・カーネル関数の分解と再生核ヒルベルト空間の特徴付け

カーネル関数はある条件のもと

$$k(x,x') = \sum_{i=1}^{\infty} \mu_i e_i(x) e_i(x')$$

のように分解できる。ただし、

- $\mu_1 \geqq \mu_2 \geqq \mu_3 \geqq \cdots \geqq 0$ （固有値）
- $(e_i)_i$ は ある測度 ν に関する $L^2(\nu)$ 内の ONS.
  $\left(\int e_i(x) e_{i'}(x) d\nu(x) = 0 \right)$ （正規直交系）.
  $(i \neq i')$

今、ある位相空間 $\mathcal{X}$ 上の（非負）Borel 測度 ν に対し、
k で決まる積分作用素を

$$T_k f(x) = \int f(y) k(y,x) d\nu(y) \qquad (f \in L^2(\nu))$$

とする。
作用素 T_k と k の分解は密接に関係している。

作用素のスペクトル理論等が示せる。

定理（Mercer 分解）
$\mathcal{X}$ をハウスドルフ位相空間とする。
ν を $\mathcal{X}$ 上の非負 Borel 測度で $\mathrm{supp}(\nu) = \mathcal{X}$ であるとする。
$k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ は連続な正定値カーネルで、

$$\int k(x,x) d\nu(x) < \infty$$

とする。
このとき、k によって決まる RKHS $\mathcal{H}$ は可分で、$\forall x \in \mathcal{X}$ において

$$k(x,x') = \sum_{i=1}^{\infty} \mu_i e_i(x) e_i(x')$$

が $\forall x' \in \mathcal{X}$ で絶対収束し、任意のコンパクト集合 $A \subset \mathcal{X}$ 上の $x' \in A$ で一様収束する。$\mu_i \searrow 0$ である。
また、この分解を用いて、T_k は

$$T_k f(x) = \sum_{i=1}^{\infty} \mu_i e_i(x) \langle e_i, f \rangle_{L^2(\nu)} \qquad (\forall x \in \mathcal{X})$$

と書ける。//

証明や関連する話題の詳細は
Steinwart & Scovel: Mercer's theorem on general domains: on the interaction between measures, kernels, and RKHSs
Constructive Approximation, 35:363-417, 2012.
を参照されたい。

① Mercerの分解を用いて、RKHSは次のようにも書ける。 $\left(\frac{0}{0}=0 \text{ とする}\right)$

$$\boxed{\begin{aligned} H &= \left\{ \sum_{i=1}^{\infty} a_i \mu_i^{\frac{1}{2}} e_i \;\middle|\; \sum_{i=1}^{\infty} a_i^2 < \infty \right\} \\ &= \left\{ \sum_{i=1}^{\infty} b_i e_i \;\middle|\; \sum_{i=1}^{\infty} \frac{b_i^2}{\mu_i} < \infty \right\} \end{aligned}}$$

$\sum_{i=1}^{\infty}$ としているが、$\mu_i=0\,(i \geq M)$ なら $\sum_{i=1}^{M}$ とする。

② $f, g \in \mathcal{H}$ を $f(x) = \sum_{i=1}^{\infty} b_i e_i(x)$, $g(x) = \sum_{i=1}^{\infty} b'_i e_i(x)$ と書いた時、

$$\boxed{\langle f, g \rangle_H = \sum_{i=1}^{\infty} \frac{b_i b'_i}{\mu_i}, \quad \|f\|_H^2 = \sum_{i=1}^{\infty} \frac{b_i^2}{\mu_i}}$$

と表現できる。これより、再生性も確認できる：

$$\langle f, \underbrace{k(x,\cdot)}_{k(x,\cdot) = \sum_{i=1} \mu_i e_i(x) e_i(\cdot)} \rangle_H = \sum_{i=1}^{\infty} \frac{b_i (\mu_i e_i(x))}{\mu_i} = \sum_{i=1}^{\infty} b_i e_i(x) = f(x)$$

係数

③ これより、$(\sqrt{\mu_i}\, e_i)_{i=1}^{\infty}$ は H の正規直交基底になっていることがわかる。

④ $\underline{\mathcal{H} = T_k^{\frac{1}{2}} L^2(\nu)}$ である。つまり $\forall f \in \mathcal{H}$ に対し、$\exists g \in L^2(\nu)$ で

$$f = T_k^{\frac{1}{2}} g \iff f = \sum_{i=1}^{\infty} \sqrt{\mu_i} \langle g, e_i \rangle_{L^2(\nu)} e_i$$

と書ける。さらに、

$$\|f\|_{\mathcal{H}} = \inf\left\{ \|g\|_{L^2(\nu)} \;\middle|\; f = T_k^{\frac{1}{2}} g,\; g \in L^2(\nu) \right\}$$

