

カーネル法と再生核ヒルベルト空間

カーネル法による教師あり学習:

ある再生核ヒルベルト空間 \mathcal{H} を用いて

$$\hat{f} = \operatorname{argmin}_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^n l(y_i, f(x_i)) + \lambda \|f\|_{\mathcal{H}}^2$$

ただし、 l は損失関数、 $\|\cdot\|_{\mathcal{H}}$ は \mathcal{H} における f のノルム。

定義 (正定値カーネル)

$k: X \times X \rightarrow \mathbb{R}$ n 次元実対称行列、 k は正定値カーネルと言う。

・ 対称性: $k(x, x') = k(x', x)$ ($\forall x, x' \in X$)

・ 正値性: $\sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j k(x_i, x_j) \geq 0$
($\forall \alpha_i \in \mathbb{R}, \forall x_i \in X, m=1, 2, \dots$)

例: (2-次元空間の内積)

$$x, y \in \mathbb{R}^d \text{ に対し } k(x, y) = x^T y = \sum_{i=1}^d x_i y_i \text{ とする。}$$

$\Rightarrow k(x, y)$ は正定値カーネル。



定義 (再生核ヒルベルト空間, Reproducing Kernel Hilbert Space, RKHS)

集合 X 上の RKHS とは、 X 上の (実数値) 関数からなるヒルベルト空間 \mathcal{H} で、 $\forall x \in X$ に対し $k_x \in \mathcal{H}$ が存在して

$$(\text{再生性}) \quad \langle f, k_x \rangle_{\mathcal{H}} = f(x) \quad (\forall f \in \mathcal{H})$$

が成り立つものとする。

特に、 $k(x, y) = \langle k_x, k_y \rangle_{\mathcal{H}}$ は正定値カーネルである。

\mathcal{H} に付随した再生核 (Reproducing kernel) とする。

定義より、RKHS \mathcal{H} に対しカーネル k が定まる。

逆に正定値カーネル k に対し、RKHS \mathcal{H} が一意に定まる。

$$k \longleftrightarrow \mathcal{H} \\ (\text{一対一})$$

定理 (Moore - Aronszajn)

$k: X \times X \rightarrow \mathbb{R}$ を正定値カーネルとする。

$\Rightarrow k$ を再性核とする RKHS が一意的に存在。 //

証明

$$H_0 := \text{span} \{ k(x, \cdot) \mid x \in X \}$$

$$= \left\{ \sum_{i=1}^m \alpha_i k(x_i, \cdot) \mid \alpha_i \in \mathbb{R}, x_i \in X, m=1, 2, \dots \right\}$$

∴ H_0 に "内積" $\langle \cdot, \cdot \rangle_{H_0}$ を次のように定める:

$$f(x) = \sum_{i=1}^n \alpha_i k(x_i, x), \quad g(x) = \sum_{j=1}^m \beta_j k(y_j, x) \in H_0 \text{ に対して}$$

$$\langle f, g \rangle_{H_0} = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j k(x_i, y_j).$$

これは f, g の表示の \mathbb{R} 係数に依存して well-defined: $f = \sum_{i=1}^n \alpha_i k(x_i, \cdot)$ と表現する時

$$\begin{aligned} \langle f, g \rangle_{H_0} &= \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j k(x_i, y_j) = \sum_{j=1}^m \beta_j f(y_j) \quad \leftarrow \text{等しい} \\ &= \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j k(x_i, y_j) = \sum_{i=1}^n \alpha_i f(x_i) \quad \leftarrow \text{等しい} \end{aligned}$$

∴ 実際には内積 $\langle \cdot, \cdot \rangle_{H_0}$ が一意に定まる。

- 線形性: $\langle af + bg, h \rangle_{H_0} = a \langle f, h \rangle_{H_0} + b \langle g, h \rangle_{H_0} \rightarrow$ 自明
- 対称性: $\langle f, g \rangle_{H_0} = \langle g, f \rangle_{H_0} \rightarrow$ 自明
- 正定性, 非退化性: $\langle f, f \rangle_{H_0} \geq 0$ かつ $\langle f, f \rangle_{H_0} = 0 \Rightarrow f = 0$

