・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

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

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）

$k : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ が次をみたすとき，k を正定値カーネルと言う．

- 対称性：$k(x, x') = k(x', x) \quad (\forall x, x' \in \mathcal{X})$
- 正値性：$\sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \geq 0$

  $(\forall \alpha_i \in \mathbb{R}, \forall x_i \in \mathcal{X}, m = 1, 2, \cdots)$

//

例：（ユークリッド空間の内積）

$x, y \in \mathbb{R}^d$ に対し，$k(x, y) = x^T y = \sum_{i=1}^{d} x_i y_i$ 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 $\mathcal{X}$ 上の RKHS とは，$\mathcal{X}$ 上の（実数値）関数からなるヒルベルト空間で，$\forall x \in \mathcal{X}$ に対して $k_x \in \mathcal{H}$ が存在して

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

が成り立つものとする．　← チェックせよ．

特に，$k(x, y) = \langle k_x, k_y \rangle_{\mathcal{H}}$ は正定値カーネルであり，$\mathcal{H}$ に付随した再生核（Reproducing kernel）とよぶ．

//

定義より，RKHS $\mathcal{H}$ に対してカーネル k が定まる．
逆に正定値カーネル k に対して，RKHS $\mathcal{H}$ が一意に定まる．

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

定理 (Moore - Aronszajn)

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

$\Rightarrow$ k を再生核とする RKHS が一意的に存在. //

証明

$$\mathcal{H}_0 := \mathrm{span}\{k(x,\cdot) \mid x\in\mathcal{X}\}$$
$$= \left\{ \sum_{i=1}^m k(x_i,\cdot)\alpha_i \;\middle|\; \alpha_i\in\mathbb{R},\ x_i\in\mathcal{X},\ m=1,2,\dots \right\}$$

とし. $\mathcal{H}_0$ に "内積" $\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$ を次のように定める:

$f(x)=\sum_{i=1}^n \alpha_i k(x_i,x)$, $g(x)=\sum_{j=1}^m \beta_j k(y_j,\cdot)\in\mathcal{H}_0$ に対し

$$\langle f,g\rangle_{\mathcal{H}_0} = \sum_{i=1}^n\sum_{j=1}^m \alpha_i\beta_j k(x_i,y_j).$$

これは f,g の表示のしかたによらず well-defined: $f=\sum_{i=1}^{n'}\alpha'_i k(x'_i,\cdot)$ とも書けた時

$$\langle f,g\rangle_{\mathcal{H}_0} = \sum_{i=1}^{n'}\sum_{j=1}^m \alpha'_i\beta_j k(x'_i,y_j) = \sum_{j=1}^m \beta_j f(y_j)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^m \alpha_i\beta_j k(x_i,y_j) = \sum_{j=1}^m \beta_j f(y_j)$$

等しい

これが実際に内積であることを示す.

- 線形性: $\langle af+bg,h\rangle_{\mathcal{H}_0} = a\langle f,h\rangle_{\mathcal{H}_0} + b\langle g,h\rangle_{\mathcal{H}_0}$ → 自明
- 対称性: $\langle f,g\rangle_{\mathcal{H}_0} = \langle g,f\rangle_{\mathcal{H}_0}$ → 自明
- 正値性, 非退化性: $\langle f,f\rangle_{\mathcal{H}_0}\geq 0$ かつ $\langle f,f\rangle_{\mathcal{H}_0}=0 \Rightarrow f=0$

まず. $\langle f,f\rangle_{\mathcal{H}_0}\geq 0$ は k の正定値性よりすぐ確認できる.

