統計解析特論 講義メモ $\boxed{\textbf{3}}$ 正定値カーネル

カーネル関数を構成する.

Definition 1 (正定値カーネル)**.** 以下を満たす関数 $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ を正定値カーネルという：

1. 対称性： $k(x, x') = k(x', x), \quad {}^\forall x, x' \in \mathcal{X}$
2. 非負定値性： ${}^\forall n = 1, 2, 3, \ldots, \quad {}^\forall x_1, \ldots, x_n \in \mathcal{X}$

$$n \times n \text{ 対称行列 } \quad K = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix} \quad \text{は非負定値行列}$$

$$\iff \quad \sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0, \quad \forall n \in \mathbb{N}, \quad \forall c_1, \ldots, c_n \in \mathbb{R}.$$

K をグラム行列という.

Mercerの定理

(大雑把には) $k(x, x')$ が正定値カーネルなら，関数 $x \mapsto \phi(x) \in \mathbb{R}^D$ が存在して

$$k(x, x') = \phi(x)^T \phi(x')$$

となる ($D = \infty$ もありえる).

- L_2 収束
- 絶対一様収束

参考文献：赤穂本，福水本.

Mercerの定理の解釈：
$k(x, x')$ が正定値カーネルなら $k(x, x') = \sum_i \phi_i(x)\phi_i(x')$.
$\Longleftrightarrow$ 統計モデル $\sum_i \phi_i(x)\theta_i$ を仮定して推定.

例 **1.** 線形，多項式，ガウシアンカーネルは正定値カーネル．
多項式 $(1+x^T x')^k$ はあとで証明

ガウシアン，ラプラス：**Bochner** の定理で証明： 積分表現を与える．

$$e^{-\gamma(x-y)^2} = \frac{1}{2\sqrt{\pi\gamma}} \int_{-\infty}^{\infty} e^{i(x-y)w} e^{-w^2/(4\gamma)} dw$$

$$e^{-\gamma|x-y|} = \frac{\gamma}{\pi} \int_{-\infty}^{\infty} e^{i(x-y)w} \frac{1}{w^2+\gamma^2} dw$$

正定値カーネル性質

- 関数 $g : \mathcal{X} \to \mathbb{R}$ に対して $g(x)g(x')$ は正定値カーネル.
- $k(x, x'),\ k_1(x, x'),\ k_2(x, x')$ が正定値カーネルのとき

1. 非負定数 $a, b \geq 0$ に対して $a + b \cdot k(x, x')$ も正定値カーネル.

2. $k_1(x, x') + k_2(x, x')$ も正定値カーネル.

3. $k_1(x, x')k_2(x, x')$ も正定値カーネル.
したがって $k(x, x')^p$, $p \in \mathbb{N}$ もカーネル関数.

- $k_1, k_2, \ldots, k_n, \ldots,$ が全て正定値カーネルのとき

$$k = \lim_{n \to \infty} k_n \quad \text{（各点収束）も正定値カーネル}$$

カーネル関数から定まる統計モデル

正定値カーネル $k(x, x')$, $x, x' \in \mathcal{X}$ から定義される関数の集合：

$$\widetilde{\mathcal{H}}_k = \left\{ \sum_{i=1}^{N} k(\cdot, \boldsymbol{z}_i)\beta_i \,\middle|\, N \in \mathbb{N}, \boldsymbol{z}_i \in \mathbb{R}^d, \beta_i \in \mathbb{R} \right\}.$$

note: 線形モデル $\{\sum_i \phi_i(x)\theta_i \,|\, \theta_i \in \mathbb{R}\}$ の別表現.

$\widetilde{\mathcal{H}}_k$ に内積を定義する：長さや角度 (直交性) を定める.
推定量の記述，解釈に役立つ.

内積

1. 対称性：$\langle f, g \rangle = \langle g, f \rangle$
2. 双線形性：$\langle af + bg, h \rangle = a\langle f, h \rangle + b\langle g, h \rangle$
3. 正値性：$\langle f, f \rangle \geq 0$
4. $\langle f, f \rangle = 0 \Leftrightarrow f = 0$.

- $\widetilde{\mathcal{H}}$ 上の内積：

$$\langle k(\cdot, x), k(\cdot, x')\rangle = k(x, x') \quad \& \quad \text{双線形性}.$$

$\|f\| = \sqrt{\langle f, f\rangle}$.

