6. Kernel Methods

/ /

§1. Overview / Definition.

(1) $k : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ ··· kernel.

$k(x, x')$

(2) Positive-definite symmetric kernel (PDS kernel).

i) (Symmetric) $k(x, x') = k(x', x)$

ii) (PD) $\forall m \in \mathbb{N}, \forall c_1, \cdots, c_m \in \mathbb{R}, \forall x_1, \cdots, x_m \in \mathcal{X}, \sum_{i,j=1}^{m} c_i c_j k(x_i, x_j) \geq 0.$

(3) Typical Example.

$\Phi : \mathcal{X} \longrightarrow \mathbb{R}^N, \quad k(x, x') = \langle \Phi(x), \Phi(x') \rangle$

↳ or a inner product space.

$$\sum_{i,j=1}^{m} c_i c_j k(x_i, x_j) = \sum_{i,j=1}^{m} c_i c_j \langle \Phi(x_i), \Phi(x_j) \rangle$$
$$= \langle \sum_{i=1}^{m} c_i \Phi(x_i), \sum_{j=1}^{m} c_j \Phi(x_j) \rangle$$
$$= \| \sum_{i=1}^{m} c_i \Phi(x_i) \|^2,$$

RKHS.

$K = U \Sigma U^T$, for $U = [u_1 \cdots u_m]$: orthonormal eigen vectors in each column

$\Sigma = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix}$ for λ_i : eigen vectors.

$\because) \; KU = U\Sigma$

Example.

① $\mathcal{X} = \mathbb{R}^n, \; c \geq 0, \; d \in \mathbb{N}$

$k(x, x') = (\langle x, x' \rangle + c)^d$

$n = 2, \; d = 2.$

$(\langle (x, y), (x', y') \rangle + c)^2 = (xx' + yy' + c)^2$

$= x^2 x'^2 + y^2 y'^2 + 2xx'yy' + 2c(xx' + 2cyy' + c^2$

$$= \langle \Phi(x,y), \Phi(x',y') \rangle$$

$$\text{for } \Phi(x,y) = \begin{bmatrix} x^2 \\ y^2 \\ \sqrt{2}xy \\ \sqrt{2c}\,x \\ \sqrt{2c}\,y \\ c \end{bmatrix}$$

② $K_0(x,x') = \langle \Phi_0(x), \Phi_0(x') \rangle_{\mathbb{R}^{n_0}}$

$K_1(x,x') = \langle \Phi_1(x), \Phi_1(x') \rangle_{\mathbb{R}^{n_1}}$

$$K(x,x') = K_0(x,x') + K_1(x,x')$$
$$= \langle \psi(x), \psi(x') \rangle_{\mathbb{R}^{n_0+n_1}}$$
$$\text{with } \psi(x) = \begin{bmatrix} \Phi_0(x) \\ \Phi_1(x) \end{bmatrix}$$

③ $K(x,x') = K_0(x,x')\,K_1(x,x')$

$$= \langle \Phi_0(x), \Phi_0(x') \rangle_{\mathbb{R}^{n_0}} \langle \Phi_1(x), \Phi_1(x') \rangle_{\mathbb{R}^{n_1}}$$
$$= \sum_{i=1}^{n_0} \Phi_{0,i}(x)\Phi_{0,i}(x') \sum_{j=1}^{n_1} \Phi_{1,j}(x)\Phi_{1,j}(x')$$
$$= \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \Phi_{0,i}(x)\Phi_{1,j}(x)\ \Phi_{0,i}(x')\Phi_{1,j}(x')$$
$$= \langle \psi(x), \psi(x') \rangle_{\mathbb{R}^{n_0 \times n_1}}$$

$$\text{for } \psi(x) = \mathrm{vec}\left(n_0 \left\{ \overbrace{\begin{bmatrix} \Phi_{0,i}(x)\Phi_{1,j}(x) \end{bmatrix}}^{n_1} \right. \right)$$

④ Gaussian Kernel. (RBF Kernel)

$k(x, x') = \exp\left(-\frac{1}{\sigma^2}\|x - x'\|_2^2\right)$

⑤ $k(x, x') = \tanh(c\langle x, x'\rangle + d)$, $c, d \geq 0$.

§2 Reproducing Kernel Hilbert Space (RKHS).

