Data Science for Actuaries (ACT6100)

Arthur Charpentier

Rappels # 3.3 (Reproducing Kernel Hilbert Space - RKHS) ⋆

automne 2020

https://github.com/freakonometrics/ACT6100/

Inner Product and Hilbert Spaces

An inner product on $\mathcal{H}$ is the application $(f,g) \mapsto \langle f,g\rangle_{\mathcal{H}}$ (taking value in $\mathbb{R}$) bilinear, symmetric, definite positive:

- $\langle f,g\rangle_{\mathcal{H}} = \langle g,f\rangle_{\mathcal{H}}$
- $\langle \alpha f+\beta g,h\rangle_{\mathcal{H}} = \alpha\langle f,h\rangle_{\mathcal{H}} + \beta\langle g,h\rangle_{\mathcal{H}}$
- $\langle f,f\rangle_{\mathcal{H}} \geq 0$ and $\langle f,f\rangle_{\mathcal{H}} = 0$ if and only if $f=0$.

Example : $\mathcal{H}=\mathbb{R}^n$, $\langle \boldsymbol{x},\boldsymbol{y}\rangle = \boldsymbol{x}^\top\boldsymbol{y}$

Example : $\mathcal{H}=\ell_2=\left\{u : \displaystyle\sum_{i=1}^{\infty} u_i^2<\infty\right\}$, $\langle u,v\rangle = \displaystyle\sum_{i=1}^{\infty} u_i v_i$

Example : $\mathcal{H}=L_2(\mu)=\left\{f : \displaystyle\int f(x)^2 d\mu(x)<\infty\right\}$,

$\langle f,g\rangle = \displaystyle\int f(x)g(x)d\mu(x)$

Note: A Hilbert space is an abstract vector space possessing the structure of an inner product.

Inner Product and Hilbert Spaces

If $\mathcal{H}$ is finite, $\mathcal{H} = \{h_1 \cdots, h_d\}$, $\langle x, y \rangle_{\mathcal{H}}$ takes value $K_{i,j}$ if $x = h_i$ and $y = h_j$. Let $\boldsymbol{K} = [K_{i,j}]$

$\boldsymbol{K}$ is a symmetric $d \times d$ matrix, $\boldsymbol{K} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$ for some orthogonal matrix $\boldsymbol{V}$ where columns are eigenvectors, and $\boldsymbol{\Lambda} = \text{diag}[\lambda_i]$ (positive values). Let

$$\Phi(x) = \big(\sqrt{\lambda_1} V_{i,1}, \sqrt{\lambda_2} V_{2,i}, \cdots, \sqrt{\lambda_d} V_{d,i}\big) \text{ if } x = h_i$$

Note that

$$K_{i,j} = [\boldsymbol{K}]_{i,j} = [\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top]_{i,j} = \sum_{l=1}^{d} \lambda_l V_{i,l} V_{l,j} = \langle \Phi(h_i), \Phi(h_j) \rangle$$

Matrix $\boldsymbol{K}$ defines an inner product, it is called a kernel. It is symmetric, associated with a positive semi-definite matrix. Then $K(u,u) \geq 0$ and $K(u,v) \leq \sqrt{K(u,u) \cdot K(v,v)}$.

Inner Product and Hilbert Spaces

Let $\Phi : u \mapsto K(\cdot, u)$, then $K(x, y) = \langle \Phi(x), \Phi(y) \rangle$
In a general setting, let $\|f\|_{\mathcal{H}} = \langle f, f \rangle_{\mathcal{H}}^{1/2}$, and define the distance from f to $\mathcal{G} \subset \mathcal{H}$

$$d(f, \mathcal{G}) = \inf_{g \in \mathcal{G}} \{\|f - g\|_{\mathcal{H}}\} = d(f, g^\star) \text{ where } g^\star \in \mathcal{G}$$

Note that $\langle g, f - g^\star \rangle_{\mathcal{H}} = 0$, $\forall g \in \mathcal{G}$. And $\mathcal{H} = \mathcal{G} \oplus \mathcal{G}^\perp$.