- $\langle , \rangle$ の定義と正定値カーネルの性質から **1,2,3** 成立.
- **1,2,3** $\Rightarrow$ シュワルツ不等式 $|\langle f, g\rangle| \leq \|f\|\|g\|$ $\Rightarrow$ **4**
- 上の内積は $f \in \widetilde{\mathcal{H}}$ の表し方に依らない **(**ことを言う必要がある**)**

例 **2.**

$$\langle k(\cdot, x_1) - 2k(\cdot, x_2), k(\cdot, x_3)\rangle = k(x_1, x_3) - 2k(x_2, x_3)$$

$$\|k(\cdot, x) - k(\cdot, x')\|^2 = k(x, x) + k(x', x') - 2k(x, x')$$

再生性

$f \in \widetilde{\mathcal{H}}_k$, $x \in \mathcal{X}$ に対して以下が成り立つ：

$$\langle f, k(\cdot, x) \rangle = f(x).$$

Proof. $f(\cdot) = \sum_{i=1}^{N} k(\cdot, z_i)\beta_i$ とすると

$$\begin{aligned}
\langle f, k(\cdot, x) \rangle &= \sum_{i=1}^{N} \beta_i \langle k(\cdot, z_i), k(\cdot, x) \rangle \\
&= \sum_{i=1}^{N} \beta_i k(z_i, x) = \sum_{i=1}^{N} k(x, z_i)\beta_i = f(x)
\end{aligned}$$

■

再生性を使うと $\langle\cdot,\cdot\rangle$ が内積であることが証明できる.
$\|f\| = 0$ のとき，任意の $x \in \mathcal{X}$ に対して

$$|f(x)| = |\langle f, k(\cdot, x)\rangle| \le \|f\|\|k(\cdot, x)\| = \|f\|\sqrt{k(x, x)} = 0$$

より f は $\mathcal{X}$ 上の零関数となる．よって $\|f\| = 0$ なら $f = 0$ となる.

線形モデルとの対応：

$$\begin{aligned}
\boldsymbol{x} &\mapsto k(\boldsymbol{x}, \cdot) \in \widetilde{\mathcal{H}}_k, \\
\boldsymbol{w} &\mapsto f \in \widetilde{\mathcal{H}}_k, \\
\boldsymbol{w}^T\boldsymbol{x} &\mapsto \langle f, k(\boldsymbol{x}, \cdot)\rangle = f(\boldsymbol{x})
\end{aligned}$$

§ 再生核ヒルベルト空間

Definition 2 (再生核ヒルベルト空間). 内積空間 $(\widetilde{\mathcal{H}}_k, \langle \cdot, \cdot \rangle)$ を完備化 $\Longrightarrow$ 再生核ヒルベルト空間 $(\mathcal{H}_k, \langle \cdot, \cdot \rangle)$, $(\widetilde{\mathcal{H}}_k \subset \mathcal{H}_k)$.

Reproducing kernel Hilbert space：RKHS と略す.

完備化：コーシー列が収束列になるように「穴を埋めた」空間.
（例：$\mathbb{Q}$ を完備化 $\Rightarrow \mathbb{R}$）

- $\mathcal{H}_k$ に対しても再生性が成り立つ：${}^\forall f \in \mathcal{H}_k,\ \langle f, k(x, \cdot) \rangle = f(x)$.
  完備化の定義から証明できる (証明略).
- (完備化しても) $f \in \mathcal{H}_k$ は $\mathcal{X}$ から $\mathbb{R}$ への関数として表現できる (証明略).

カーネル関数を用いた推定

1. カーネル関数 $k(x, x')$ を定める
2. $k(x, x')$ から **RKHS** $\mathcal{H}_k$ が定まる
3. データから回帰関数 $f(x) + b \in \mathcal{H}_k + \mathbb{R}$ を推定する.

カーネル関数 k を用いて推定
$\Longleftrightarrow$ 統計モデルとして **RKHS** $\mathcal{H}_k$ を仮定

正定値カーネルに対応する関数空間

- m 次多項式カーネル：$k(x,x') = (1+xx')^m$, $x, x' \in \mathbb{R}$.
  $\mathcal{H}_k$：m 次以下の多項式のなす線形空間
- ガウシアン・カーネル：$k(x,x') = \exp\{-\gamma(x-x')^2\}$, $x, x' \in \mathbb{R}$.

$$\mathcal{H}_k = \left\{ f \in L_2(\mathbb{R}) \,\middle|\, \int |\mathcal{F}(f)(w)|^2 e^{w^2/(4\gamma)} dw < \infty \right\}$$

$\mathcal{F}(f)$: f のフーリエ変換．周波数成分が指数的に減衰する関数の集合

表現定理

データ $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ に対して，学習アルゴリズム

$$\min_{f,b} L(f(x_1), \ldots, f(x_n), b; D) + \Omega(\|f\|^2), \quad f \in \mathcal{H}_k, \quad b \in \mathbb{R}$$

を考える．Ω は単調増加関数とする **(**$\Omega(\|f\|^2)$ は正則化項に対応**)**.
このとき f の最適解は

$$f(\cdot) = \sum_{i=1}^{n} k(\cdot, x_i)\beta_i, \quad \beta_i \in \mathbb{R}$$

と表せる.

