Kernel:

let $\mathcal{X}$ be an non-empty set. A function $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a kernel if

$\exists$ $\mathbb{R}$- Hilbert space $\mathcal{H}$ and $\phi: \mathcal{X} \to \mathcal{H}$ s.t. $k(x,y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$ $\forall$ $x, y \in \mathcal{X}$

Remark: $k(\cdot,\cdot)$ is symmetric since $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is

construct kernel:

let $K \in \mathbb{R}^{n \times n}$ $\quad K_{ij} = k(x_i, x_j)$

$$\begin{bmatrix} K \end{bmatrix} = \underbrace{\begin{array}{c} u_i \\ u_j \end{array}\begin{bmatrix} \cdots \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \quad \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} \cdots \end{bmatrix}}_{U^T}$$

since K is symmetric, it has real eigenvalues

consider feature $\phi(x_i) = [\sqrt{\lambda_1} u_{i1} \; \cdots \; \sqrt{\lambda_n} u_{in}]$

then $\langle \phi(x_i), \phi(x_j) \rangle = k(x_i, x_j)$

PSD Functions

A symmetric function $k: \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ is PSD

if $K \in \mathbb{R}^{n \times n}$ where $K_{ij} = k(x_i, x_j)$ is PSD

lemma.

let $\mathcal{H}$ be any Hilbert space, $\mathcal{X}$ be an non-empty set

$\phi: \mathcal{X} \to \mathcal{H}$, then

1° $k(x,y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$ is a PSD function

2° if k is a psd function, $\exists$ $\phi: \mathcal{X} \mapsto \mathcal{H}$ s.t. $k(x,y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$

1° take $K_{ij} = k(x_i, x_j)$ $\quad K = U \Lambda U^T$

if $\lambda_k < 0$, let $z = \sum_{i=1}^{n} u_{ik} \phi(x_i)$

$$\begin{bmatrix} \; | & \cdots & u_k & \cdots & | \; \end{bmatrix}$$

then
$$\begin{aligned} \|z\|_2^2 &= \left(\sum_{i=1}^{n} u_{ik} \phi(x_i)\right)^T \left(\sum_{i=1}^{n} u_{ik} \phi(x_i)\right) \\ &= \sum_i \sum_j u_{ik} u_{jk} \langle \phi(x_i), \phi(x_j) \rangle \\ &= u_k K u_k \\ &= \lambda \|u_k\|_2^2 < 0 \quad \text{impossible} \end{aligned}$$

Eg. polynomial kernal

$$k(x,y) = (x^T y)^2 = \sum_i x_i z_i \sum_j x_j z_j = \sum_i \sum_j (x_i x_j)(z_i z_j)$$

$$\phi(x) = [x_1 x_1 \;\; x_1 x_2 \;\; x_1 x_3 \;\; x_2 x_1 \;\; \cdots \;\; x_3 x_3]$$

$$k(x,y) = (x^T y + c)^2 \qquad \phi(x) = [x_1 x_1 \;\; x_1 x_2 \;\; x_2 x_1 \;\; x_2 x_2 \;\; \sqrt{2c}\, x_1 \;\; \sqrt{2c}\, x_2 \;\; c]$$

$$k(x,y) = (x^T z + c)^d \qquad \phi(x) = [\text{all polynomial term up to degree } d]$$

Eg. Gaussian kernel

$$K(x,z) = \exp\left(-\frac{\|x-z\|_2^2}{2\sigma^2}\right)$$

$$= \exp\left(-\frac{\|x\|_2^2}{2\sigma^2}\right) \cdot \exp\left(-\frac{\|z\|_2^2}{2\sigma^2}\right) \exp\left(\frac{x^T z}{\sigma^2}\right)$$

$$= \exp\left(-\frac{\|x\|_2^2}{2\sigma^2}\right) \cdot \exp\left(-\frac{\|z\|_2^2}{2\sigma^2}\right) \left[1 + \frac{1}{\sigma^2}(x^T z) + \frac{1}{2!\,\sigma^2}(x^T z)^2 + \cdots \frac{1}{n!\,\sigma^2}(x^T z)^n \cdots\right]$$

$$= \left\langle \exp\left(-\frac{\|x\|_2^2}{2\sigma^2}\right) \begin{bmatrix} 1 \\ \sigma [x] \\ \frac{\sigma}{\sqrt{2!}} \begin{bmatrix} x_1 x_1 \\ x_1 x_2 \\ \vdots \end{bmatrix} \\ \frac{\sigma}{\sqrt{3!}} \begin{bmatrix} \vdots \\ \vdots \end{bmatrix} \end{bmatrix}, \; \cdots \right\rangle$$