April 30, 2020

Kernel: $k(x,y) = k(y,x)$ $\in \mathbb{R}^d$ measure similarity

Feature map: $\psi_1, \dots, \psi_m$ $\quad \psi_j : \mathbb{R}^d \to \mathbb{R}$

$\psi : \mathbb{R}^d \to \mathbb{R}^m \quad \psi(x) = \begin{pmatrix} \psi_1(x) \\ \vdots \\ \psi_m(x) \end{pmatrix}$ $\quad m > n = \#$ of data points

Data: $X = \{x_1, \dots, x_n\}$ $\quad x_i \in \mathbb{R}^d$

$\min_c \|c\|_2^2$ s.t. $\Psi c = f_X$ $\quad \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}$

$$\Psi = \begin{array}{c c} & m \\ n & \begin{bmatrix} \psi(x_1)^T \\ \vdots \\ \psi(x_n)^T \end{bmatrix} \end{array}$$

$\hat{f} = \sum_{j=1}^m c_j \psi_j \quad \hat{f} \approx f$

if $\psi_1, \dots, \psi_m$ are orthonormal basis for some inner product space H

$\min \|\hat{f}\|_H^2$ s.t. $\hat{f}_X = f_X$

$k_y \in H \qquad k_y = \sum_j \underbrace{\psi_j(y)}_{\text{coeff}} \; \underbrace{\psi_j}_{\text{function}}$

$$(x) = \sum_j \psi_j(y)\,\psi_j(x)$$

$$= k(y, x)$$

$$\langle k_y, k_x \rangle_H$$

$$= \Big\langle \sum_i \psi_i(y)\,\psi_i,\ \sum_j \psi_j(x)\,\psi_j \Big\rangle_H$$

$$= \sum_{i,j} \psi_i(y)\,\psi_j(x)\,\underbrace{\cancel{\langle \psi_i, \psi_j \rangle_H}}_{\delta_{ij}}$$

ℓ_2

$$= \sum_j \psi_j(y)\,\psi_j(x)$$

$$\boxed{k_y(x) = k(y, x) = \langle \psi(y), \psi(x) \rangle_{\ell_2} = \langle k_y, k_x \rangle_H}$$

RKHS: $g(x) = \langle g, k_x \rangle_H$ for any $g \in H$

Note: need to be careful in infinite dimensions

Note! $x_1, \ldots, x_n \qquad (K_{xx})_{ij} = k(x_i, x_j)$

$K_{xx} = \underline{\Psi}\,\underline{\Psi}^T \qquad n \times m \ \boxed{\Psi}$ positive semi-def.