- Kernels $\equiv$ Elegant and powerful technique to deal with linear models

- Linear models
  - examples of feature maps
  - kernel will allow to deal with infinite dimensional feature maps

- problems when linear models are used
  - interpolation / approximation
  - supervised learning
  - solution of PDE / optimal control
  - integration problems

- Case study: interpolation

$$\operatorname*{argmin}_{w \in \mathbb{R}^d} \sum (w^T\phi(x_i) - y_i)^2 + \lambda\|w\|^2 = \|\Phi w - y\|^2 + \lambda\|w\|^2$$

Define $\Phi = \begin{pmatrix} \phi(x_1)^T \\ \vdots \\ \phi(x_n)^T \end{pmatrix}$ $\qquad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$

$$\hat{w} = \operatorname*{argmin}_{w \in \mathbb{R}^d} w(\Phi^T\Phi + \lambda I)w - 2\, w^T\Phi^T y$$

SHERMAN - WOODBURY FORMULA

$$\hat{w} = (\Phi^T\Phi + \lambda I)^{-1}\Phi^T y = \Phi^T(\Phi\Phi^T + \lambda I)^{-1} y$$

$$\hat{f}(x) = \phi(x)^T\hat{w} = (\Phi\,\phi(x))\,(\Phi\Phi^T + \lambda I)^{-1} y$$

$$\mathbb{R}^n \ni \Phi\,\phi(x) = \begin{pmatrix} \phi(x_1)^T\phi(x) \\ \vdots \\ \phi(x_n)^T\phi(x) \end{pmatrix} \qquad \Phi\Phi^T \in \mathbb{R}^{n\times n} \qquad \Phi\Phi^T_{ij} = \phi(x_i)^T\phi(x_j)$$

Kernel $\quad k(x,x') = \phi(x)^T\phi(x')$

$$\hat{f}(x) = v(x)^T\left(K + \lambda I\right)^{-1} y$$

$$v(x) = \begin{pmatrix} k(x_1,x) \\ \vdots \\ k(x_n,x) \end{pmatrix} \qquad K_{ij} = k(x_i,x_j)$$

no dependence on d !!!

Example polynomials

$$\underbrace{\phi(x)}_{O(p)} = \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^p \end{pmatrix} \mapsto \underbrace{k(x,x')}_{O(1)} = \phi(x)^T\phi(x') = 1 + xx' + x^2x'^2 = \sum_{t=1}^{p}(xx')^t = \frac{1-(xx')^{p+1}}{1-xx'}$$

Infinite feature maps

$$\phi(x) = \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \end{pmatrix} \mapsto k(x,x') = \frac{1}{1-xx'}$$

$$\phi(x) = \begin{pmatrix} e^{-\frac{x^2}{2}} \\ \frac{x}{\sqrt{1!}}e^{-\frac{x^2}{2}} \\ \frac{x^2}{\sqrt{2!}}e^{-\frac{x^2}{2}} \\ \frac{x^3}{\sqrt{3!}}e^{-\frac{x^2}{2}} \\ \vdots \end{pmatrix} \mapsto k(x,x') = \sum_{t\in\mathbb{N}} e^{-\frac{x^2}{2}}\frac{(xx')^t}{t!}e^{-\frac{x'^2}{2}} = e^{-\frac{(x-x')^2}{2}}$$

Gaussian Kernel

We don't need explicit feature maps anymore.

Kernels

Let $K: X \times X \to \mathbb{R}$ be a function s.t.

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \geq 0 \quad \forall \alpha_1 .. \alpha_n \in \mathbb{R},\ x_1 ... x_n \in X,\ n \in \mathbb{N}$$

Aronszajn 1950

K is P.D. $\iff$ $\exists\ \phi : X \to H$, H Hilbert space

$$K(x, x') = \langle \phi(x), \phi(x') \rangle_H$$

Bochner theorem

$K(x,y) = v(x-y)$.

K is P.D. $\iff$ $\mathcal{F}\{v\}(\omega) \geq 0$ and $\mathcal{F}\{v\} \in L^1(\mathbb{R}^d)$

$$v(x-y) = \int e^{i\omega(x-y)} \mathcal{F}\{v\}(\omega)\, d\omega = \int \underbrace{\sqrt{v}(\omega)\, e^{i\omega x}} \sqrt{v}(\omega)\, e^{i\omega y}\, d\omega$$

$\phi : X \to L^2(\mathbb{R}^d)$

$\phi(x) = \sqrt{v}(\cdot)\, e^{i \cdot x}$

$$= \langle \phi(x), \overline{\phi(y)} \rangle_{L^2(\mathbb{R}^d)}$$

More generally given $\phi : X, \Omega \to \mathbb{R}$ with $\phi(x, \cdot) \in L^2(\Omega, \mu)$

$$K(x,y) = \langle \phi(x, \cdot), \phi(y, \cdot) \rangle_{L^2(\Omega, d\mu)}$$

Mercer Theorem

Given finite mesure ν on Ω

define $L : L^2(\nu) \to L^2(\nu)$ linear operator

$$(Lf)(x) = \int_\Omega k(x,y)\, f(y)\, d\nu(y)$$

When $\int |k(x,y)|^2 \, d\nu(y)\, d\nu(x) < \infty$ then $tr(L^2) < \infty$

and $(Lf)(x) = \sum_{j \in \mathbb{N}} \lambda_j \psi_j(x) \langle \psi_j(x'), f \rangle$ for some basis $(\psi_n)_{n \in \mathbb{N}} \in L^2(\nu)$

in particular $k(x,x') = \langle \phi(x), \phi(x') \rangle$, $\phi(x) = \begin{pmatrix} \sqrt{\lambda_1}\, \psi_1(x) \\ \sqrt{\lambda_2}\, \psi_2(x) \\ \vdots \end{pmatrix}$

Properties of kernels

K_1, K_2 kernel

↳ • $K_1 + K_2$ is a kernel

• $K_1 \cdot K_2$

• $K_1(g(\cdot), g(\cdot))$

K_1, K_2 p.d. matrix

↳ $K_1 + K_2$ p.d.