これより. $\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$ は Cauchy - Shwartz の不等式が成り立つ

実際, $\forall f,g\in\mathcal{H}_0$ に対し. ($a,b\in\mathbb{R}$, $\|f\|_{\mathcal{H}_0}^2 := \langle f,f\rangle_{\mathcal{H}_0}$)

$$0\leq \langle af-bg,af-bg\rangle_{\mathcal{H}_0} = a^2\|f\|_{\mathcal{H}_0}^2 - 2ab\langle f,g\rangle_{\mathcal{H}_0} + b^2\|g\|_{\mathcal{H}_0}^2$$

より. $2ab\langle f,g\rangle_{\mathcal{H}_0} \leq a^2\|f\|_{\mathcal{H}_0}^2 + b^2\|g\|_{\mathcal{H}_0}^2$

$\|f\|_{\mathcal{H}_0}^2, \|g\|_{\mathcal{H}_0}^2 > 0$ なら, $a=\frac{1}{\|f\|_{\mathcal{H}_0}}$, $b=\frac{1}{\|g\|_{\mathcal{H}_0}}$ とおけば

$$\langle f,g\rangle_{\mathcal{H}_0} \leq \|f\|_{\mathcal{H}_0}\|g\|_{\mathcal{H}_0}$$

を得る. $\|f\|_{\mathcal{H}_0}=0$, または $\|g\|_{\mathcal{H}_0}=0$ のときは. $\langle f,g\rangle_{\mathcal{H}_0}=0$ がすぐ示せる.

ここで. $\langle f,f\rangle_{\mathcal{H}_0}=0$ なら. $\forall x\in\mathcal{X}$ で $f(x)=\langle f,k(x,\cdot)\rangle_{\mathcal{H}_0}$ より.

Cauchy-Schwartz を用いて

$$|f(x)| = |\langle f,k(x,\cdot)\rangle_{\mathcal{H}_0}| \leq \underbrace{\|f\|_{\mathcal{H}_0}}_{0}\cdot\sqrt{k(x,x)} = 0$$

より. $f(x)=0$ ($\forall x\in\mathcal{X}$). つまり $f=0$ (関数として) を得る.

$\langle\cdot,\cdot\rangle_{\mathcal{H}_0}$ が内積であることが示せたので、これで定まるノルム $\|\cdot\|_{\mathcal{H}_0}$ で $\mathcal{H}_0$ を完備化できる。これを $\mathcal{H}$ とおく

完備化の定義より、$\mathcal{H}$ の内積は次のように定まる：
$f, g \in \mathcal{H}$ に対し、$\mathcal{H}_0$ 内のコーシー列 $(f_n)_n, (g_n)_n$ が存在して、
$\langle f, g\rangle_{\mathcal{H}} := \lim_{n\to\infty} \langle f_n, g_n\rangle_{\mathcal{H}_0}$ と定義される。
これはコーシー列の取り方によらない。
$f \in \mathcal{H}$ に対して、f に対応するコーシー列 $(f_m)_m \in \mathcal{H}_0$ を用いて

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

とすれば $\mathcal{H}$ の各元 $f \in \mathcal{H}$ に対し実数値関数 $x \mapsto f(x)$ が定まる。
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
ともに f に対応するとき ←コーシー列

$$|\langle f_m, k(\cdot,x)\rangle_{\mathcal{H}_0} - \langle g_m, k(\cdot,x)\rangle_{\mathcal{H}_0}|$$
$$\leq \|k(\cdot,x)\|_{\mathcal{H}_0} \|f_m - g_m\|_{\mathcal{H}_0} \longrightarrow 0$$ である。

$\mathcal{H} \ni f \mapsto (x \mapsto f(x)) \in \mathbb{R}^{\mathcal{X}}$ の対応が単射であれば 空間とみなせる)
集合として $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ と見なせる。(つまり $\mathcal{H}$ は実数値関数のヒルベルト
これは $f(x) = 0\ (\forall x \in \mathcal{X}) \Rightarrow \|f\|_{\mathcal{H}} = 0$ が示せればよい。
f に対応するコーシー列 $(f_n)_n$ を取って来る。f_n 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 を持って来て、
$\forall n \geq N$ で $\|f_n - f_N\|_{\mathcal{H}_0} \leq \frac{\varepsilon}{A}$ とできる (ε は任意)。すると、$\forall n \geq N$ で
← コーシー列なので。