Riesz representation theorem For any continuous linear functionals L from $\mathcal{H}$ into the field $\mathbb{R}$, there exists a unique $g_L \in \mathcal{H}$ such that $\forall f \in \mathcal{H}$, $\langle g_L, f \rangle_{\mathcal{H}} = Lf$.

Consider the case where $\mathcal{H} = \mathbb{R}^n$. Let $\boldsymbol{\Sigma}$ denote some symmetric $n \times n$ positive definite matrix. Then

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{\Sigma}} = \boldsymbol{x}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y} \text{ is an inner product on } \mathbb{R}^n.$$

Inner Product and Hilbert Spaces

Note that if $\boldsymbol{\sigma}_i$ denote columns of $\boldsymbol{\Sigma} = [\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \cdots, \boldsymbol{\sigma}_n]$

$$\langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle_{\boldsymbol{\Sigma}} = \boldsymbol{\sigma}_i^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\sigma}_{\boldsymbol{j}} = \Sigma_{i,j},$$

and more generally, $\langle \boldsymbol{\sigma}_i, \boldsymbol{x} \rangle_{\boldsymbol{\Sigma}} = x_i$

The space $\mathcal{H}$ of functions $\mathbb{R}^p \to \mathbb{R}$ is a Reproducing Kernel Hilbert Space (RKHS) if there is an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ such that $\mathcal{H}$ with $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is an Hilbert space, and for all $\boldsymbol{x} \in \mathbb{R}^p$, linear functional $\delta_{\boldsymbol{x}} : \mathcal{H} \to \mathbb{R}$ defined as $\delta_{\boldsymbol{x}}(f) = f(\boldsymbol{x})$ is bounded.

Thus, $\mathcal{H}$ is a RKHS if and only if $\forall f \in \mathcal{H}$ and $\boldsymbol{x} \in \mathbb{R}^p$, there exists $M_{\boldsymbol{x}}$ such that $|f(\boldsymbol{x})| \leq M_{\boldsymbol{x}} \cdot \|f\|_{\mathcal{H}}$.
From Riesz theorem, there exists a unique $\zeta_{\boldsymbol{x}} \in \mathcal{H}$ associated with $\delta_{\boldsymbol{x}}$, i.e. $\langle \zeta_{\boldsymbol{x}}, f \rangle_{\mathcal{H}} = f(\boldsymbol{x})$

Function $\boldsymbol{x} \mapsto \zeta_{\boldsymbol{x}}$ is called reproducing function in $\boldsymbol{x}$ and $K : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ defined as $K(\boldsymbol{x}, \boldsymbol{y}) = \zeta_{\boldsymbol{x}}(\boldsymbol{t})$ is the reproducing kernel of $\mathcal{H}$.

Inner Product and Hilbert Spaces

Observe that $\langle K(\boldsymbol{x},\cdot),K(\boldsymbol{y},\cdot)\rangle_{\mathcal{H}} = K(\boldsymbol{x},\boldsymbol{y})$.

The kernel is unique, and is (semi-)definite positive.
If $\mathcal{H}$ is a closed subspace of Hilbert space $\mathcal{X}$. For any function $f\in\mathcal{X}$, $\boldsymbol{x}\mapsto\langle f,K(\cdot,\boldsymbol{x})\rangle_{\mathcal{X}}$ is the projection of f on $\mathcal{H}$.
Note that conversely, Moore-Aronszajn's theorem allows to create a RKHS from a definite positive kernel K.

Inner Product and Hilbert Spaces

Mercer's kernel Let μ denote some measure on $\mathbb{R}^p$ and $\mathcal{H} = L^2(\mathbb{R}^p, \mu)$, define

$$(L_K f)(\boldsymbol{x}) = \int K(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) d\mu(\boldsymbol{y})$$

which is a compact bounded linear operator, self-adjoint and positive. Let $\lambda_1 \geq \lambda_2 \geq \cdots$ denote eigenvalues of L_k, with (orthonormal) eigenvectors $\psi_1, \psi_2, \cdots$, then

$$K(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=1}^{p} \lambda_k \psi_k(\boldsymbol{x}) \psi_k(\boldsymbol{y}) = \Psi(\boldsymbol{x})^\top \Psi(\boldsymbol{y}) = \langle \Psi(\boldsymbol{x}), \Psi(\boldsymbol{y}) \rangle_{L^2}$$

where $\Psi(\boldsymbol{x}) = (\sqrt{\lambda_k} \psi_k(\boldsymbol{x}))$.