K_1, K_2 p.d. $\Rightarrow K_1 \circ K_2$ is p.d

Hadamard product

↳ $K_1 = \langle \phi(x), \phi(x) \rangle_H$

$K_1(g(y), g(y')) = \langle \phi'(y), \phi'(y) \rangle_H \quad \phi' = \phi(g(\cdot))$

Interesting subcases

- $K((x,y),(x',y')) = K_1(x,x') \cdot K_2(y,y')$

- Let $i_\Omega : \Omega \to X$ with $\Omega \subseteq X$
  $x \mapsto x$

  $K_\Omega(x,x') = K(i_\Omega(x), i_\Omega(x))$

Kernels on weird spaces

- strings
- graphs
- manifolds

Sobolev Spaces and convolution spaces

$W_2^m(\mathbb{R}^d) = \{ f : \mathbb{R}^d \to \mathbb{R} \mid \|f\|_{W_2^m}^2 = \sum_{|\alpha| \leq m} \int \left| \frac{\partial^\alpha f}{\partial x} \right|^2 dx < \infty \}$

↳ Charact. in terms of Potentials $W_2^m(\mathbb{R}^d) = \{ f \mid f = g * J, g \in L^2 \}$

- RKHS

Let $\phi: X \to V$ for V Hilbert space, then

$$w \in V \longrightarrow w^T\phi(\cdot): X \to \mathbb{R}$$

$$\mathcal{H} = \{ w^T\phi(\cdot) \mid w \in V \}$$

Definition based on the Kernel

$$\mathcal{H}_0 = \text{span}\left\{ \sum_{i=1}^{n} \alpha_i K(x, x_i) \mid \alpha_i \in \mathbb{R},\ x_i \in X \right\}$$

$\langle \cdot, \cdot \rangle \qquad \langle K(x, \cdot), K(z, \cdot) \rangle = K(x, z)$

$\forall f, g \in H_0$ $\langle \cdot, \cdot \rangle$ is well defined

$$\mathcal{H} = \overline{\mathcal{H}_0}^{\langle \cdot, \cdot \rangle} \qquad \text{norm } \|f\|_H^2 = \langle f, f \rangle$$

Evaluation functional

$$f \in H_0, \qquad K(x, \cdot) \in \mathcal{H}_0$$

$$\langle f, K(x, \cdot) \rangle = \sum \alpha_i \langle K(x_i, \cdot), K(x, \cdot) \rangle$$

$$= \sum \alpha_i K(x_i, x) = f(x)$$

$K(x, \cdot)$ is the evaluation functional

Algebraic properties of $\mathcal{H}$

$$K = K_1 + K_2 \rightarrow H = H_1 \oplus H_2$$

$$\|h\|_H^2 = \min_{\substack{f \in H_1 \\ g \in H_2 \\ f+g=h}} \|f\|_{H_1}^2 + \|g\|_{H_2}^2$$

$$K = K_1 \cdot K_2 \rightarrow \mathcal{H} = \mathcal{H}_1 \otimes H_2$$

$$\text{norm } \|f \otimes g\|_H = \|f\|_{H_1} \|g\|_{H_2}$$

$$\underbrace{K_1 \leq K_2}_{\sum \alpha_i \alpha_j K_1(x_i, x_j) \leq \sum \alpha_i \alpha_j K_2(x_i, x_j)} \rightarrow \mathcal{H}_1 \subseteq \mathcal{H}_2 \quad \|\cdot\|_{H_1} \geq \|\cdot\|_{H_2}$$

When $\mathcal{F}$ is a RKHS

Abstract Hilbert space $(\mathcal{F}, \langle \cdot, \cdot \rangle)$

$\mathcal{F}$ is RKHS iff

i.e. $\quad |f(x)| \leq M_x \|f\|_H \quad \forall f \quad \forall x \quad (*)$

Denote by $ev_x : \mathcal{F} \to \mathbb{R}$ the evaluation functional

$ev_x(f) = f(x)$.

1. $ev_x(\lambda f + \mu g) = \lambda f(x) + \mu g(x) = \lambda ev_x(f) + \mu ev_x(g)$
2. by $(*)$ it is also uniformly bounded then by Riesz theorem $ev_x \in \mathcal{F}$.

Then $\quad K(x, x') = \langle ev_x, ev_{x'} \rangle_{\mathcal{F}}$

References

Aronszajn, Nachman. "Theory of reproducing kernels." Transactions of the American mathematical society 68.3 (1950): 337-404
Berlinet, Alain, Christine Thomas-Agnan. Reproducing kernel Hilbert spaces in probability and statistics. Springer Science & Business Media, 2011.
Paulsen, Vern I., Mrinal Raghupathi. An introduction to the theory of reproducing kernel Hilbert spaces. Cambridge University Press, 2016.