RKHS Notes

Justin - Ds.

PDS Kernels: $K: \mathcal{X}\times\mathcal{X} \to \mathbb{R}$

if $\forall \{x_1, \cdots, x_n\} \subseteq \mathcal{X}$,

the matrix $K = (K(x_i, x_j))_{m\times m}$ is symmetric and positive semi-definite

$$K = \begin{pmatrix} K(x_1,x_1) & K(x_1,x_2) & \cdots & K(x_1,x_m) \\ K(x_2,x_1) & K(x_2,x_2) & \cdots & K(x_2,x_m) \\ \vdots & \vdots & & \vdots \\ K(x_m,x_1) & K(x_m,x_2) & \cdots & K(x_m,x_m) \end{pmatrix}$$

核矩阵 (kernel matrix / Gram matrix)

$\begin{cases} \text{对称} \\ \text{半正定} \begin{cases} \mu_i \geq 0,\ K\text{的特征值非负} \\ \forall c,\ c^T K c = \sum_i \sum_j c_i c_j K(x_i, x_j) \geq 0 \end{cases} \end{cases}$

(Mercer's Theorem)

$K: \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ continuous symmetric

$$K(x, x') = \sum_{n=0}^{\infty} a_n \phi_n(x) \phi_n(x') \quad \text{with } a_n > 0$$

iff $\forall$ square integrable function C, s.t. $\iint_{\mathcal{X}\times\mathcal{X}^2} c(x) c(x') K(x, x') dx dx' \geq 0$

K是核函数 $\Leftrightarrow$ K是PDS核 (正定核)

$\Leftrightarrow$ K是 SPSD矩阵 (半正定矩阵)

Examples:

① Polynomial Kernel: $\forall\ x, x' \in \mathbb{R}^N$, $\cancel{K(x,x')} = (x \cdot x' + c)^d$ (PDS核)

$\Phi(\cdot) \in \mathbb{H} \in \mathbb{R}^{\binom{N+d}{d}}$

e.g. $N=2$, $d=2$, $K(x, x') = (x_1 x_1' + x_2 x_2' + c)^2$

$$= x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2 x_1 x_1' x_2 x_2' + 2c x_1 x_1' + 2c x_2 x_2' + c^2$$

$$= \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2} x_1 x_2 \\ \sqrt{2c}\, x_1 \\ \sqrt{2c}\, x_2 \\ c \end{pmatrix} \cdot \begin{pmatrix} \text{the} \\ \text{same} \\ \text{right} \end{pmatrix} \qquad C_4^2 = 6$$

② Gaussian kernel (RBF)

$$K(x,x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \qquad \sum_{n=0}^{\infty} \frac{(x\cdot x')^n}{n!\,\sigma^{2n}} = \exp\left(\frac{x\cdot x'}{\sigma^2}\right)$$

③ Laplace Kernel

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

(L_1-正则化)

④ Sigmoid kernel

$$K(x,x') = \tanh(a(x\cdot x')+b)$$

Lemma: (Cauthy-Schwarz inequality for PDS kernel)

$$\forall x, x' \in \mathcal{X},\quad K(x,x')^2 \leq K(x,x)\cdot K(x',x')$$

Proof: Kernel matrix $\mathbb{K} = \begin{pmatrix} K(x,x) & K(x,x') \\ K(x',x) & K(x',x') \end{pmatrix}$ $\quad K: PDS \Rightarrow \mathbb{K}: SPSD$

eigenvalue $\lambda_1, \lambda_2 \geq 0$ $\quad \Rightarrow \lambda_1 \cdot \lambda_2 = \underline{\det(\mathbb{K}) \geq 0}$ $\quad \Rightarrow \det(K) = \_\_\_ - \_\_\_ \geq 0$ □

Thm $\boxed{K: \mathcal{X}\times\mathcal{X} \to \mathbb{R} \text{ is PDS kernel}} \Rightarrow \exists$ Hilbert Space $\mathbb{H}$, mapping $\Phi: \mathcal{X} \to \mathbb{H}$

s.t. $\forall x, x' \in \mathcal{X},\ K(x,x') = \langle \Phi(x), \Phi(x') \rangle$

Reproducing Kernel Hilbert Space

(RKHS) 再生核

$\mathcal{H}$ follows "reproducing property"

$$\forall h \in \mathbb{H},\ \forall x \in \mathcal{X},\ \underset{\text{值}}{h(x)} = \langle \underset{\text{函数}}{h}, \underset{\text{参数}}{K(x,\cdot)} \rangle$$