また $\langle f, f \rangle_{H_0} \geq 0$ は k の正定値性から確認できる。

∴ $\langle \cdot, \cdot \rangle_{H_0}$ は Cauchy-Schwarz の不等式が成り立つ。

実際、 $\forall f, g \in H_0$ に対して

$$(a, b \in \mathbb{R}, \|f\|_{H_0}^2 := \langle f, f \rangle_{H_0})$$

$$0 \leq \langle af - bg, af - bg \rangle_{H_0} = a^2 \|f\|_{H_0}^2 - 2ab \langle f, g \rangle_{H_0} + b^2 \|g\|_{H_0}^2$$

$$\text{より } 2ab \langle f, g \rangle_{H_0} \leq a^2 \|f\|_{H_0}^2 + b^2 \|g\|_{H_0}^2$$

$$\|f\|_{H_0}^2, \|g\|_{H_0}^2 \geq 0 \text{ 故に } a = \frac{1}{\|f\|_{H_0}}, b = \frac{1}{\|g\|_{H_0}} \text{ とすれば}$$

$$\langle f, g \rangle_{H_0} \leq \|f\|_{H_0} \|g\|_{H_0}$$

∴ $\|f\|_{H_0} = 0$ かつ $\|g\|_{H_0} = 0$ のときは $\langle f, g \rangle_{H_0} = 0$

が成り立つ。

∴ $\langle f, f \rangle_{H_0} = 0$ ならば $\forall x \in X$ に対して $f(x) = \langle f, k(x, \cdot) \rangle_{H_0} = 0$ となる。

Cauchy-Schwarz の場合

$$|f(x)| = |\langle f, k(x, \cdot) \rangle_{H_0}| \leq \|f\|_{H_0} \cdot \sqrt{k(x, x)} = 0$$

∴ $f(x) = 0$ ($\forall x \in X$)。つまり $f = 0$ (関数 \mathbb{R}) となる。

$\langle \cdot, \cdot \rangle_{\mathcal{H}}$ の内積 \mathcal{H}_0 による \mathcal{H} の内積 $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$ を用いて \mathcal{H} のノルム $\| \cdot \|_{\mathcal{H}}$ を定義する。このノルムによる \mathcal{H} の完備化 \mathcal{H}_0 を \mathcal{H} の完備化と呼ぶ。

完備化の定義より、 \mathcal{H} の内積は次のように定義する：

$f, g \in \mathcal{H}$ に対し、 \mathcal{H}_0 内のコーシー列 $(f_n)_n, (g_n)_n$ が存在して、

$$\langle f, g \rangle_{\mathcal{H}} := \lim_{n \rightarrow \infty} \langle f_n, g_n \rangle_{\mathcal{H}_0} \text{ と定義される。}$$

これはコーシー列の取り方によらない。

$f \in \mathcal{H}$ に対し、 f に対応するコーシー列 $(f_m)_m \in \mathcal{H}_0$ を用いて

$$f(x) = \lim_{m \rightarrow \infty} \langle f_m, k(x, \cdot) \rangle_{\mathcal{H}_0}$$

とすれば、 \mathcal{H} の各元 $f \in \mathcal{H}$ に対し実数値関数 $x \mapsto f(x)$ が定義される。

これはコーシー列の取り方によらない。実際、 $(f_m)_m, (g_m)_m \in \mathcal{H}_0$ が

それぞれ f に対応するときは

(コーシー列)

$$\begin{aligned} & | \langle f_m, k(\cdot, x) \rangle_{\mathcal{H}_0} - \langle g_m, k(\cdot, x) \rangle_{\mathcal{H}_0} | \\ & \leq \| k(\cdot, x) \|_{\mathcal{H}_0} \| f_m - g_m \|_{\mathcal{H}_0} \rightarrow 0 \text{ である。} \end{aligned}$$

$\mathcal{H} \ni f \mapsto (x \mapsto f(x)) \in \mathbb{R}^X$ の対応は単射である。これは空間 \mathcal{H} と \mathbb{R}^X の間の対応 $f \mapsto (f(x))$ を示す。