(1) PDS Kernel k

$k(x, x') = \langle \Phi(x), \Phi(x') \rangle_{RKHS_k}$

Theorem RKHS is canonical.

$$\begin{tikzcd} & \mathbb{H}' \\ \mathcal{X} \arrow[ur, "\Phi'"] \arrow[dr, "\Phi"'] & \\ & RKHS_k \arrow[uu, dashed, "\exists!\psi"'] \end{tikzcd}$$

$k(x, x') = \langle \Phi(x), \Phi(x') \rangle_{RKHS_k}$

$= \langle \Phi'(x), \Phi'(x) \rangle_{\mathbb{H}'}$

Then, $\exists! \psi$ that is ① linear

② preserves inner product

③ $\Phi' = \psi \circ \Phi$.

Note. Denote $\mathbb{H} = RKHS_k$.

(Step 1) Construct inner product space $\mathbb{H}_0$

(Step 2) $\mathbb{H}$, the completion of $\mathbb{H}_0$.

① $\mathbb{H}_0 \subseteq [\mathcal{X} \to \mathbb{R}]$ as $\mathbb{H}_0 = \{\sum_{i=1}^{m} a_i k(x_i, \cdot) \mid m \in \mathbb{N}, x_1, \cdots, x_m \in \mathcal{X}, a_1, \cdots, a_m \in \mathbb{R}\}$.

$= \mathrm{span}\{k(x, \cdot) \mid x \in \mathcal{X}\}$.

$\mathbb{H}_0$ is a vector space over $\mathbb{R}$, with point-wise addition

" constant multiplication.

$\frac{1}{(\ln^2 K)^{1/3}} = \frac{1}{\ln}$.

$K = \frac{n}{2} \times L^3$

/ /

(The inner product)

Define a inner product on $\mathbb{H}_0$ as
$$\left\langle \sum_{i=1}^{m} a_i k(x_i,\cdot), \sum_{j=1}^{n} b_j k(x_j,\cdot) \right\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j k(x_i,x_j).$$

- The inner product is well-defined.

If $\sum_{i=1}^{m} a_i k(x_i,\cdot) = \sum_{\ell=1}^{p} c_\ell k(x_\ell,\cdot)$, then
$$\begin{aligned}\left\langle \sum_{\ell=1}^{p} c_\ell k(x_\ell,\cdot), \sum_{j=1}^{n} b_j k(x_j,\cdot) \right\rangle &= \sum_{\ell=1}^{p}\sum_{j=1}^{n} c_\ell b_j k(x_\ell,x_j) \\ &= \sum_{j=1}^{n} b_j \left(\sum_{\ell=1}^{p} c_\ell k(x_\ell,x_j)\right) \\ &= \sum_{j=1}^{n} b_j \left(\sum_{i=1}^{m} a_i k(x_i,x_j)\right) \\ &= \left\langle \sum_{i=1}^{m} a_i k(x_i,\cdot), \sum_{j=1}^{n} b_j k(x_j,\cdot) \right\rangle.\end{aligned}$$

Thus, inner prod. independent of representation of the first component $\sum_{i=1}^{m} a_i k(x_i,\cdot)$.

Similar for the 2nd component.

- Inner product is symmetric.

$$\begin{aligned}\left\langle \sum a_i k(x_i,\cdot), \sum b_j k(x_j,\cdot) \right\rangle &= \sum_i\sum_j a_i b_j k(x_i,x_j) \\ &= \sum_i\sum_j a_i b_j k(x_j,x_i) \\ &= \sum_j b_j a_i k(x_j,x_i) \\ &= \left\langle \sum b_j k(x_j,\cdot), \sum a_i k(x_i,\cdot) \right\rangle\end{aligned}$$

- $\forall h \in \mathbb{H}_0$, $\langle h,h\rangle \geq 0$, and $\langle h,h\rangle = 0 \Leftrightarrow h = 0$.

"Reproducing property"

Theorem	k: PSD kernel on $\mathcal{X}$. Then, $\exists$ hilbert space $\mathbb{H}$ (RKHS) and a function $\Phi: \mathcal{X} \to \mathbb{H}$ s.t.

① $k(x, x') = \langle \Phi(x), \Phi(x') \rangle_{\mathbb{H}}$

② $\mathbb{H} \subseteq [\mathcal{X} \to \mathbb{R}]$

③ $\Phi: \mathcal{X} \to \mathbb{H}$

$\Phi(x) = k(x, \cdot)$

④ [Reproducing Property] $\forall h \in \mathbb{H}, \ x \in \mathcal{X}$

$h(x) = \langle h, \Phi(x) \rangle_{\mathbb{H}} = \langle h, k(x, \cdot) \rangle$

Note

$D \in Pr(\mathcal{X})$.