PDS核 ⇒ RKHS

(正定对称)

Proof: 证明 Hilbert Space
- 内积空间: $\langle \cdot, \cdot \rangle$ (正定, 对称, 线性)
- 完备 (Complete): all cauchy sequence converges

$\forall x \in \mathcal{X}$, $\underline{\Phi(x)}: \mathcal{X} \to \mathbb{R}^{\mathcal{X}}$, $\forall x' \in \mathcal{X}$, $\underline{\Phi(x)}(x') = K(x, x')$

映射　　　函数(复合)

$\mathbb{H}_0 = \left\{ \sum_{i \in I} a_i \Phi(x_i) : a_i \in \mathbb{R}, x_i \in \mathcal{X}, |I| < \infty \right\}$ (linear combination of all functions $\Phi(x)$)

$\Rightarrow$ 证明 $\mathbb{H}_0$ 是内积空间

Define inner product

$f = \sum_{i \in I} a_i \Phi(x_i)$, $g = \sum_{j \in J} b_j \Phi(x_j)$ $\Rightarrow$ $\underbrace{\langle f, g \rangle = \sum_{i \in I} \sum_{j \in J} a_i b_j K(x_i, x_j')}_{\text{内积定义}} = f(x)$

$= \sum_{j \in J} b_j \cdot \underbrace{\sum_{i \in I} a_i K(x_i, x_j')}_{\text{确定}}$

$= \sum_{j \in J} b_j \cdot \underbrace{\sum_{i \in I} a_i \Phi(x_i)}_{f}(x_j')$

$= \sum_{j \in J} b_j \cdot f(x_j')$ $\left(= \sum_{i \in I} a_i\, g(x_i) \right)$

(验证内积：正定，对称，线性)

① definiteness: K PDS

$\langle f, f \rangle = \sum_{i,j \in I} a_i a_j K(x_i, x_j) \geq 0$

② Symmetry: K PDS

③ bi-linearity $\langle \cdot, \cdot \rangle$:

$\langle f + f', g \rangle = \sum_{j \in J} b_j (f(x_j) + f'(x_j)) = \sum_{j \in J} b_j f(x_j') + \sum_{j \in J} b_j f'(x_j')$

$\forall f_1, \cdots, f_m \in \mathbb{H}$, $c_1, \cdots, c_m \in \mathbb{R}$, $\sum_{i \in I} \sum_{j \in J} c_i c_j \langle f_i, f_j \rangle = \langle \sum c_i f_i, \sum c_j f_j \rangle \geq 0$

$\Rightarrow \langle \cdot, \cdot \rangle$ is an inner product $\Rightarrow$ $\mathbb{H}_0$ is an inner product space (pre-Hilbert)

Completion Theorem (完备化定理):

任意度量空间，都存在一个完备的度量空间 $\mathcal{H}$, s.t $\mathcal{H}_0$ 是 $\mathcal{H}$ 的一个稠密子空间

$\mathcal{H}_0 \xrightarrow{\text{completion}} \mathcal{H}$: Hilbert space　　　$\mathcal{H}_0 \subseteq \mathcal{H}$

Reproducing property: $h = \sum_{i \in I} a_i \Phi(x_i) \in \mathcal{H}$

$\forall x \in \mathcal{X}$: $h(x) = \sum_{i \in I} a_i \Phi(x_i)(x) = \sum_i a_i K(x_i, x) = \langle h, \Phi(x) \rangle$

PDS Kernel K $\Rightarrow$ RKHS $\mathcal{H}$ $\begin{cases} \mathcal{H}: \text{feature space} \\ \Phi: \text{feature mapping} \end{cases}$

不需找 Φ, 只需找 K

不一定唯一

$\|\cdot\|_{\mathcal{H}} = \sqrt{\langle \cdot, \cdot \rangle_{\mathcal{H}}}$

Property of kernel:

① Normalization $K(x,x') \xrightarrow{\text{normalize}} \tilde{K}(x,x') = \begin{cases} 0, & \text{if } K(x,x)=0 \text{ or } K(x',x')=0 \\ \dfrac{K(x,x')}{\sqrt{K(x,x)\cdot K(x',x')}}, & \text{otherwise} \end{cases}$