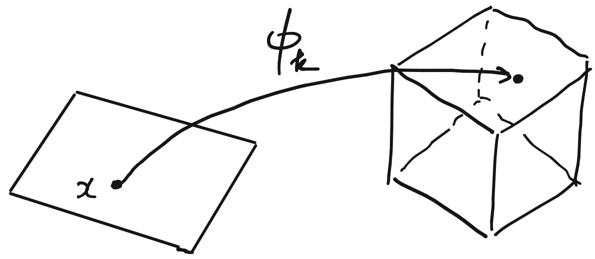
がわかる。
$\mu_i \searrow 0$ であることと、多くの場合で e_i は i が大きくなるにつれ高周波成分をもつことになるので、f は $g \in L^2(\nu)$ に平滑化作用素 $T_k^{\frac{1}{2}}$ を作用させて滑らかにさせたものになっている。

⑤ $\phi_k(x) = \left(\sqrt{\mu_i}\, e_i(x)\right)_{i=1}^{\infty}$ と書くと、

$$k(x, x') = \sum_{i=1}^{\infty} \left(\sqrt{\mu_i}\, e_i(x)\right)\left(\sqrt{\mu_i}\, e_i(x')\right)$$

$$= \langle \phi_k(x), \phi_k(x') \rangle_{\ell^2}$$ ← $\phi_k(x)$ と $\phi_k(x')$ の内積

と書ける。この ϕ_k は x の特徴写像とみなせる。



Mercer展開を用いて、カーネル法による学習は次のように書ける：

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

$$= \min_{(a_j)_{j=1}^{\infty}} \frac{1}{n} \sum_{i=1}^{n} \ell\left(y_i, \sum_{j=1}^{\infty} a_j \sqrt{\mu_j}\, e_j(x_i)\right) + \lambda \sum_{j=1}^{\infty} a_j^2$$

$$= \min_{(b_j)_{j=1}^{\infty}} \frac{1}{n} \sum_{i=1}^{n} \ell\left(y_i, \sum_{j=1}^{\infty} b_j\, e_j(x_i)\right) + \lambda \sum_{j=1}^{\infty} \frac{b_j^2}{\mu_j}$$

このように書くと、カーネル法が（無限次元）線形モデルを用いた学習法であることがわかる。
また、正則化項はリッジ正則化になっている。

※ Mercer展開は $\mathcal{H}$ が可分なら常に可能。$(e_i)_i$ を $\mathcal{H}$ の任意の完全正規直交系（CONS）とすると、

$$k(x, y) = \sum_i e_i(x) e_i(y) \qquad (\forall x, y \in \mathcal{X})$$

が成り立つ（$k(x, \cdot) = \sum_i e_i(x) e_i(\cdot)$ が $\mathcal{H}$ 内の収束の意味でも成り立つ）
先の分解は T_k に付随して決めたので ν に依存する。
しかし、k や RKHS 自体は ν に依存しない。つまり、先の分解は $(\mu_i, e_i)_i$ の選び方として $L^2(\nu)$ の正規直交系を用いた特殊な例にすぎない

・ガウス過程とRKHS

再生核ヒルベルト空間はガウス過程の特徴付けにも有用である.

<u>定義</u>

$\mathbb{R}^m$値確率変数 $(X_1, \dots, X_m)$ が多変量正規分布に従うとは, $\forall t_1, \dots, t_m \in \mathbb{R}$ を用いて,

$$t_1 X_1 + \cdots + t_m X_m$$

が $\mathbb{R}$ 上の正規分布に従うことと定義する.

↑ 分散0も許す.

平均 : $\mu = (E[X_1], \dots, E[X_m])^\top \in \mathbb{R}^m$

共分散 : $\Sigma = \left(E[X_i X_j] - E[X_i] E[X_j] \right)_{i,j} \in \mathbb{R}^{m \times m}$

を用いて, $N(\mu, \Sigma)$ と書く. (Σ はランク落ちしても良い)

//

<u>定義</u>

ある確率過程 $(f_x)_{x \in \mathcal{X}}$ を考える.

任意の $x_1, \dots, x_m \in \mathcal{X}$ (m は任意) に対し,

$$(f_{x_1}, \dots, f_{x_m}) \in \mathbb{R}^m$$

がある多変量正規分布に従うとき, $f = (f_x)_{x \in \mathcal{X}}$ は<u>ガウス過程</u>に従うという.

//

$m(x) = E[f_x]$: 平均関数

$k(x, x') = E[(f_x - m(x))(f_{x'} - m(x'))]$: 共分散関数

を用いて, $\mathcal{GP}(m, k)$ と書く. (← 正規分布は平均と分散のみで決まる.)

(注) $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ は正定値対称カーネルになっている.