$$\|f_n\|_{\mathcal{H}_0}^2 = \langle f_n - f_N, f_n\rangle_{\mathcal{H}_0} + \langle f_N, f_n\rangle_{\mathcal{H}_0}$$
$$\leq \|f_n - f_N\|_{\mathcal{H}_0} \|f_n\|_{\mathcal{H}_0} + \langle \underbrace{\sum_{j=1}^{m} \alpha_j k(x_j,\cdot)}_{f_N \text{をこう書く。}}, f_n\rangle_{\mathcal{H}_0}$$
$$\leq \frac{\varepsilon}{A}\cdot A + \underbrace{\sum_{j=1}^{m} \alpha_j f_n(x_j)}_{\to 0\ (n\to\infty)}$$

より、$\limsup_n \|f_n\|_{\mathcal{H}_0}^2 \leq \varepsilon$。

ε は任意なので、$\lim_{n\to\infty} \|f_n\|_{\mathcal{H}_0}^2 = 0 \Longrightarrow \|f\|_{\mathcal{H}} = 0$。
よって、$\mathcal{H}$ は実数値関数のなすヒルベルト空間となる。

あとは、$\mathcal{H}$が再生核を持つことを示せば良い。

$\forall x \in \mathcal{X}$において、$(k(x,\cdot))_m$なるコーシー列で
$k_x \in \mathcal{H}$を定義すると、

$$\langle f, k_x \rangle_{\mathcal{H}} = \underbrace{\lim_{m\to\infty}}_{\text{完備化の定義}} \langle f_m, k(x,\cdot) \rangle_{\mathcal{H}_0} = f(x) \quad (\text{再生性})$$

となり、$\mathcal{H}$はRKHSになる。

再生核は $\langle k_x, k_y \rangle_{\mathcal{H}} = \lim_{m\to\infty} \langle k(x,\cdot), k(y,\cdot) \rangle_{\mathcal{H}_0}$
$= k(x,y)$

より、$\mathcal{H}$はkを再生核としてもつ。

(一意性)

$\mathcal{H}'$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、$\mathrm{span}\{k(x,\cdot) \mid x \in \mathcal{X}\} \subset \mathcal{H}'$である。
(関数の集合として)

また、$\forall f, g \in \mathrm{span}\{k(x,\cdot) \mid x \in \mathcal{X}\}$に対し、

$$\langle f, g \rangle_{\mathcal{H}} = \langle f, g \rangle_{\mathcal{H}'} = \langle f, g \rangle_{\mathcal{H}_0}$$

であることも再生性よりわかる。よって、$\mathcal{H}'$の完備性より

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

かつ $\forall f, g \in \mathcal{H}$に対し、$\langle f, g \rangle_{\mathcal{H}} = \langle f, g \rangle_{\mathcal{H}'}$でもある。

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

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

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

よって、$\mathcal{H} = \mathcal{H}'$(関数の集合として)であり、$\mathcal{H}$と$\mathcal{H}'$はヒルベルト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.

$\forall x \in \mathcal{X}$ において $e_x : f \mapsto f(x)$ が $\mathcal{H}$ のノルムについて連続

$\Longleftrightarrow$ $\mathcal{H}$ は RKHS //

(証明) ($\Longrightarrow$) Rieszの表現定理より $\exists k_x \in \mathcal{H}$ で:

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

と書ける. これは再生性に他ならない.