$k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$\Phi': Pr(\mathcal{X}) \to \mathbb{H}$

$\Phi'(D) \in \mathbb{H} \subseteq [\mathcal{X} \to \mathbb{R}]$ GAN.

$\Phi'(D)(x) = \mathbb{E}_{y \sim D}[k(x, y)]$

↳ kernel mean embedding

$\underbrace{d(D, D')}_{\hookrightarrow \text{mmd}} = \| \Phi'(D) - \Phi'(D') \|_{\mathbb{H}} = \sqrt{\langle \quad , \quad \rangle_{\mathbb{H}}}$.

(pf of thm)

$\mathbb{H}_0 = \text{span}\{k(x, \cdot) \mid x \in \mathcal{X}\} = \{\sum_{i=1}^{m} a_i k(x_i, \cdot) \mid m \in \mathbb{N}, a_i \in \mathbb{R}, x_i \in \mathcal{X}\}$.

We proved $\mathbb{H}_0$ has well-defined symmetric operation $\langle \cdot, \cdot \rangle_{\mathbb{H}}$:

$\langle \sum_{i=1}^{m} a_i k(x_i, \cdot), \sum_{j=1}^{n} b_j k(x_j, \cdot) \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j k(x_i, x_j)$.

Bilinear: $\langle c_1 h_1 + c_2 h_2, h \rangle = c_1 \langle h_1, h \rangle + c_2 \langle h_2, h \rangle$ by definition.

It remains to check

$\begin{cases} \forall h \in \mathbb{H}_0, \ \langle h, h \rangle \geq 0. & (1) \\ \langle h, h \rangle = 0 \Rightarrow h = 0 & (2) \end{cases}$

/ /

① Let $h = \sum_{i=1}^{m} a_i k(x_i, \cdot)$. Then,

$\langle h, h \rangle = \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j k(x_i, x_j) \geq 0$ as k is PSD.

② Before proving $[\langle h, h \rangle = 0 \Rightarrow h = 0]$, we prove reproducing property (RP) of $\mathbb{H}_0$.

Let $h = \sum_{i=1}^{m} a_i k(x_i, \cdot) \in \mathbb{H}_0$, $x \in \mathcal{X}$.

$h(x) = \sum_{i=1}^{m} a_i k(x_i, x) = \langle \sum a_i k(x_i, \cdot), k(x, \cdot) \rangle$

$= \langle h, \Phi(x) \rangle$

③ CS-ineq. holds $\forall h, h' \in \mathbb{H}_0$: $(\langle h, h' \rangle)^2 \leq \langle h, h \rangle \langle h', h' \rangle$.

③-1 : $\langle \cdot, \cdot \rangle : \mathbb{H}_0 \times \mathbb{H}_0 \to \mathbb{R}$ is PDS kernel.

$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \langle h_i, h_j \rangle = \langle \sum_{i=1}^{m} a_i h_i, \sum_{j=1}^{m} a_j h_j \rangle \geq 0$.

Null

③-2 : For any PDS kernel $k' : Z \times Z \to \mathbb{R}$, CS holds.

$K' = \begin{bmatrix} k'(x,x) & k'(x,y) \\ k'(x,y) & k'(y,y) \end{bmatrix}$, $\underbrace{\forall \text{ eigenvalue} \geq 0}_{\text{As } k' \text{ is PDS}} \Rightarrow \det(K') \geq 0$.

④ To prove (2), assume $\langle h, h \rangle = 0$. For $x \in \mathcal{X}$,

$h(x)^2 \underset{\text{RP}}{=} \langle h, k(x, \cdot) \rangle^2 \underset{\text{CS}}{\leq} \langle h, h \rangle \langle k(x, \cdot), k(x, \cdot) \rangle = 0$.

Thus, $h = 0$.

§2 Construction on Kernels

(1) Given a PDS kernel k, construct normalized kernel k':

$$k'(x,y) = \begin{cases} \dfrac{k(x,y)}{\sqrt{k(x,x)}\sqrt{k(y,y)}} & \text{if } k(x,x) \neq 0,\ k(y,y) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Why PDS?

① Symmetry easy

② PD

By RKHS thm, $k(x,y) = \langle \Phi(x), \Phi(y) \rangle_{\mathbb{H}}$.

Let $\Phi'(x) = \begin{cases} 0 & \text{if } \Phi(x) = 0 \\ \frac{\Phi(x)}{\|\Phi(x)\|} & \text{otherwise.} \end{cases}$

Then, $\|\Phi(x)\|^2 = k(x,x) \Rightarrow \Phi(x) = \sqrt{k(x,x)}$.