<u>カーネル関数 k を用いて f を陽に表したい.</u>

$\Longrightarrow$ Karhunen-Loève 展開

定理 (Karhunen-Loève展開)

$f \sim GP(0, k)$ とする.
k は、先の Mercer展開の定理の条件をみたすとする.
k の Mercer展開を用いて、(f の適当なバージョンが)

$$f_x = \sum_{i=1}^{\infty} \xi_i \sqrt{\mu_i} e_i(x)$$

← RKHSの正規直交基底に i.i.d. のガウス係数 ξ_i をかけたもの

$$\xi_i \sim N(0,1) \quad (\text{i.i.d.})$$

と書ける.

確認

$$E[f_x f_{x'}] = E\left[\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \xi_i \xi_j \sqrt{\mu_i} e_i(x) \sqrt{\mu_j} e_j(x')\right]$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \underbrace{E[\xi_i \xi_j]}_{= \delta_{ij}} \sqrt{\mu_i \mu_j} e_i(x) e_j(x')$$

$$= \sum_{i=1}^{\infty} \mu_i e_i(x) e_i(x') = k(x, x')$$

//

定理 (Driscollの定理)

$\mathcal{H}$ が無限次元

$\Longleftrightarrow$ (f のあるバージョン $\tilde{f}$ が存在して)

$\tilde{f} \in \mathcal{H}$ が確率1で成り立つ.

//

(確認) KL-展開を用いて.

$$\|f\|_{\mathcal{H}}^2 = \sum_{i=1}^{\infty} \xi_i^2 \, \mathbb{1}(\mu_i \neq 0)$$

と書ける.

無限次元: $\forall i$ で $\mu_i \neq 0$ なら $\displaystyle\sum_{i=1}^{\infty} \xi_i^2 = \infty$ (a.s.)

無限

有限次元: $\exists I$ で $\mu_i = 0$ $(\forall i > I)$ なら $\displaystyle\sum_{i=1}^{I} \xi_i^2 < \infty$ (a.s.)

有限

例　$[0,1]$ 上の Wiener 過程 $(W_t)_{t\in[0,1]}$

$E[W_t]=0$
$W_0=0$ (a.s.)
$W_t - W_s \sim N(0, t-s)$

$$k(t,s) = E[W_t W_s] = \min(t,s) \quad (t,s\in[0,1])$$

(W_t) は加法過程.

$$\int_0^1 k(t,s)\, e(t)\, dt = \mu e(s)$$

を解くと,

$$\begin{cases} \mu_i = \dfrac{4}{(2i-1)^2\pi^2} \\ e_i(t) = \sqrt{2}\sin\left(\dfrac{(2i-1)\pi}{2}t\right) \end{cases}$$

$(i=1,2,\dots)$

が 固有値, 固有関数になる.
よって. KL-展開より. Wiener 過程は

$$W_t = \sum_{i=1}^{\infty} \sqrt{\mu_i}\, e_i(t)\cdot \xi_i = \sum_{i=1}^{\infty} \frac{2\sqrt{2}}{\pi(2i-1)} \sin\left(\frac{(2i-1)\pi}{2}t\right)\xi_i$$

for $\xi_i \sim N(0,1)$ (i.i.d.).

なる表現を持つ.

○ ガウス過程回帰

$$y_i = f^{\circ}(x_i) + \varepsilon_i$$

$(\varepsilon_i \sim N(0, \sigma^2))$

なるモデルを考える. $D_n = (x_i, y_i)_{i=1}^{n}$ を用いて f° を推定したい.

ベイズ事前分布 $\pi(df)$: ガウス過程 $GP(0, k)$

f の尤度 : $\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - f(x_i))^2\right)$

$\Longrightarrow$ ベイズ事後分布

$$\pi(df \mid D_n) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - f(x_i))^2\right) \pi(df)$$

ある $x \in \mathcal{X}$ に対し.

$$k_{x,D^n} := (k(x, x_1), \dots, k(x, x_n))^T \in \mathbb{R}^n$$

$$K := (k(x_i, x_j))_{i,j} \in \mathbb{R}^{n \times n}$$

とすると, f の事後分布は

$Y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$

- 平均 : $\hat{m}(x) = k_{x,D^n}^T (K + \sigma^2 I)^{-1} Y$
- 共分散 : $\hat{k}(x, x') = k(x, x') - k_{x,D^n}^T (K + \sigma^2 I)^{-1} k_{x,D^n}$

なるガウス過程で与えられる. (全てガウス分布の条件付き確率等の計算で求まる.)

平均に注目すると, これはカーネルリッジ回帰の解と一致する.

$$\hat{\alpha} = \underset{\alpha \in \mathbb{R}^n}{\mathrm{argmin}} \sum_{i=1}^{n} (y_i - k_{x_i,D^n} \alpha)^2 + \lambda \alpha^T K \alpha$$

$\Rightarrow \hat{\alpha} = (K + \lambda I)^{-1} Y$ より,

$$\hat{f}(x) = \sum_{i=1}^{n} k(x, x_i) \hat{\alpha}_i$$
$$= k_{x,D_n}^T (K + \lambda I)^{-1} Y$$
$$= \hat{m}(x)$$

if $\lambda = \sigma^2$