Proof. $S = \mathrm{span}\,\{k(\cdot, x_1),\, k(\cdot, x_2),\, \ldots,\, k(\cdot, x_n)\}$ とする．
直交補空間を $\mathcal{K}^{\perp}$ とする． $f = f_1 + f_2,\ f_1 \in \mathcal{S},\ f_2 \in \mathcal{S}^{\perp}$． このとき

$$
\begin{aligned}
& f(x_i) = \langle f_1 + f_2, k(\cdot, x_i)\rangle = \langle f_1, k(\cdot, x_i)\rangle = f_1(x_i), \\
& \|f\|^2 = \|f_1\|^2 + \|f_2\|^2 \geq \|f_1\|^2.
\end{aligned}
$$

したがって

$$
\begin{aligned}
& L(f(x_1), \ldots, f(x_n), b; D) + \Omega(\|f\|^2) \\
\geq\ & L(f_1(x_1), \ldots, f_1(x_n), b; D) + \Omega(\|f_1\|^2).
\end{aligned}
$$

すなわち，部分空間 $\mathrm{span}\{k(\cdot, x_1), \ldots, k(\cdot, x_n)\}$ 上に f の最適解が存在する． ■

$$
\begin{aligned}
&\min_{f,b} L(f(x_1), \ldots, f(x_n), b; D) + \Omega(\|f\|^2), \quad f \in \mathcal{H}_k \\
\Leftrightarrow\ &\min_{\beta,b} L\Big(\sum_{i=1}^{n} k(x_1, x_i)\beta_i, \ldots, \sum_{i=1}^{n} k(x_n, x_i)\beta_i, b; D\Big) + \Omega(\boldsymbol{\beta}^\top K \boldsymbol{\beta})
\end{aligned}
$$

ただし $K_{ij} = k(x_i, x_j)$.

- 表現定理：
  $\mathcal{H}_k$ が無限次元でも，有限次元の最適化問題を解くことで，厳密な最適解が得られる.
- さまざまな損失関数で成立

$k(x,x')$：が正定値カーネルでないとき：

$$\text{統計モデル：}\ \mathcal{H}' = \left\{ f(x) = \sum_{i=1}^{N} k(x,z_i)\beta_i \,\middle|\, N\in\mathbb{N},\ \beta_i\in\mathbb{R},\ z_i\in\mathcal{X} \right\}.$$

- 正則化：$\langle f,f\rangle$ ではなく $\sum_k \beta_k^2$ など.
- 内積空間と対応せず，統計的性質を調べにくい.

カーネルパラメータの選択

カーネルパラメータ：カーネル関数を決めるパラメータ.

- 多項式カーネルの次数 $m \in \mathbb{N}$.
- ガウシアン・カーネルの幅パラメータ $\gamma > 0$.

正則化パラメータ，カーネルパラメータ：K-**fold CV** などで決める.

カーネルパラメータの選択：ガウシアン・カーネル

- 正則化パラメータの候補：$\lambda_1, \ldots, \lambda_p$
- カーネル幅パラメータの候補：$\gamma_1, \ldots, \gamma_q$
- (λ_i, γ_j) を用いて回帰関数を学習する．このときの予測誤差の推定値を **cross validatoin** で計算：$\widehat{e}_{\mathrm{CV}}(\lambda_i, \gamma_j)$.
- $\widehat{e}_{\mathrm{CV}}(\lambda_i, \gamma_j)$, $i = 1, \ldots, p, j = 1, \ldots, q$. のなかで最小値を達成する (λ, γ) を用いた回帰関数を推定量とする.

K-**fold CV** を用いるとき pqK 回の推定を行う.

ヒューリスティクス(発見的解法)

- 交差検証法は一般に計算コストが大きい.
- モデルパラメータが複数ある場合：以下のような工夫をする
  - 並列計算
  - いくつかのパラメータを適当な値に固定,
    残りのパラメータを交差検証法で決める.

ガウシアンカーネルの γ の選び方

ガウシアンカーネルの γ をヒューリスティックに決める.