It follows that $k'(x,y) = \langle \Phi'(x), \Phi'(y) \rangle$.

Thus, k' is PDS.

Example

$$k(x,y) = \exp\left(\frac{\langle x,y \rangle}{\sigma^2}\right)$$

$$k'(x,y) = e^{\frac{\langle x,y \rangle}{\sigma^2} - \frac{\langle x,x \rangle}{2\sigma^2} - \frac{\langle y,y \rangle}{2\sigma^2}}$$

$$= e^{-\frac{1}{2\sigma^2}\|x-y\|^2}.$$

Why $k(x,y) = \exp\left(\frac{\langle x,y \rangle}{\sigma^2}\right)$ is a PDS kernel?

(2) Operation Producing New Kernel.

k_0, k_1 : PDS kernel on $\mathcal{X}$.

$k_+(x,y) = k_0(x,y) + k_1(x,y)$ is PDS kernel

$k_\times(x,y) = k_0(x,y) \times k_1(x,y)$ is PDS kernel.

Symmetry is easy.

PD? $\sum_{i,j=1}^{m} a_i a_j k_+(x_i, x_j) = \sum_{i,j=1}^{m} a_i a_j k_0(x_i, x_j) + a_i a_j k_1(x_i, x_j) \geq 0$.

$K_0 \in \mathbb{R}^{m \times m}$, $(K_0)_{ij} = k_0(x_i, x_j)$, $K_0 = U \Sigma U^T$ for Σ: diagonal, with nonneg. entries

U: orthonormal.

$(K_0)^{1/2} = U \Sigma^{1/2} U^T \Rightarrow (K_0)^{1/2} (K_0)^{1/2} = K_0$.

$\sum_{i,j=1}^{m} a_i a_j k_\times(x_i, x_j) = \sum a_i a_j k_0(x_i, x_j) k_1(x_i, x_j)$.

/ /

$$= \sum a_i a_j (K_0)_{ij}\, k_1(x_i, x_j)$$
$$= \sum a_i a_j (K_0^{\frac{1}{2}} K_0^{\frac{1}{2}})_{ij}\, k_1(x_i, x_j)$$
$$= \sum_{i,j} \sum_{l} a_i a_j (K_0^{\frac{1}{2}})_{il} (K_0^{\frac{1}{2}})_{lj}\, k_1(x_i, x_j)$$
$$= \sum_{l} \underbrace{\Big(\sum_{i,j} (a_i K_0^{\frac{1}{2}})_{il} \, (a_j K_0^{\frac{1}{2}})_{lj} \Big) k_1(x_i, x_j)}_{\geq 0 \ \forall l}.$$

≥ 0.

(3) Limit

$k, k_0, k_1, k_2, \cdots : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ s.t. $\forall x, x' \in \mathcal{X}$,
$k_i(x, x') \longrightarrow k(x, x')$ as $i \longrightarrow \infty$.
If k_i's are PDS, then k also.

$$k(x, x') = \lim_i k_i(x, x') = \lim_i k_i(x', x) = k(x', x)$$

$m \in \mathbb{N}$, $c_1, \cdots, c_m \in \mathbb{R}$, $x_1, \cdots, x_m \in \mathcal{X}$.

$$\sum_{i,j} c_i c_j k(x_i, x_j) = \sum_{i,j} c_i c_j \lim_n k_n(x_i, x_j)$$
$$= \lim_n \underbrace{\sum_{i,j} c_i c_j k_n(x_i, x_j)}_{\geq 0 \ \forall n}$$

(4) Power Series

$\sum a_n x^n$, $a_n \geq 0$. radius of convergence $(-\rho, \rho)$.
$k(x, x') \in (-\rho, \rho)$ $\forall x, x' \in \mathcal{X}$, k : PDS
Then, $k'(x, x') = \sum a_n k(x, x')^n$ is also PDS.

$\because$) $a_n k(x, x')$ is PDS $\forall n$.

$\sum_{n=1}^{m} a_n k(x, x')$ is PDS $\forall m$

$\lim_{m \to \infty} \sum_{n=1}^{m} a_n k(x, x')$ is PDS by (3)

Exercise	$k(x,x') = \exp\left(-\frac{1}{2\sigma^2}\lVert x-x'\rVert^2\right)$ is PDS.

k is a normalization of $k'(x,x') = e^{\frac{1}{\sigma^2}\langle x,x'\rangle}$

$$k'(x,x') = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{1}{\sigma^2}\langle x,x'\rangle\right)^n$$
$$= \sum_{n=0}^{\infty} \frac{1}{n!\,\sigma^{2n}}\langle x,x'\rangle^n.$$

$$a_n = \frac{1}{n!\,\sigma^{2n}}, \quad k''(x,x') = \langle x,x'\rangle \text{ is PDS.}$$

By example (4), k' is PDS

$\Rightarrow k$ is PDS.