$\forall x \in \mathcal{X}$, $K(x,x)=1$ (归一)

$$\exp\left[-\frac{\|x-x'\|^2}{2\sigma^2}\right] \longleftarrow \underbrace{\exp\left(\frac{x\cdot x'}{\sigma^2}\right)}_{K(x,x')} \qquad \frac{K(x,x')}{\sqrt{K(x,x)K(x',x')}} = \frac{e^{\frac{x\cdot x'}{\sigma^2}}}{e^{\frac{\|x\|^2}{2\sigma^2}}\cdot e^{\frac{\|x'\|^2}{2\sigma^2}}} = e^{\frac{-\|x-x'\|^2}{2\sigma^2}}$$

Proposition: K PDS kernel $\xrightarrow[\text{normalization}]{}$ $\tilde{K}$ PDS kernel

Proof: $\{x_1,\cdots,x_m\} \subseteq \mathcal{X}$, $c \in \mathbb{R}^m$

$$\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{m} c_i c_j \cdot \tilde{K}(x_i,x_j)}_{\text{PSD}} = \sum\sum c_i c_j \frac{K(x_i,x_j)}{\sqrt{K(x_i,x_i)K(x_j,x_j)}} \qquad (K>0)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{c_i c_j \langle \Phi(x_i), \Phi(x_j)\rangle}{\|\Phi(x_i)\|_{\mathcal{H}}\cdot\|\Phi(x_j)\|_{\mathcal{H}}} = \left\|\sum \frac{c_i \Phi(x_i)}{\|\Phi(x_i)\|_{\mathcal{H}}}\right\|_2$$

② def (emperical feature map), given $\{x_1,\cdots,x_m\} \subseteq \mathcal{X}$,

$$\Phi: \mathcal{X} \to \mathbb{R}^m, \ \forall x \in \mathcal{X}, \ \Phi(x) = \begin{bmatrix} K(x,x_1) \\ K(x,x_2) \\ \vdots \\ K(x,x_m) \end{bmatrix}$$

已知K,寻找Φ

"K-similarity" $(\Phi(x)(x_i) = x_i)$

$\Phi(x_i) = K\cdot e_i$ (核矩阵第 i 列)

$\langle \Phi(x_i), \Phi(x_j)\rangle = (Ke_i)^T\cdot(Ke_j)$

$= e_i^T K^2 e_j = (K^2)_{ij}$

$$K = \begin{pmatrix} K(x_1,x_1) & K(x_1,x_2) & \cdots & K(x_1,x_m) \\ K(x_2,x_1) & K(x_2,x_2) & \cdots & K(x_2,x_m) \\ \vdots & \vdots & & \vdots \\ K(x_m,x_1) & K(x_m,x_2) & \cdots & K(x_m,x_m) \end{pmatrix}$$

核矩阵 (kernel matrix / Gram matrix)

$\Rightarrow$ the kernel matrix associated with Φ is K^2

Let $(K^-)^{\frac{1}{2}}$, where K^- is pseudo-inverse of K

↘ by SVD of K^-

K invertible: $K^- = K^{-1}$

K non-inverse: $KK^- = I$

$K^- = (K^TK)^{-1}K^T$

define $\psi(x) = (K^{-})^{\frac{1}{2}} \Phi(x)$

$\Rightarrow \langle \psi(x_i), \psi(x_j) \rangle = ((K^{-})^{\frac{1}{2}} K e_i)^T ((K^{-})^{\frac{1}{2}} \cdot K e_j)$

$= e_i \cdot K \cdot (K^{-})^{\frac{1}{2}} \cdot (K^{-})^{\frac{1}{2}} \cdot K \cdot e_j$

$= K_{ij}$

$\Rightarrow$ the kernel matrix associate with ψ is K

define $\Omega(x) = K^{-} \Phi(x)$, $\forall x \in X$

$\langle \Omega(x_i), \Omega(x_j) \rangle = e_i^T K^T (K^{-}) \cdot (K^{-}) \cdot K \cdot e_j = e_i^T e_j = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$

$\Rightarrow$ the kernel matrix associated with Ω is I

③ Closure under: 求和, 求积, 直积   极限   复合   幂级数

闭包性质   sum   product   Cartesian product   pointwise limit $(K_n)_{n \geq 1}$ power series

proof: e.g. $\left.\begin{array}{l} c^T K c \geq 0 \\ c^T K' c \geq 0 \end{array}\right\} \Rightarrow c^T (K + K') c \geq 0$ $\square$