集合 $\mathcal{H} \subset \mathbb{R}^X$ と見なせる。(ただし \mathcal{H} は実数値関数のヒルベルト空間) である。これは $f(x) = 0 \ (\forall x \in X) \Rightarrow \| f \|_{\mathcal{H}} = 0$ であることを示す。

f に対応するコーシー列 $(f_n)_n$ を取れる。これはコーシー列なので、

$\exists A \geq \sup_n \| f_n \|_{\mathcal{H}_0}$ である。十分大きな N を持つとき、

$$\forall n \geq N \text{ とき } \| f_n - f_N \|_{\mathcal{H}_0} \leq \frac{\varepsilon}{A} \text{ とできる (任意性)。} \text{ したがって } \forall n \geq N \text{ とき } \| f_n - f_N \|_{\mathcal{H}_0} \leq \frac{\varepsilon}{A} \text{ とできる。}$$

$$\begin{aligned} \| f_n - f_N \|_{\mathcal{H}_0}^2 &= \langle f_n - f_N, f_n - f_N \rangle_{\mathcal{H}_0} \\ &= \langle f_n - f_N, f_n \rangle_{\mathcal{H}_0} + \langle f_N, f_n - f_N \rangle_{\mathcal{H}_0} \\ &\leq \| f_n - f_N \|_{\mathcal{H}_0} \| f_n \|_{\mathcal{H}_0} + \left\langle \sum_{j=1}^m d_j k(x_j, \cdot), f_n - f_N \right\rangle_{\mathcal{H}_0} \\ &\leq \frac{\varepsilon}{A} \cdot A + \sum_{j=1}^m d_j \underbrace{\langle f_n(x_j), f_n - f_N \rangle_{\mathcal{H}_0}}_{f_N \text{ を用いて} } \end{aligned}$$

$$\text{すなわち } \limsup_n \| f_n - f_N \|_{\mathcal{H}_0}^2 \leq \varepsilon \text{ である。} \quad \downarrow \quad 0 \text{ (} n \rightarrow \infty \text{)}$$

$$\varepsilon \text{ は任意なので } \lim_{n \rightarrow \infty} \| f_n - f_N \|_{\mathcal{H}_0}^2 = 0 \Rightarrow \| f \|_{\mathcal{H}} = 0$$

よって、 \mathcal{H} は実数値関数のなすヒルベルト空間である。

よして、 H は再生核を持つことを示せばよい。

$\forall x \in X$ において、 $(k(x, \cdot))_m$ は Γ - \mathcal{B} 列で
 $k_x \in H$ を定義する。

$$\langle f, k_x \rangle_H = \lim_{m \rightarrow \infty} \langle f_m, k(x, \cdot) \rangle_{H_0} = f(x) \quad (\text{再生性})$$

\perp 完備化の定義

となり、 H は RKHS となる。

$$\text{再生核は } \langle k_x, k_y \rangle_H = \lim_{m \rightarrow \infty} \langle k(x, \cdot), k(y, \cdot) \rangle_{H_0} \\ = k(x, y)$$

より、 H は k を再生核として持つ。

(一意性)

H' を k を再生核として持つ別の RKHS とする。

再生核の定義より、 $\text{Span}\{k(x, \cdot) \mid x \in X\} \subset H'$ である。
(関数の集合として)

また、 $\forall f, g \in \text{Span}\{k(x, \cdot) \mid x \in X\}$ に対して、

$$\langle f, g \rangle_H = \langle f, g \rangle_{H'} = \langle f, g \rangle_{H_0}$$

であることは再生性よりわかる。よって、 H' の完備性より

$$H = \overline{\text{Span}\{k(x, \cdot) \mid x \in X\}} \subset H' \quad (\text{関数の集合として})$$

から $\forall f, g \in H$ に対して、 $\langle f, g \rangle_H = \langle f, g \rangle_{H'}$ である。

今、 $f \in H'$ が $f \perp H$ なら、 $k(x, \cdot) \in H$ ($\forall x$) より

$$f(x) = \langle f, k(x, \cdot) \rangle_{H'} = 0 \quad (\forall x \in X)$$

となり、 $\|f\|_{H'} = 0$ となる ($f = 0 \times f$ かつ $\|f\|_{H'} = 0 \times \|f\|_{H'} = 0$)