(5) Empirical Kernel Map and empirically induced kernel.

k : PDS kernel. $k(x,x') = \langle \Phi(x), \Phi(x')\rangle_{RKHS_k}$

$x_1, \cdots, x_m \in \mathcal{X}$.

Goal : $k'(x,x') = \langle \Phi'(x), \Phi'(x')\rangle_{\mathbb{R}^m}$

$\Phi' : \mathcal{X} \to \mathbb{R}^m$

such that $k'(x_i,x_j) = k(x_i,x_j) \quad \forall i,j \in [m]$

$\Phi'' : \mathcal{X} \longrightarrow \mathbb{R}^m$ (Empirical Kernel Map)

$\Phi''(x) = (k(x,x_1), \cdots, k(x,x_m))$

$k''(x,x') = \langle \Phi''(x), \Phi''(x')\rangle_{\mathbb{R}^m}$

$$\Rightarrow k''(x_i,x_j) = \langle \Phi''(x_i)\, \Phi''(x_j)\rangle_{\mathbb{R}^m}$$
$$= \sum_{l=1}^{m} k(x_i,x_l)\, k(x_l,x_j)$$

/ /

$$= \sum_{l=1}^{m} K_{il} K_{lj} = (K^2)_{ij} \neq K_{ij}.$$

Note: By Eigen-decomposition, $K = U\Sigma U^T$, U: orthonormal

$K^+ = U\Sigma^+ U^T$, with $\Sigma^+ = \begin{bmatrix} \sigma_1^{-1} & & & & \\ & \ddots & & & \\ & & \sigma_r^{-1} & & \\ & & & 0 & \\ & & & & \ddots \ 0 \end{bmatrix}$

$(K^+)^{1/2} = U(\Sigma^+)^{1/2} U^T.$

$$(K^+)^{1/2}(K^+)^{1/2} = K^+$$
$$KK^+K = K$$

$\Phi' : \mathcal{X} \to \mathbb{R}^m$, by $\Phi'(x) = (K^+)^{1/2}\Phi''(x)$

$$k'(x, x') = \langle \Phi'(x), \Phi'(x') \rangle_{\mathbb{R}^m}$$

$$\begin{aligned}
k'(x_i, x_j) &= \langle (K^+)^{1/2}\Phi''(x_i), (K^+)^{1/2}\Phi''(x_j) \rangle_{\mathbb{R}^m} \\
&= \Phi''(x_i)^T ((K^+)^{1/2})^T (K^+)^{1/2} \Phi''(x_j) \\
&= \Phi''(x_i)^T K^+ \Phi''(x_j) \\
&= (Ke_i)^T K^+ (Ke_j) \\
&= e_i^T K K^+ K e_j \\
&= e_i^T K e_j \\
&= K_{ij} = k(x_i, x_j)
\end{aligned}$$

§ Kernel Trick & Representer Theorem.

(1) $$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$$

s.t. $0 \le \alpha_i \le C$ for $\forall i \in [m]$ and $\sum \alpha_i y_i = 0$.

α_i^*: optimal solution

/ /

$$h(x) = \mathrm{sgn}\Big(\sum \alpha_i^* y_i \underbrace{\langle x_i, x\rangle}_{k(x_i,x)} + y_{i_0} - \sum \alpha_i^* y_i \underbrace{\langle x_i, x_{i_0}\rangle}_{k(x_i,x_{i_0})}\Big)$$

$0 < \alpha_{i_0}^* < C$

RKHS

(2) $k \Rightarrow \mathrm{RKHS}_k = \mathbb{H} \subseteq [\mathcal{X} \to \mathbb{R}]$

$\underset{h\in\mathbb{H}}{\mathrm{argmin}}\ F(h)$

$F(h) = G(\|h\|_{\mathbb{H}}) + L(h(x_1), \cdots, h(x_m))$.

Then, $\exists$ a minimizer h^* of the form

$h^*(x) = \sum_{i=1}^{m} c_i k(x_i, x)$.

Theorem

(Representer Theorem)

$k \cdots$ PDS kernel on $\mathcal{X}$.