($\Longleftarrow$) RKHSの定義からすぐわかる. なぜなら

$\forall x \in \mathcal{X}$ に対し. $\exists k_x \in \mathcal{H}$ で:

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

であるが. これは f について連続

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. $f' \to f \in \mathcal{H}$ なら. $|f(x) - f'(x)| \to 0$ である. //

例

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

とおくと、kは正定値カーネル。$k(x,\cdot)$の弱微分は $\mathbb{1}_{[0,x]}$

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

$$= \int_0^y f'(y)\,dy = f(y) \qquad (f \text{は} f(0)=0 \text{とする})$$

$f(x) = \int_0^x F(u)\,du,\ g(x) = \int_0^x G(u)\,du$ に対して、
$\langle f,g\rangle_H := \int_0^1 F(u)G(u)\,du$ としている。

表現定理

定理

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

ある単調増大な $\psi: [0,\infty) \to \mathbb{R}$ を用いて $\psi(\|f\|_{\mathcal{H}})$ とおきかえても同様に成り立つ。

はある $\alpha_i \in \mathbb{R}$ を用いて

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

と表現される。 //

(証明) $H_1 = \mathrm{span}\{k(x_1,\cdot), \ldots, k(x_n,\cdot)\}$ とすると、
$^\forall f \in \mathcal{H}$は

$$f = f_1 + f_2 \qquad (f_1 \in H_1,\ f_2 \in H_1^{\perp})$$

と書ける。
今 $^\forall x_i\ (i=1,2,\ldots,n)$ に対し

$$f(x_i) = \langle f, k(x_i,\cdot)\rangle_{\mathcal{H}} = \langle f_1, k(x_i,\cdot)\rangle_{\mathcal{H}} + \underbrace{\langle f_2, k(x_i,\cdot)\rangle_{\mathcal{H}}}_{=0}$$

$$= \langle f_1, k(x_i,\cdot)\rangle_{\mathcal{H}} \qquad (\because f_2 \perp H_1)$$

であることに注意する。
また、

$$\|f\|_{\mathcal{H}}^2 = \|f_1\|_{\mathcal{H}}^2 + \|f_2\|_{\mathcal{H}}^2 \text{ でもあるので}$$

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

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

となり、この最小化は $f_2 = 0$ において達成される。
つまり、$f \in \mathcal{H}_1$ であり、$f = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot)$ と書ける。

//

表現定理より、RKHS上の学習は

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

と有限次元最適化問題に帰着される。

● 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 を再生核として持つ RKHS

$k = k_1 + k_2$ を再生核として持つ RKHS は

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

なる空間で. $\mathcal{H}$ のノルム $\|\cdot\|_{\mathcal{H}}$ は

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

で与えられる. //

証明

$\mathcal{F} = \mathcal{H}_1 \oplus \mathcal{H}_2$ を 集合 $\mathcal{H}_1 \times \mathcal{H}_2$ に以下のノルムを入れたヒルベルト空間とする:

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

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

ここで. $u: \mathcal{F} \longrightarrow \mathcal{H}$, $(f_1, f_2) \longmapsto f_1 + f_2$ として. $N = u^{-1}(0)$ とする.

u は定義より. 線形かつ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. $((f_n, -f_n))_{n=1}^{\infty} \in N$ が $(f_1, f_2) \in \mathcal{F}$ に収束するとき.
$f_n \longrightarrow f_1$ in $\mathcal{H}_1$, $-f_n \longrightarrow f_2$ in $\mathcal{H}_2$ である.

このとき.
$$f_n(x) = \langle f_n, k_1(x, \cdot) \rangle_{\mathcal{H}_1} \longrightarrow \langle f_1, k_1(x, \cdot) \rangle_{\mathcal{H}_1} = f_1(x)$$
$$-f_n(x) = \langle -f_n, k_2(x, \cdot) \rangle_{\mathcal{H}_2} \longrightarrow \langle f_2, k_2(x, \cdot) \rangle_{\mathcal{H}_2} = f_2(x)$$

と各点収束するので. $f_1(x) = -f_2(x)$ $({}^\forall x \in \mathcal{X})$ である.
つまり. $f_1 + f_2 = 0$ で $(f_1, f_2) \in N$ である. よって. N は閉.

