Recall. Reproducing Kernel Hilbert Space.

Idea. for linear regression. $W = X^T \alpha$ $\{x_1 \cdots x_n\}$ are n data.

$$\Downarrow \quad f(x) = \langle w, x \rangle$$

$$w = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$$

f is linear combination of $k(x_i, \cdot)$

For kernel space: final function $f = \alpha_1 k_{x_1} + \alpha_2 k_{x_2} + \cdots + \alpha_n k_{x_n}$

basis function

Kernel: $k(x,y) = \langle k_x, k_y \rangle$ $f(x) = \langle f, k_x \rangle$, k_x is the mapping from f to $f(x)$

$$f(y) = \langle f, k_y \rangle = \langle \alpha_1 k_{x_1} + \alpha_2 k_{x_2} + \cdots + \alpha_n k_{x_n}, k_y \rangle$$

$$= \alpha_1 \langle k_{x_1}, k_y \rangle + \alpha_2 \langle k_{x_2}, k_y \rangle + \cdots + \alpha_n \langle k_{x_n}, k_y \rangle$$

$f(\cdot) = \sum \alpha_i k(x_i, \cdot)$ as basis function

$$= \alpha_1 k(x_1, y) + \alpha_2 k(x_2, y) + \cdots + \alpha_n k(x_n, y)$$

§ Hilbert Space. (Vector space $x_1, \cdots, x_n \in V \Rightarrow \alpha_1 x_1 + \cdots + \alpha_n x_n \in V$ + inner product)

$\langle \cdot, \cdot \rangle$ is a function

$\langle f, g \rangle \to \mathbb{R}$

f and g lies in a vector space

$$\begin{cases} \langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle = \alpha_1 \langle f_1, g \rangle + \alpha_2 \langle f_2, g \rangle \\ \langle f, g \rangle = \langle g, f \rangle \\ \langle f, f \rangle_H \geq 0 \end{cases}$$

$= \|f\|_H^2$

$$\text{why} \quad K = \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & & & \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{pmatrix} \text{ is P.S.D. ?}$$

Example Functions $(f+g)(x) = f(x) + g(x)$ $\forall x \in \mathbb{R}^d$

examples $\rightarrow$ $\langle f, g \rangle_{L_2} = \int f(x) g(x) \, dx$ $\quad \langle f, f \rangle = \|f\|_{L_2} = \int f^2 dx$

$\rightarrow$ $\langle f, g \rangle = \int f(x) g(x) + f'(x) g'(x) \, dx$

for a Reproducing kernel Hilbert space

$-\sup_x \|k_x\| \leq B$

$f(x) \leq \|k_x\| \cdot \|f\|_H$

Reproducing k_x : is a (bounded) mapping $f \to f(x)$

$\langle f, k_x \rangle = f(x)$

- The function space is a Hilbert space

$\Rightarrow f = \sum \alpha_i k_{x_i}$ means

$f(y) = \sum \alpha_i \langle k_{x_i}, k_y \rangle = \sum \alpha_i k(x_i, y)$

- kernel : $k(x,y) = \langle k_x, k_y \rangle$

??

$\boxed{\text{Not Required}}$ If we have an innerproduct, can we write down the kernel? $\Rightarrow$ Not always, in many cases, $\|k_x\|$ is ∞

In some of cases, we can do this.

Examples $\langle f, g \rangle_H = \int_0^1 f'(x) g'(x) \, dx$, $f, g: [0,1] \to \mathbb{R}$

Claim $k(x,z) = \min\{x, z\}$ $f(0) = 0$.

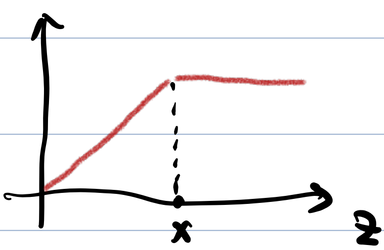
$k_x = k(x, \cdot)$, $\langle f, k_x \rangle = f(x)$

Check $\langle f, k_x \rangle = f(x)$

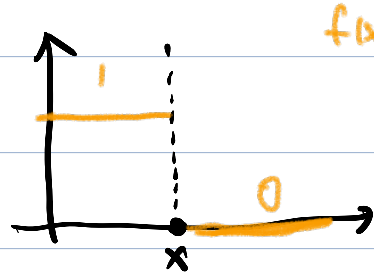
$= \int_0^1 f'(z) k_x'(z) dz$ Integral by parts!! $-\int_0^1 f(z) k_x''(z) dz$