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\{-\gamma \|\boldsymbol{x} - \boldsymbol{x}'\|^2\}$$

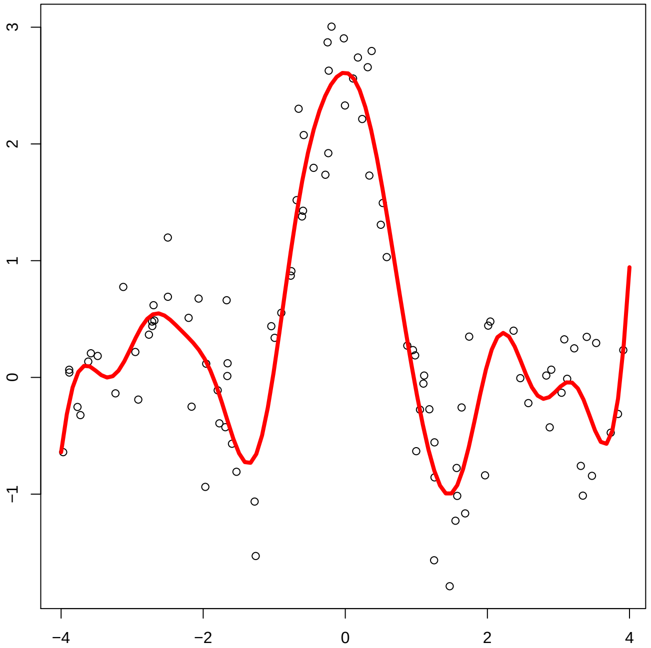
- 計算の安定化のため，$\gamma\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2$ が「ほどほどの値」を取るようにする.

$$d^2 := \mathtt{median}\{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 \,:\, i < j\} \quad (\text{中央値})$$

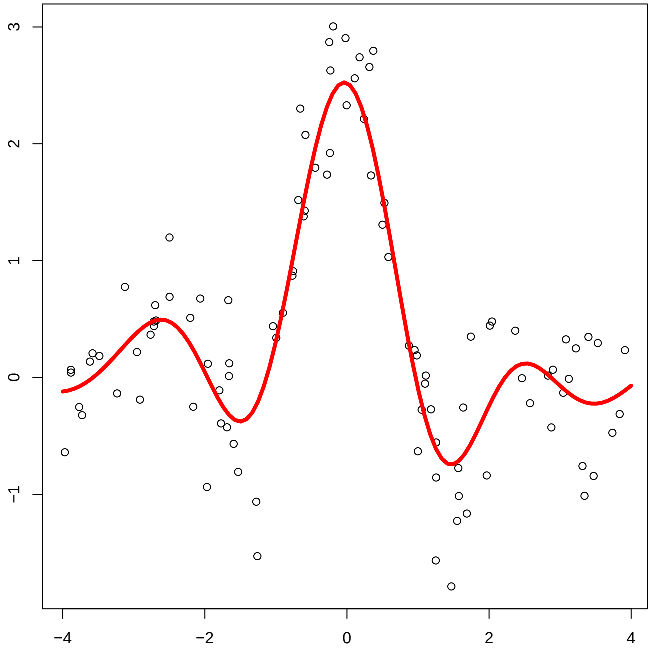
として $\gamma = \dfrac{1}{d^2}$ とする
  - $\gamma\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2$ は 1 のまわりに分布
  - グラム行列の数値計算が安定

推定結果

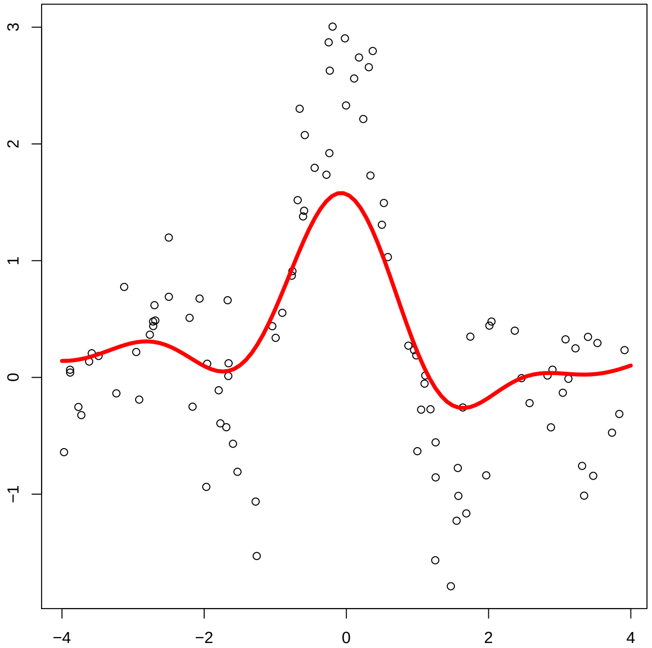
- カーネルパラメータ：$\texttt{sigma} = 0.98$ **(median-heuristics)**
- 正則化パラメータ λ：いろいろ変える.



$(\texttt{sigma}, \lambda) = (0.98, 0.1^5)$



$(\texttt{sigma}, \lambda) = (0.98, 0.78)$



$(\texttt{sigma}, \lambda) = (0.98, 10)$