よって、 $H = H'$ (関数の集合として) であり、 H と H' は \mathcal{B} 上の
空間として同型である。

//

定理

ある実数値関数のヒルベルト空間 \mathcal{H} を考える。

$\forall x \in X$ に対し $e_x: f \mapsto f(x)$ が \mathcal{H} のノルムに $\|\cdot\|_{\mathcal{H}}$ による連続

$\Leftrightarrow \mathcal{H}$ は RKHS

(証明) (\Rightarrow) Riesz の表現定理より $\exists k_x \in \mathcal{H}$ として

$$e_x(f) = \langle f, k_x \rangle_{\mathcal{H}}$$

と書ける。これは再生性にも他ならない。

(\Leftarrow) RKHS の定義から分かる。なぜなら

$\forall x \in X$ に対し $\exists k_x \in \mathcal{H}$ として

$$f(x) = \langle f, k_x \rangle_{\mathcal{H}}$$

であるから、これは f の $\|\cdot\|_{\mathcal{H}}$ による連続

実際、 $|f(x) - f'(x)| = |\langle f - f', k_x \rangle_{\mathcal{H}}|$

$$\leq \|f - f'\|_{\mathcal{H}} \sqrt{\langle k_x, k_x \rangle_{\mathcal{H}}}$$

より、 $f' \rightarrow f \in \mathcal{H}$ ならば $|f(x) - f'(x)| \rightarrow 0$ である。

例

$$k(x, y) = \min(x, y) \quad (0 \leq x, y \leq 1)$$

とすると、 k は正定値カーネル。 $k(x, \cdot)$ の弱微分集合 $\mathcal{H} = \mathcal{H}[0, 1]$

$$\langle f, k(x, \cdot) \rangle_{\mathcal{H}} = \int_0^1 f'(y) \mathbb{1}_{[0, x]}(y) dy \quad (x \in [0, 1])$$

$$= \int_0^x f'(y) dy = f(x) \quad (f \text{ は } f(0)=0 \text{ とおす})$$

$$\uparrow \quad f(x) = \int_0^x F(u) du, \quad g(x) = \int_0^x G(u) du \text{ とおす。}$$

$$\langle f, g \rangle_{\mathcal{H}} = \int_0^1 F(u)G(u) du \quad \in L^2 \text{ 内積。}$$

表現定理

定理

$$\hat{f} = \arg \min_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^n \ell(y_i, f(x_i)) + \lambda \|f\|_{\mathcal{H}}^2$$

は ある $\alpha_i \in \mathbb{R} \subseteq \mathbb{A} \cup \mathbb{Z}$

$$\hat{f}(x) = \sum_{i=1}^n \alpha_i k(x_i, x)$$

と表現できる。 //

(証明) $\mathcal{H}_1 = \text{span} \{k(x_1, \cdot), \dots, k(x_n, \cdot)\}$ とすると。

$\forall f \in \mathcal{H}$ は

$$f = f_1 + f_2 \quad (f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_1^\perp)$$

と書ける。

今 $\forall x_i \ (i=1, 2, \dots, n)$ は

$$\begin{aligned} f(x_i) &= \langle f, k(x_i, \cdot) \rangle_{\mathcal{H}} = \langle f_1, k(x_i, \cdot) \rangle_{\mathcal{H}} + \underbrace{\langle f_2, k(x_i, \cdot) \rangle_{\mathcal{H}}}_{=0} \\ &= \langle f_1, k(x_i, \cdot) \rangle_{\mathcal{H}} \end{aligned} \quad (\because f_2 \perp \mathcal{H}_1)$$

よって \mathbb{Z} は注意する。

また、

$$\|f\|_{\mathcal{H}}^2 = \|f_1\|_{\mathcal{H}}^2 + \|f_2\|_{\mathcal{H}}^2 \quad \mathbb{Z} \in \text{あるの } \mathbb{Z}:$$

$$\frac{1}{n} \sum_{i=1}^n \ell(y_i, f(x_i)) + \lambda \|f\|_{\mathcal{H}}^2$$

$$= \frac{1}{n} \sum_{i=1}^n \ell(y_i, f_1(x_i)) + \lambda (\|f_1\|_{\mathcal{H}}^2 + \|f_2\|_{\mathcal{H}}^2)$$