Inner Product and Hilbert Spaces

Example : Consider the space $\mathcal{H}$ defined as

$$\mathcal{H}_1 = \Big\{f : [0,1] \to \mathbb{R} \text{ continuously differentiable, with } f' \in L^2([0,1]) \text{ and } f(0) = 0\Big\}$$

$\mathcal{H}_1$ is an Hilbert space on $[0,1]$ with inner product

$$\langle f, g\rangle_{\mathcal{H}_1} = \int_0^1 f'(t)g'(t)dt$$

with (definite positive) kernel $K_1(x,y) = \min\{x,y\}$:

$$\langle f, K(x,\cdot)\rangle_{\mathcal{H}_1} = \int_0^1 f'(t)\underbrace{\frac{\partial K_1(t,x)}{\partial x}}_{=\mathbf{1}_{[0,x]}(t)} = \int_0^x f'(t)dt = f(x)$$

Inner Product and Hilbert Spaces

Example : Consider the Sobolev space $W^1([0,1])$ defined as

$$W^1([0,1]) = \Big\{f:[0,1]\to\mathbb{R} \text{ continuously differentiable, } \text{with } f'\in L^2([0,1])\Big\}$$

Observe that $W^1([0,1]) = \mathcal{H}_0 \oplus \mathcal{H}_1$ where

$$\mathcal{H}_0 = \Big\{f:[0,1]\to\mathbb{R} \text{ continuously differentiable, with } f'=0\Big\}$$

The later is an Hilbert space with kernel $K_0(x,y)=1$.
One can consider kernel $K(x,y)=K_0(x,y)+K_1(x,y)$ (related to linear splines). More generally, consider

$$\mathcal{H}_2 = \Big\{f:[0,1]\to\mathbb{R} \text{ twice cont. diff., } \text{with } f''\in L^2([0,1] \text{ and } f'(0)=0)\Big\}$$

Then $\langle f,g\rangle_{\mathcal{H}_2} = \displaystyle\int_0^1 f''(t)g''(t)dt$ is an inner product, with kernel

$$K_2(x,y) = \int_0^1 (x-t)_+(y-t)_+dt$$

Inner Product and Hilbert Spaces

Consider some Hilbert space $\mathcal{H}$ with kernel K and some functional $\Psi : \mathbb{R}^{n+1} \to \mathbb{R}$, increasing in its last argument.
Given $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$, $\min_{f\in\mathcal{H}} \left\{\Psi(f(\boldsymbol{x}_1), \cdots, f(\boldsymbol{x}_n), \|f\|_{\mathcal{H}})\right\}$ admits solution

$$\forall \boldsymbol{x},\ f(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}_i, \boldsymbol{x})$$

A classical expression for Ψ is, for some convex function ψ,

$$\Psi(f(\boldsymbol{x}_1), \cdots, f(\boldsymbol{x}_n), \|f\|_{\mathcal{H}}) = \psi(\boldsymbol{y}, f(\boldsymbol{x}_1), \cdots, f(\boldsymbol{x}_n)) + \lambda\|f\|_{\mathcal{H}}$$

$$\Psi(f(\boldsymbol{x}_1), \cdots, f(\boldsymbol{x}_n), \|f\|_{\mathcal{H}}) = \sum_{i=1}^{n} \ell(\boldsymbol{y}, f(\boldsymbol{x}_i)) + \lambda\|f\|_{\mathcal{H}}$$