$\mathbb{H} \cdots$ RKHS for k, $x_1, \cdots, x_m \in \mathcal{X}$.

$\underset{h\in\mathbb{H}}{\mathrm{argmin}}\ F(h)$, where $F(h) = G(\|h\|_{\mathbb{H}}) + L(h(x_1), \cdots, h(x_m))$

$G: \mathbb{R} \to \mathbb{R}$ is nondecreasing

If $\exists$ soln of the problem, then $\exists$ soln of the form

$\sum_{i=1}^{m} \alpha_i k(x_i, \cdot)$ for some $\alpha_i \in \mathbb{R}$.

If G is increasing, every soln of the problem has the form above.

(pf)

Let h be the soln of the optimization problem.

$\mathbb{H}_0 = \mathrm{span}\{k(x_i,\cdot) \mid i \in [m]\}$

$= \{\sum_{i=1}^{m} \alpha_i k(x_i, \cdot) \mid \alpha_i \in \mathbb{R}\}$

$h_0 = \mathrm{proj}_{\mathbb{H}_0} h$, $h = h_0 + h_0^{\perp}$

i.e. $h_0 \in \mathbb{H}_0$, $h_0^\perp \in \mathbb{H}_0^\perp$ with $h_0 + h_0^\perp = h$

Then, $\|h\|^2 = \|h_0\|^2 + \|h_0^\perp\|^2$. Hence, $G(\|h_0\|^2) \le G(\|h\|^2)$.

It suffices to prove $L(h_0(x_1), \cdots, h_0(x_m)) = L(h(x_1), \cdots, h(x_m))$.

By Reproducing property,
$$\begin{aligned} h(x_i) &= \langle h, k(x_i, \cdot) \rangle_{\mathbb{H}} \\ &= \langle h_0 + h_0^\perp, k(x_i, \cdot) \rangle \\ &= \langle h_0, k(x_i, \cdot) \rangle + \langle h_0^\perp, k(x_i, \cdot) \rangle \\ &= h_0(x_i) + 0 = h_0(x_i). \end{aligned}$$

Hence, we are done!

If $\|h_0^\perp\| \neq 0$, then $\|h_0\| < \|h\|$
$$\Rightarrow G(\|h_0\|^2) < G(\|h\|^2).$$

§3 Negative Definite Symmetric Kernel (NDS)

(1) **Motivation**

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$$

PDS kernel → Generalization of inner prod.

NDS kernel → " distance.

(2) **Definition**

$k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is NDS if

① $k(x, x') = k(x', x)$ $\forall x, x' \in \mathcal{X}$.

② $\forall m \in \mathbb{N}$, $\forall c_1, \cdots, c_m \in \mathbb{R}$ s.t. $\sum_{i=1}^m c_i = 0$, and $x_1, \cdots, x_n \in \mathcal{X}$,
$$\sum_{i,j=1}^m c_i c_j k(x_i, x_j) \le 0$$

/ /

Note

$*$ NDS is a generalization of distance squared.

$*$ NDS $k \Leftrightarrow \exp(-tk)$ is NDS for any $t>0$.

Example

$\mathcal{X}=\mathbb{R}^n$, $k(x,x')=\|x-x'\|^2$ is NDS.

① Symmetry trivial

② $\sum c_i=0$, $x_1,\cdots,x_m\in\mathbb{R}^m$.

$$\sum c_i c_j k(x_i,x_j)=\sum c_i c_j\|x_i-x_j\|^2 \qquad \langle x_i-x_j, x_i-x_j\rangle$$

$$=\sum c_i c_j\langle x_i-x_j, x_i-x_j\rangle$$

$$=\sum c_i c_j\left(\langle x_i,x_i\rangle-2\langle x_i,x_j\rangle+\langle x_j,x_j\rangle\right)$$

$$=\sum_{i=1}^{n} c_i\langle x_i,x_i\rangle\cancel{\left(\sum_{j=1}^{m} c_j\right)}^{0}$$

$$-2\sum_{i,j=1}^{m}\langle x_i,x_j\rangle c_i c_j$$

$$+\sum_{j=1}^{m} c_j\langle x_j,x_j\rangle\cancel{\left(\sum_{i=1}^{m} c_i\right)}^{0}$$

$$=-2\left\langle\sum_i c_i x_i,\sum_j c_j x_j\right\rangle$$

$$=-2\left\|\sum_i c_i x_i\right\|^2\le 0.$$

!