と仮定。この最適化問題は $f_2 = 0$ にある u_2 達成できる。

つまり、 $f \in \mathcal{H}$ とあり、 $f = \sum_{i=1}^n \alpha_i k(x_i, \cdot)$ と書ける。

表現定理より、RKHS上の学習は

$$\min_{\alpha_i \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n \ell(y_i, \sum_{j=1}^n \alpha_j k(x_j, x_i)) + \lambda \sum_{i,j=1}^n \alpha_i \alpha_j k(x_i, x_j)$$

と有限次元最適化問題に帰着できる。

● カーネルの和とそのRKHS

定理

k_1, k_2 : 正定値カーネル,

$\mathcal{H}_1, \mathcal{H}_2$: k_1, k_2 の再生核 $\mathcal{H}_1, \mathcal{H}_2$ 持つ RKHS

$k = k_1 + k_2$ の再生核 \mathcal{H} 持つ RKHS (非

$\mathcal{H} = \{ f = f_1 + f_2 \mid f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_2 \}$

核空間 \mathcal{H} のノルム $\|\cdot\|_{\mathcal{H}}$ (非

$$\|f\|_{\mathcal{H}}^2 = \inf_{\substack{f = f_1 + f_2 \\ f_1 \in \mathcal{H}_1, f_2 \in \mathcal{H}_2}} \{ \|f_1\|_{\mathcal{H}_1}^2 + \|f_2\|_{\mathcal{H}_2}^2 \} \quad (\forall f \in \mathcal{H})$$

と等値する。

証明

$\mathcal{F} = \mathcal{H}_1 \oplus \mathcal{H}_2$ の集合 $\mathcal{H}_1 \times \mathcal{H}_2$ 上に以下のノルム $\|\cdot\|_{\mathcal{F}}$ を \mathcal{F} 空間

とす:

$$\|(f_1, f_2)\|_{\mathcal{F}}^2 = \|f_1\|_{\mathcal{H}_1}^2 + \|f_2\|_{\mathcal{H}_2}^2 \quad ((f_1, f_2) \in \mathcal{F})$$

$$(\langle (f_1, f_2), (f'_1, f'_2) \rangle_{\mathcal{F}} = \langle f_1, f'_1 \rangle_{\mathcal{H}_1} + \langle f_2, f'_2 \rangle_{\mathcal{H}_2})$$

ここで $u: \mathcal{F} \rightarrow \mathcal{H}$ とし $N = u^{-1}(0)$ とす。
 $(f_1, f_2) \mapsto f_1 + f_2$

u は定義より線形写像 \mathcal{F} から \mathcal{H} への全射であることがわかる。

N は \mathcal{F} の閉部分空間であることもわかる。

なぜなら $((f_n, -f_n))_{n=1}^{\infty} \in N$ かつ $(f_1, f_2) \in \mathcal{F}$ に収束するとす。

$f_n \rightarrow f_1$ in $\mathcal{H}_1, -f_n \rightarrow f_2$ in \mathcal{H}_2 とす。

$$\begin{aligned} \text{このとき} \quad f_n(x) &= \langle f_n, k_1(x, \cdot) \rangle_{\mathcal{H}_1} \rightarrow \langle f_1, k_1(x, \cdot) \rangle_{\mathcal{H}_1} = f_1(x) \\ -f_n(x) &= \langle -f_n, k_2(x, \cdot) \rangle_{\mathcal{H}_2} \rightarrow \langle f_2, k_2(x, \cdot) \rangle_{\mathcal{H}_2} = f_2(x) \end{aligned}$$

と各点収束するから $f_1(x) = -f_2(x)$ ($\forall x \in X$) とす。

つまり $f_1 + f_2 = 0$ なる $(f_1, f_2) \in N$ とす。よって N は閉。

N^\perp を F 内の N の直交空間とす。 $v \in U$ の N^\perp への制限とす。
 とす。 $v: N^\perp \rightarrow H$ は 全単射 とす。