$N^\perp$を$\mathcal{F}$内でのNの直交空間とする。vをuの$N^\perp$への制限とする。
すると、$\underline{v : N^\perp \longrightarrow \mathcal{H}\text{ は 全単射}}$である。

vを用いて、$\mathcal{H}$に内積

$$\langle f, g \rangle_{\mathcal{H}} := \langle v^{-1}(f), v^{-1}(g) \rangle_{\mathcal{F}}$$

と定義する。あとは$\mathcal{H}$がこの内積を用いて

1. $k = k_1 + k_2$を再生核として持つRKHS
2. $\mathcal{H}$のノルムは定理のステートメントのように与えられる。

ことを示す。

1. $f \in \mathcal{H}$に対し、$v^{-1}(f) = (f_1, f_2)$（特に$f = f_1 + f_2$）とする。
また、$\forall x \in \mathcal{X}$に対し、$k(x,\cdot) = k_1(x,\cdot) + k_2(x,\cdot) \in \mathcal{H}$である。
$(k'(x,\cdot), k''(x,\cdot)) = v^{-1}(k(x,\cdot)) \in N^\perp \subset \mathcal{F}$
とおく。
ここで、
$$k(x,\cdot) = k_1(x,\cdot) + k_2(x,\cdot)$$
$$= k'(x,\cdot) + k''(x,\cdot)$$
より、$(k_1(x,\cdot) - k'(x,\cdot)) + (k_2(x,\cdot) - k''(x,\cdot)) = 0$
なので、$(k_1(x,\cdot) - k'(x,\cdot), k_2(x,\cdot) - k''(x,\cdot)) \in N$
今、$v^{-1}(f) = (f_1, f_2) \in N^\perp$なので、

$$\langle (f_1, f_2), (k_1(x,\cdot) - k'(x,\cdot), k_2(x,\cdot) - k''(x,\cdot)) \rangle_{\mathcal{F}}$$
$$= \langle f_1, k_1(x,\cdot) - k'(x,\cdot) \rangle_{\mathcal{H}_1} + \langle f_2, k_2(x,\cdot) - k''(x,\cdot) \rangle_{\mathcal{H}_2}$$
$$= 0$$
となり、特に、$\langle f_1, k'(x,\cdot) \rangle_{\mathcal{H}_1} + \langle f_2, k''(x,\cdot) \rangle_{\mathcal{H}_2} = f_1(x) + f_2(x)$.

$\langle \cdot, \cdot \rangle_{\mathcal{H}}$の定義より
$$\begin{aligned}
\langle f, k(x,\cdot) \rangle_{\mathcal{H}} &= \langle v^{-1}(f), v^{-1}(k(x,\cdot)) \rangle_{\mathcal{F}} \\
&= \langle (f_1, f_2), (k'(x,\cdot), k''(x,\cdot)) \rangle_{\mathcal{F}} \\
&= \langle f_1, k'(x,\cdot) \rangle_{\mathcal{H}_1} + \langle f_2, k''(x,\cdot) \rangle_{\mathcal{H}_2} \\
&= f_1(x) + f_2(x) = f(x)
\end{aligned}$$
である。$\longrightarrow$ $\mathcal{H}$はkを再生核として持つRKHS.

2.

$f \in \mathcal{H}$ を一つ固定する. $f = f_1 + f_2$ $((f_1, f_2) \in \mathcal{F})$ とする.

$(g_1, g_2) = (f_1, f_2) - v^{-1}(f) \in N$ とおく.