Lemma

$k:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ symmetric, $x_0\in\mathcal{X}$

$k':\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ by $k'(x,x')=k(x,x_0)+k(x',x_0)-k(x,x')-k(x_0,x_0)$

Then, ① k' is symmetric.

② k is NDS $\Leftrightarrow k'$ is PDS.

(pf)

① Trivial

② $(\Leftarrow)$ k: PDS, $\sum_{i=1}^{m} c_i=0$, $x_1,\cdots,x_m\in\mathbb{R}$.

$k(x,x')=k(x,x_0)+k(x',x_0)-k'(x,x')-k(x_0,x_0)$.

$$\sum_{i,j} c_i c_j k(x_i,x_j)=\sum_{i,j} c_i c_j\left(k(x_i,x_0)+k(x_j,x_0)-k'(x_i,x_j)-k(x_0,x_0)\right)$$

$$=-\sum_{i,j} c_i c_j k'(x_i,x_j)\ge 0.$$

$\Rightarrow \sum c_i c_j k(x_i,x_j)\le 0$

/ /

$(\Rightarrow)$ k: NDS, $c_1, \cdots, c_m \in \mathbb{R}$, $x_1, \cdots, x_m \in \mathcal{X}$.

$c_0 = -\sum_{i=1}^{m} c_i$, $x_0 \in \mathcal{X}$.

$$\sum_{i,j=0}^{m} c_i c_j k(x_i, x_j) = \sum_{i,j=0}^{m} c_i c_j \left(k(x_i, x_0) + k(x_j, x_0) - k'(x_i, x_j) - k(x_0, x_0) \right)$$

$$= -\sum_{i,j=0}^{m} c_i c_j k'(x_i, x_j) = -\sum_{i,j=1}^{m} c_i c_j k'(x_i, x_j) \le 0.$$

$k'(x_0, x') = k'(x, x_0) = 0$, by def.

Hence, k' is PDS.

Theorem 6.17 NDS $k \Leftrightarrow \exp(-tk)$ for $t>0$ is PDS $\forall t>0$.

(pf) $(\Rightarrow)$ $x_0 \in \mathcal{X}$. k' as in lemma.

k' is PDS $\Rightarrow$ tk' is PDS for $t>0$

$\Rightarrow k'' = \exp(tk')$ is PDS

$$\| \quad \frac{\exp(-tk(x,x'))\ \exp(-tk(x_0,x_0))}{\exp(-tk(x,x_0))\ \exp(-tk(x',x_0))}$$

$$k'''(x,x') = \frac{\langle \exp(-tk(x,x_0)),\ \exp(-tk(x',x_0)) \rangle}{\exp(-tk(x_0,x_0))} \text{ is PDS. (check def.!)}$$

$k''(x,x')k'''(x,x') = \exp(-tk(x,x'))$ is PDS for being multiplication of PDS.

$(\Leftarrow)$ Pick $c_1, \cdots, c_m \in \mathbb{R}$ s.t. $\sum_{i=1}^{m} c_i = 0$, $x_1, \cdots, x_m \in \mathcal{X}$.

$\exp(-kt)$ is PDS

$$\sum_{i,j=1}^{m} c_i c_j \underbrace{\left(\exp(-tk(x_i, x_j)) \right)}_{1 - tk(x_i,x_j) + O(t^2)} \ge 0.$$

$$\Big(\cancel{\sum_{i,j=1}^{m} c_i c_j}^{\,0}\Big) - t \sum_{i,j=1}^{m} c_i c_j k(x_i, x_j) + O(t^2) \ge 0.$$

By $\epsilon \to 0$, we see $\sum_{i,j=1}^{m} c_i c_j k(x_i, x_j) \le 0$.
i.e. k is NDS.

Theorem $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$: NDS kernel. $k(x, x') = 0 \Leftrightarrow x = x'$. Then,

$\exists$ a Hilbert space $\mathbb{H}$ and a map $\Phi : \mathcal{X} \to \mathbb{H}$ such that
$k(x, x') = \|\Phi(x) - \Phi(x')\|_{\mathbb{H}}^2 = \langle \Phi(x) - \Phi(x'), \Phi(x) - \Phi(x') \rangle_{\mathbb{H}}$

(pf) Let $x_0 \in \mathcal{X}$. $k'(x, x') = k(x, x_0) + k(x', x_0)$
$\qquad - k(x, x') - k(x_0, x_0)$.

k' is PDS. $\mathbb{H}'$: RKHS
$\Phi' : \mathcal{X} \to \mathbb{H}'$: Associated Feature map
$\Phi'(x) = k(\cdot, x)$.