$v \in U$ と、 H 上の内積

$$\langle f, g \rangle_H := \langle v^{-1}(f), v^{-1}(g) \rangle_F$$

と定義する。 あるいは H 上の内積を用いて

1. $\mathcal{K} = \mathcal{K}_1 + \mathcal{K}_2$ を再生核と \mathcal{K} 持 $\mathcal{R}(H)$
2. H のノルムは定数 α のスケールを付した \mathcal{K} のノルムと一致する。

ことを示す。

1. $f \in H$ (実数) $v^{-1}(f) = (f_1, f_2)$ (特に $f = f_1 + f_2$) とす。
 また $\forall x \in X$ とす、 $\mathcal{K}(x, \cdot) = \mathcal{K}_1(x, \cdot) + \mathcal{K}_2(x, \cdot) \in H$ とす。

$$(\mathcal{K}'(x, \cdot), \mathcal{K}''(x, \cdot)) = v^{-1}(\mathcal{K}(x, \cdot)) \in N^\perp \subset F$$

とす。

$$\begin{aligned} \text{つまり} \quad \mathcal{K}(x, \cdot) &= \mathcal{K}_1(x, \cdot) + \mathcal{K}_2(x, \cdot) \\ &= \mathcal{K}'(x, \cdot) + \mathcal{K}''(x, \cdot) \end{aligned}$$

$$\text{よって} \quad (\mathcal{K}_1(x, \cdot) - \mathcal{K}'(x, \cdot)) + (\mathcal{K}_2(x, \cdot) - \mathcal{K}''(x, \cdot)) = 0$$

$$\text{よって} \quad (\mathcal{K}_1(x, \cdot) - \mathcal{K}'(x, \cdot), \mathcal{K}_2(x, \cdot) - \mathcal{K}''(x, \cdot)) \in N$$

$$\text{よって} \quad v^{-1}(f) = (f_1, f_2) \in N^\perp \text{ なること。}$$

$$\begin{aligned} &\langle (f_1, f_2), (\mathcal{K}_1(x, \cdot) - \mathcal{K}'(x, \cdot), \mathcal{K}_2(x, \cdot) - \mathcal{K}''(x, \cdot)) \rangle_F \\ &= \langle f_1, \mathcal{K}_1(x, \cdot) - \mathcal{K}'(x, \cdot) \rangle_{H_1} + \langle f_2, \mathcal{K}_2(x, \cdot) - \mathcal{K}''(x, \cdot) \rangle_{H_2} \\ &= 0 \end{aligned}$$

$$\text{よって、特に} \quad \langle f_1, \mathcal{K}'(x, \cdot) \rangle_{H_1} + \langle f_2, \mathcal{K}''(x, \cdot) \rangle_{H_2} = f_1(x) + f_2(x)$$

$\langle \cdot, \cdot \rangle_H$ の定義より

$$\begin{aligned} \langle f, \mathcal{K}(x, \cdot) \rangle_H &= \langle v^{-1}(f), v^{-1}(\mathcal{K}(x, \cdot)) \rangle_F \\ &= \langle (f_1, f_2), (\mathcal{K}'(x, \cdot), \mathcal{K}''(x, \cdot)) \rangle_F \\ &= \langle f_1, \mathcal{K}'(x, \cdot) \rangle_{H_1} + \langle f_2, \mathcal{K}''(x, \cdot) \rangle_{H_2} \\ &= f_1(x) + f_2(x) = f(x) \end{aligned}$$

よって、 $\rightarrow H$ は \mathcal{K} を再生核と \mathcal{K} 持 $\mathcal{R}(H)$ 。

2.

$f \in \mathcal{H}$ である \rightarrow 固定する. $f = f_1 + f_2$ ($(f_1, f_2) \in \mathcal{F}$) とする.
 $(g_1, g_2) = (f_1, f_2) - v^{-1}(f) \in \mathcal{N}$ とおく.