$= f(x)$

$k_x(z) \rightarrow$



$k_x'(z)$

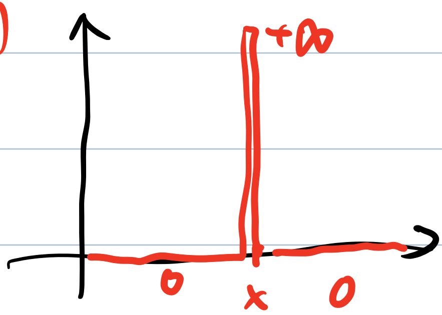


①

$\int f'(z) k_x'(z) dz$

$= \int_0^x f'(z) dz = f(x)$

$-k_x''(z)$
$\|$
δ_x
dirac



$\to \int f \cdot k_x'' = f(x)$

$\boxed{k_x'' = -\delta_x}$ ②

$f(x) = \int f \cdot \delta_x$ $\to$ $\langle f, \delta_x \rangle_{L_2} = f(x)$

$\langle f, k_x \rangle_H = f(x)$

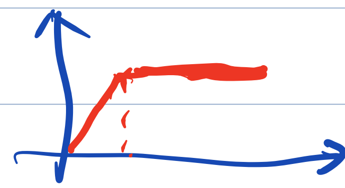
Smoothness of Reproducing Kernel Hilbert Space

[Not Required]

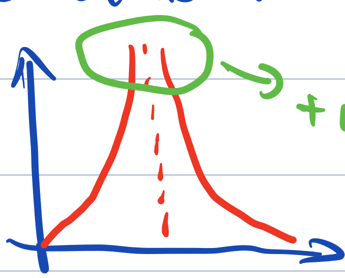
Kernel is the (Green's Function)

$$\langle f, g \rangle_H = \int f' g' \to k_x'' = -\delta_x$$

in 1 dimension, the solution is



but in 2 dimension



+∞

→ the kernel is not bounded

⇒ It's not reproducing!

$$\langle f, g \rangle_H = \int f'' g'' \to k_x'''' = \delta_x$$

In 1-3 dimension It is the bounded, in 4-dimen It's not!

Informal: In R^d dimension, $f^{(d/2)}$ should exist

Representer Theorem.

$\langle f, k_x \rangle$ $\langle f, k_{x_n} \rangle$ $\langle f, f \rangle_H$

If one minimizing

$$f^* = \underset{f \in H}{\operatorname{argmin}} \, L\left(f(x_1), f(x_2), \cdots, f(x_n) \right) + \Omega\left(\|f\|_H^2 \right)$$

→ function space (may be infinite dimensional)

Then. $f^* = \sum_{i=1}^{n} \alpha_i k_{x_i}$

or $f^*(\cdot) = \sum_{i=1}^{n} k(x_i, \cdot)$

" Dual Solution only depend on the number of data"

- Not just holds for RKHS. It holds for a lot of hypothesis space, even Neural Networks.

§ Rademacher Complexity of a Kernel Class.

Intrinsic assumption $\frac{1}{n}\sum_{i=1}^{n} k(x_i,x_i)$ uniformly bound

$$\hat{R}_{S_n}\left(\{f\in\mathcal{F} \mid \underbrace{\|f\|_H \le M}_{\langle f,f\rangle_H \le M^2}\}\right) \le \frac{M}{n}\sqrt{\sum_{i=1}^{n} k(x_i,x_i)}$$

$$= \frac{M}{\sqrt{n}}\sqrt{\frac{1}{n}\sum_{i=1}^{n} k(x_i,x_i)}$$

↳ This is the trace of Kernel Matrix.

Proof. $X_1 \cdots X_n$ is data

$$\text{LHS} = \frac{1}{n}\mathbb{E}_\sigma\left[\sup_{\{f\in\mathcal{F} \mid \|f\|_H\le M\}} \sum_{i=1}^{n}\sigma_i f(x_i)\right]$$

$$= \frac{1}{n}\mathbb{E}_\sigma\left[\sup_{\{f\in\mathcal{F} \mid \|f\|_H\le M\}} \sum_{i=1}^{n}\sigma_i \langle f, k_{x_i}\rangle_H\right]$$

$$= \frac{1}{n}\mathbb{E}_\sigma\left[\sup_{\{f\in\mathcal{F} \mid \|f\|_H\le M\}} \langle f, \sum_{i=1}^{n}\sigma_i k_{x_i}\rangle_H\right]$$

$$\le \frac{1}{n} M\, \mathbb{E}_\sigma\left\|\sum_{i=1}^{n}\sigma_i k_{x_i}\right\| \quad (\text{This is because } \|f\|_H \le M)$$

$$= \frac{1}{n} M\, \mathbb{E}_\sigma\sqrt{\langle \sum_{i=1}^{n}\sigma_i k_{x_i}, \sum_{i=1}^{n}\sigma_i k_{x_i}\rangle} \quad \left(\|f\| = \sqrt{\langle f,f\rangle}\right)$$

$$\le \frac{1}{n} M\sqrt{\mathbb{E}_\sigma \langle \sum_{i=1}^{n}\sigma_i k_{x_i}, \sum_{i=1}^{n}\sigma_i k_{x_i}\rangle} \quad (\sqrt{\cdot} \text{ is a concave} \text{ Jensen's Inequality})$$

$$= \frac{1}{n} M\sqrt{\mathbb{E}\sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_i\sigma_j \langle k_{x_i}, k_{x_j}\rangle^{= k(x_i,x_j)}}$$

$\mathbb{E}\,\sigma_i\sigma_j = 0$ if $i\ne j$

$$= \frac{1}{n} M\sqrt{\sum_{i=1}^{n} k(x_i, x_i)}$$

$f(\cdot)$ is concave.