Inner Product and Hilbert Spaces

Assume that $y_i = m(x_i) + \varepsilon_i$, where $m \in W_2([0,1])$, then polynomial splines of degree 2 is the solution of

$$\min_{m \in W_2} \left\{ \frac{1}{n} \sum_{i=1}^{n} \big(y_i - m(x_i)\big)^2 + \nu \int_0^1 [m''(t)]^2 dt \right\}$$

then $m^\star(x) = \beta_0 + \beta_1 x + \sum_{i=1}^{n} \gamma_i K_2(x_i, x)$ Note that one can use a matrix representation

$$\min \left\{ \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Q}\gamma\right)^\top \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Q}\gamma\right) + n\nu\gamma^\top \boldsymbol{Q}\gamma \right\}$$

where $\boldsymbol{Q} = [K_1(x_i, x_j)]$. If $\boldsymbol{M} = \boldsymbol{Q} + n\nu\mathbb{I}$,

$$\boldsymbol{\beta}^\star = \left(\boldsymbol{X}^\top \boldsymbol{M}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{M}^{-1} \boldsymbol{y}$$

$$\gamma^\star = \boldsymbol{M}^{-1}\left(\mathbb{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{M}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{M}^{-1}\right)\boldsymbol{y}$$

Inner Product and Hilbert Spaces

Kimeldorf & Wahba's representation theorem

Consider a kernel K and $\mathcal{H}_K$ the associated RKHS. For any (convex) loss function $\ell : \mathbb{R}^2 \to \mathbb{R}$, the solution

$$m^\star \in \underset{m\in\mathcal{H}_K}{\text{argmin}} \sum_{i=1}^{n} \ell(y_i, m(\boldsymbol{x}_i)) + \|m\|^2_{\mathcal{H}_K}$$

can be expressed

$$m^\star(\cdot) = \sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}_i, \cdot)$$

See Kimeldorf & Wahba (1971, Some results on Tchebycheffian spline functions), and some slides.

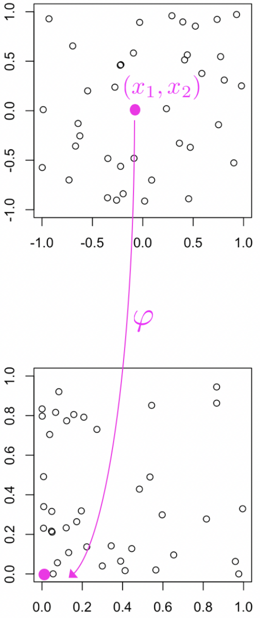
Inner Product and Hilbert Spaces

For more technical (mathematical) results, see Wahba (1990, Spline Models for Observational Data)

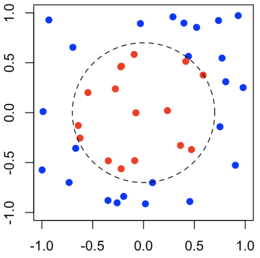
We've seen that in many cases, $K(\boldsymbol{x},\boldsymbol{y}) = \varphi(\boldsymbol{x})^\top\varphi(\boldsymbol{y})$
for some $\varphi:\mathbb{R}^p\rightarrow\mathbb{R}^q$ (here $q=p$)

Consider for example $\varphi:\begin{pmatrix}x_1\\x_2\end{pmatrix}\mapsto\begin{pmatrix}x_1^2\\x_2^2\end{pmatrix}$

so that $K(\boldsymbol{x},\boldsymbol{y}) = x_1^2y_1^2+x_2^2y_2^2$



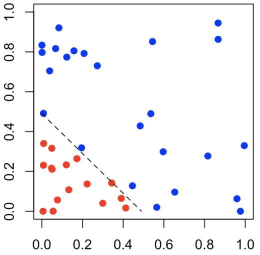
Inner Product and Hilbert Spaces



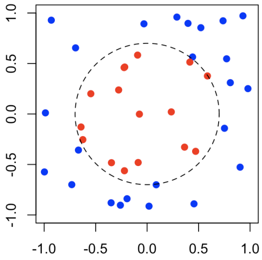
Consider $\varphi : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1^2 \\ x_2^2 \end{pmatrix}$

$K(\boldsymbol{x}, \boldsymbol{y}) = x_1^2 y_1^2 + x_2^2 y_2^2$

From data $(y_i, \boldsymbol{x}_i)$, transform the covariates into $(y_i, \phi(\boldsymbol{x}_i))$, and use a (classical) linear model



Inner Product and Hilbert Spaces



Consider $\varphi : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ x_1^2 + x_2^2 \end{pmatrix}$

$$K(\boldsymbol{x}, \boldsymbol{y}) = x_1 y_1 + (x_1^2 + x_2^2)(y_1^2 + y_2^2)$$