\mathcal{F} の ノルム について

$$\|(f_1, f_2)\|_{\mathcal{F}}^2 = \|f_1\|_{\mathcal{H}_1}^2 + \|f_2\|_{\mathcal{H}_2}^2$$

である. $(g_1, g_2) \in \mathcal{N}$, $v^{-1}(f) \in \mathcal{N}^\perp$ であるから

$$\|(f_1, f_2)\|_{\mathcal{F}}^2 = \|(g_1, g_2) + v^{-1}(f)\|_{\mathcal{F}}^2$$

\leftarrow 直交性

$$= \|(g_1, g_2)\|_{\mathcal{F}}^2 + \|v^{-1}(f)\|_{\mathcal{F}}^2$$

である. \Rightarrow (4) 成立.

$$\|f\|_{\mathcal{H}}^2 = \|v^{-1}(f)\|_{\mathcal{F}}^2 \leq \|(f_1, f_2)\|_{\mathcal{F}}^2$$

$$= \|f_1\|_{\mathcal{H}_1}^2 + \|f_2\|_{\mathcal{H}_2}^2$$

が 成立 する. 等号は $(f_1, f_2) = v^{-1}(f)$ のとき 成立 する. //

補題

\mathcal{H} : \mathcal{R} 上 再生核 となる RKHS

\mathcal{H}_λ : $\lambda \mathcal{R}$ 上 再生核 となる RKHS ($\lambda > 0$)

$$\Rightarrow \|f\|_{\mathcal{H}}^2 = \frac{1}{\lambda} \|f\|_{\mathcal{H}_\lambda}^2 //$$

(証明は略)

• スパース推定 (Multiple Kernel Learning) の応用

k_1, \dots, k_M : M 個のカーネル. \Rightarrow これらの 線形結合 によって "良い" カーネル を 得たい.

\mathcal{H}_m : k_m の RKHS

$f \in \mathcal{H}_1 \oplus \dots \oplus \mathcal{H}_M$ とする. $1 \leq p \leq 2$ とする.

$$\min_{(f_m)} \left\{ \left(\sum_{m=1}^M \|f_m\|_{\mathcal{H}_m}^p \right)^{\frac{1}{p}} \mid f = \sum_{m=1}^M f_m, f_m \in \mathcal{H}_m \right\}$$

$$= \min_{(\lambda_m)} \left\{ \|f\|_{\mathcal{H}_{\mathcal{R}}} \mid \mathcal{R} = \sum_{m=1}^M \lambda_m k_m, \sum_{m=1}^M \lambda_m^{\frac{p}{2-p}} = 1, \lambda_m \geq 0 \right\}$$

$p=2$ のときは $\max(\lambda_m) = 1$ とする.

特2.

$$\sum_{m=1}^M \lambda_m k_m, \sum_{m=1}^M \lambda_m = 1, \lambda_m \geq 0 \text{ (凸組合)}$$

$p=1$:

$$\min \left\{ \|f\|_{\mathcal{H}_k} \mid k = \overbrace{\text{Conv}(\{k_m\}_m)} \right\}$$

$$= \min \left\{ \|f_1\|_{\mathcal{H}_1} + \dots + \|f_M\|_{\mathcal{H}_M} \mid f = \sum_{m=1}^M f_m, f_m \in \mathcal{H}_m \right\}$$

\Rightarrow

$$\min_{\substack{f_m \in \mathcal{H}_m \\ (m=1, \dots, M)}} L\left(\sum_{m=1}^M f_m\right) + \lambda \underbrace{\sum_{m=1}^M \|f_m\|_{\mathcal{H}_m}}_{\text{クルーゴL1正則化}}$$

$$= \min_{\substack{f \in \mathcal{H}_k \\ k \in \text{Conv}(\{k_m\}_{m=1}^M)}} L(f) + \lambda \|f\|_{\mathcal{H}_k}$$

⊛ カーネルの凸組合の学習 \Leftrightarrow クルーゴL1正則化 (スパース正則化)

$p=2$:

$$\min_{\substack{f_m \in \mathcal{H}_m \\ (m=1, \dots, M)}} L\left(\sum_{m=1}^M f_m\right) + \lambda \sum_{m=1}^M \|f_m\|^2$$

$$= \min_{\substack{f \in \mathcal{H}_k \\ k = \sum_{m=1}^M k_m}} L(f) + \lambda \|f\|_{\mathcal{H}_k}^2$$

⊛ カーネルの単純和を用いた学習 \Leftrightarrow クルーゴL2正則化

Micchelli & Pontil: Learning the kernel function via regularization.
Journal of Machine Learning Research, 9, 1099-1125 (2005)