$c_1 f(x) + c_2 f(y) \le f(c_1 x + c_2 y)$

Matches "Least square's variance is the trace of covariance matrix"

§ Spectral View of RKHS (Informal)

(HW 7, Sobolev)

The Complexity is the trace(K)

$$= \sum_{i=1}^{\infty} \lambda_j(K) < \infty$$

$\Downarrow$

The eigenvalues should be decay fast enough!!

Example Shift-Invariant Kernel

$\Rightarrow$ eigenvectors: Fourier basis!!

" eigenvalue will mean Fourier coef "

$\Rightarrow$ "rough understanding"

If a function lies in a RKHS, "fast enough"

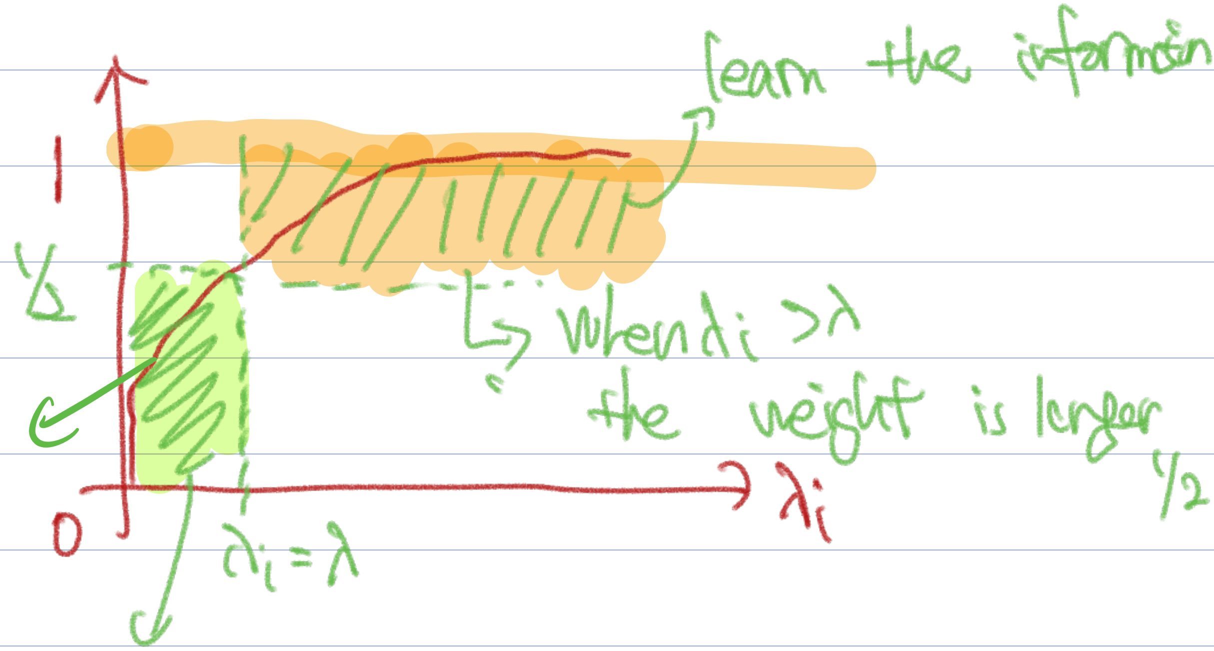
then the Fourier coef should decay

$$\alpha = (XX^T + \lambda I)^{-1} X^T y \quad \text{(Ridge Regression)}$$

If $XX^T = \sum \lambda_i u_i u_i^T$ is the eigendecomposition.

$$\langle x, \alpha \rangle = \sum_{i=1}^{n} \underbrace{\frac{\lambda_i}{\lambda_i + \lambda}}_{\text{weight}} \underbrace{\langle y, u_i \rangle}_{\text{how much you have on } u_i}$$

$$\frac{\lambda_i}{\lambda_i+\lambda}$$



neglect

if $\lambda_i < \lambda$, weight is smaller than 1/2

We only learn the part $\lambda_i > \lambda$

$$\Rightarrow \sum_{\lambda_i > \lambda} d_i < \infty$$

"mis-specification" $f \notin H$

$A: x \to Ay$ Consider matrix as a linear mapping.

$$K = \begin{pmatrix} k(x_1, x_1) & & \\ & \ddots & \\ & & k(x_n, x_n) \end{pmatrix}$$

$\underline{g = Kf}$

$$f = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} \quad g = \begin{bmatrix} g(x_1) \\ \vdots \\ g(x_n) \end{bmatrix}$$

$$\downarrow$$

$$g(x_j) = \frac{1}{n} \sum_{i=1}^{n} k(x_j, x_i) f(x_i)$$

$$\Downarrow$$

$$g(x) = \int k(x, y) f(y)\, dy$$

in population. "kernel matrix" as the integral operator

$$f \to \underline{g} = \int k(x, y) f(y)\, dy.$$