$\mathcal{F}$ のノルムより

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

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

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

← 直交性

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

である. これらより

$$\|f\|_{\mathcal{H}}^2 = \|v^{-1}(f)\|_{\mathcal{F}}^2 \leq \|(f_1, f_2)\|_{\mathcal{F}}^2 = \|f_1\|_{\mathcal{H}_1}^2 + \|f_2\|_{\mathcal{H}_2}^2$$

が成り立ち、等号は $(f_1, f_2) = v^{-1}(f)$ のときのみ成立する. //

補題

$\mathcal{H}$: k を再生核とする RKHS
$\mathcal{H}_\lambda$: λk を再生核とする RKHS (ただし $\lambda > 0$)

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

(証明は略)

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

$k_1, \ldots, k_M$: M 個のカーネル. これらの線形結合で"良い"カーネルを得たい.

$\mathcal{H}_m$: k_m の RKHS

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

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

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

k のRKHS

$p=2$ のときは、$\max(\lambda_m) = 1$ とする.

特に.

$\sum_{m=1}^{M}\lambda_m k_m,\ \sum_{m=1}^{M}\lambda_m=1,\ \lambda_m\geq 0$ (凸結合)

$p=1$:

$$\min\left\{\|f\|_{\mathcal{H}_k} \;\middle|\; k=\overbrace{\mathrm{conv}(\{k_m\}_m)}\right\}$$

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

$\Longrightarrow$

$$\min_{\substack{f_m\in\mathcal{H}_m\\(m=1,\dots,M)}} L\left(\sum_{m=1}^{M}f_m\right)+\lambda\underbrace{\sum_{m=1}^{M}\|f_m\|_{\mathcal{H}_m}}_{\text{グループ}L_1\text{正則化}}$$

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

(※) カーネルの凸結合の学習 $\Longleftrightarrow$ グループL_1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 $\Longleftrightarrow$ グループL_2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: $k=\sum\lambda_m k_m$ に対し. $\because \|f_m\|_{\mathcal{H}_{\lambda_m k_m}}^2=\lambda_m^{-1}\|f_m\|_{\mathcal{H}_m}^2$

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

であることより. λ_m を条件 $\left(\sum_{m=1}^{M}\lambda_m^{\frac{p}{2-p}}=1,\ \lambda_m\geq 0\right)$ のもとで最小化すれば. 左辺の中身 $=\left(\|f_m\|_{\mathcal{H}_m}^p\right)^{\frac{1}{p}}$ 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 Λ が存在して，

$$\psi(x)=\int \exp(\sqrt{-1}\,x^T z)\,d\Lambda(z)$$

が成り立つことである． //

(証明は略)．

例 (Gaussian カーネル)

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

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

定理 Λ が連続な密度関数 p を持つとする：

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

このとき，対応する RKHS $\mathcal{H}$ は

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

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

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

(証明の概略)

$k(x,y)$ が $\mathcal{H}$ の再生核となることを示す

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

より、 $\hat{k}(\cdot,y)(\omega)=\exp(-i\omega^T y)\,p(\omega)$

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

$$= \int \frac{\hat{f}(\omega)\,\overline{\exp(-i\omega^T y)\,p(\omega)}}{p(\omega)}\,d\omega$$

$$= \int \hat{f}(\omega)\exp(i\omega^T y)\,d\omega$$

$$= f(y)$$

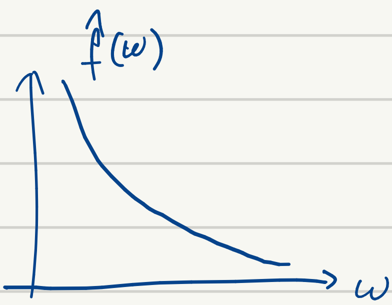
$\longrightarrow$ これは再生性に他ならない.

例 (Gaussianカーネル)

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

$$= \int (2\pi)^d\,\hat{f}(\omega)\,\overline{\hat{g}(\omega)}\,\underbrace{\exp\left(\frac{\sigma^2}{2}\|\omega\|^2\right)}_{\text{高周波成分には大きな重み}}\,d\omega$$

//



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

高周波成分が急減少しないといけない.

つまり、$f\in\mathcal{H}$ は滑らかな関数になる.