証明は省略: $k = \sum \lambda_m k_m$ として

$$\|f\|_{\mathcal{H}_k}^2 = \min \left\{ \sum_{m=1}^M \lambda_m^{-1} \|f_m\|_{\mathcal{H}_m}^2 \mid f = \sum_{m=1}^M f_m \right\}$$

2. ある $\Sigma \in \mathcal{H}$. $\lambda_m \geq 0$ 条件 $(\sum_{m=1}^M \lambda_m^{-\frac{p}{2}} = 1, \lambda_m \geq 0)$ のもとで最小化
すれば、左辺の中身 = $(\|f_m\|_{\mathcal{H}_m}^p)^{\frac{1}{p}}$ を得る。

• 平行移動不変カーネルの特徴付け.

$k(x, y) = \psi(x - y)$ と書ける時. k は平行移動不変カーネルと言う.

定理 (Bochner)

\mathbb{R}^d 上の連続関数 ψ を用いて, $k(x, y) = \psi(x - y)$ と書ける時, k が正定値カーネルとなるための必要+十分条件は, \mathbb{R}^d 上のある 非負有限 Borel 測度 μ が存在して,

$$\psi(x) = \int \exp(i x^T z) d\mu(z)$$

が成り立つことである. //

(証明は略).

例 (Gaussian カーネル)

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

$$\Rightarrow \mu(d\omega) = \underbrace{\frac{1}{(2\pi)^d} \exp\left(-\frac{\sigma^2}{2} \|\omega\|^2\right)}_{\substack{! \\ p(\omega) \geq 0}} d\omega$$

定理

μ が連続な密度関数 p を持つとき:

$$k(x, y) = \int \exp(i \omega^T (x - y)) \underline{p(\omega)} d\omega.$$

このとき, \mathcal{H} は $\mathcal{R} \mathcal{H} \mathcal{F} \mathcal{H}$ (は)

$$\mathcal{H} = \left\{ f \in L^2(\mathbb{R}^d) \mid \int \frac{|\hat{f}(\omega)|^2}{p(\omega)} d\omega < \infty \right\},$$

$$\langle f, g \rangle_{\mathcal{H}} = \int \frac{\hat{f}(\omega) \overline{\hat{g}(\omega)}}{p(\omega)} d\omega$$

と与えられる. f は \hat{f} として, $\hat{f}(\omega) = \frac{1}{(2\pi)^d} \int \exp(-i \omega^T z) f(z) dz.$

(証明の概略)

$k(x, y)$ が \mathcal{H} の再生核となることを示す

$$k(x, y) = \int \exp(i\omega^T x) \exp(-i\omega^T y) p(\omega) d\omega$$

$$\text{よって } \hat{k}(\cdot, y)(\omega) = \exp(-i\omega^T y) p(\omega)$$

$$\begin{aligned} \Rightarrow \langle f, k(y, \cdot) \rangle_{\mathcal{H}} &= \int \frac{\hat{f}(\omega) \overline{\hat{k}(\cdot, y)(\omega)}}{p(\omega)} d\omega \\ &= \int \frac{\hat{f}(\omega) \overline{\exp(-i\omega^T y) p(\omega)}}{p(\omega)} d\omega \\ &= \int \hat{f}(\omega) \exp(i\omega^T y) d\omega \\ &= f(y) \end{aligned}$$

→ 此の再生性(は)他ならない。

例 (Gaussian カーネル)

$$\langle f, g \rangle_{\mathcal{H}} = \int \frac{\hat{f}(\omega) \overline{\hat{g}(\omega)}}{\left(\frac{1}{2\pi}\right)^d \exp\left(-\frac{\sigma^2}{2} \|\omega\|^2\right)} d\omega$$

$$= \int (2\pi)^d \hat{f}(\omega) \overline{\hat{g}(\omega)} \underbrace{\exp\left(\frac{\sigma^2}{2} \|\omega\|^2\right)}_{\uparrow} d\omega$$

高周波成分には大きな重み。

$$\Rightarrow \|f\|_{\mathcal{H}}^2 < \infty \text{ となるためには}$$

高周波成分が急減少する必要がある。

つまり、 $f \in \mathcal{H}$ は滑らかな関数になる。