Let $\mathbb{H} = \mathbb{H}'$, $\Phi(x) = \Phi'(x)/\sqrt{2}$.

$$
\begin{aligned}
\|\Phi(x) - \Phi(x')\|_{\mathbb{H}}^2 &= \frac{1}{2} \langle \Phi'(x) - \Phi'(x'), \Phi(x) - \Phi'(x') \rangle_{\mathbb{H}} \\
&= \frac{1}{2} \left(k'(x, x) - 2k'(x, x') + k'(x', x') \right) \\
&= \frac{1}{2} \Big(\cancel{2k(x, x_0)} - \cancel{k(x, x)} - \cancel{k(x_0, x_0)} \\
&\qquad - 2\left(\cancel{k(x, x_0)} + \cancel{k(x', x_0)} - k(x, x') - \cancel{k(x_0, x_0)} \right) \\
&\qquad + k\left(\cancel{2k(x', x_0)} - \cancel{k(x', x')} - \cancel{k(x_0, x_0)} \right) \Big) \\
&= k(x, x')
\end{aligned}
$$

§ Application of a PDS kernel k via Bochner's Theorem

(1) $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
$\Phi : \mathcal{X} \to \mathbb{R}^p$ $\qquad$ m: nbr of data points.
$(p \ll m)$

/ /

$k(x,x') \approx \langle \Phi(x), \Phi(x') \rangle$.

$x \longmapsto \mathrm{sgn}\langle w, \Phi(x) \rangle$.

Theorem $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ PDS kernel.

$k(x,x') = G(x-x')$ s.t. $G(0)=1$ and G is continuous.

($\Rightarrow k$ is translation invariant, i.e. $k(x+t, x'+t) = k(x,x')$).

Then, $\exists$ a probability distribution $p \in Pr(\mathcal{X})$ s.t.

$$k(x,x') = \int_{\mathcal{X}} e^{i\langle w, x-x' \rangle} p(dw)$$

Q1 What does it mean

Q2. How can we use it?

Example $x, x' \in \mathbb{R}^n$,

(Gaussian Kernel) $k(x,x') = e^{-\frac{\|x-x'\|^2}{2}}$. $p(x) = \frac{1}{(2\pi)^{n/2}} e^{\left(-\frac{\|x\|^2}{2}\right)}$

(Laplacian Kernel) $k(x,x') = e^{-\|x-x'\|_1^2}$, $p(x) = \prod_{i=1}^{n} \frac{1}{\pi(1+x_i^2)}$

$k(x,x') = \prod_{i=1}^{n} \frac{1}{1+(x_i-x_i')^2}$, $p(x) = e^{-\|x\|_1^2}$.

By thm $k(x,x') = \mathbb{E}_{w \sim p(w)} \left[\cos\langle w, x-x' \rangle \right]$

$\cos\langle w, x-x' \rangle = \cos(\langle w, x \rangle + \langle w, x' \rangle)$

$= \cos\langle w, x \rangle \cos\langle w, x' \rangle + \sin\langle w, x \rangle \sin\langle w, x' \rangle$

$= \langle (\cos\langle w, x \rangle, \sin\langle w, x \rangle), (\cos\langle w, x' \rangle, \sin\langle w, x' \rangle) \rangle$

/ /

$$= \langle \Phi_w(x), \Phi_w(x') \rangle.$$

$$w_1, \dots, w_p \overset{\text{iid}}{\sim} p(w)$$

$$x \mapsto \left(\frac{1}{\sqrt{p}}\Phi_{w_1}(x), \frac{1}{\sqrt{p}}\Phi_{w_2}(x), \dots, \frac{1}{\sqrt{p}}\Phi_{w_p}(x)\right) = \Phi_{w_1,\dots,w_p}(x)$$

$$K'_{w_1,\dots,w_p}(x,x') = \langle \Phi_{w_1,\dots,w_p}(x), \Phi_{w_1,\dots,w_p}(x') \rangle$$
$$= \frac{1}{p}\sum_{i=1}^{p} \langle \Phi_{w_i}(x), \Phi_{w_i}(x') \rangle$$

$$\mathbb{E}_{w_1,\dots,w_p}[K'_{w_1,\dots,w_p}(x,x')] = K(x,x')$$

$$\mathbb{V}[\qquad\qquad] = \frac{1}{p}\mathbb{V}_{w\sim p(w)}[K'_w(x,x')] \longrightarrow 0 \text{ as } p \to \infty.$$