm của RKHS mà đang "phạt" toàn bộ các đạo hàm.

Kernel Matern: Đây là Kernel có thể định nghĩa một chuỗi các Kernel để $\hat{q}(\omega) \propto^{(*)} \frac{1}{\left(\alpha^2+\|\omega\|_2^2\right)^s}$ với $s > \frac{d}{2}$ nhằm đảm bảo tính khả tích của phép biến đổi Fourier. Các Kernel Matern đều tương ứng với không gian Sobolev theo thứ tự s.

(*) ∝: Tỉ lệ với

HẾT PHẦN 1