第6章　正則化とカーネル法

6.1　導入

訓練データの実現値を $\mathfrak{t}_n = \{(\boldsymbol{x}_j, y_j);\, j = 1, 2, \ldots, n\}$ とする．ただし $\boldsymbol{x}_j \in \mathbb{R}^d$, $y_j \in \mathbb{R}$ である．予測関数 $g_{\mathfrak{t}}$ をみつけるために探索する関数族 $\mathcal{G} = \{g : \mathbb{R}^d \mapsto \mathbb{R}\}$ を定めて, $g \in \mathcal{G}$ に関して

$$\widehat{\mathsf{Err}}_{\mathfrak{t}}(g) := \frac{1}{n}\sum_{j=1}^{n}\bigl(y_j - g(\boldsymbol{x}_j)\bigr)^2$$

を最小にする関数 $g_{\mathfrak{t}} \in \mathcal{G}$ をみつけたい．関数族 $\mathcal{G}$ をすべての $\mathbb{R}^d$ 上の実数値関数全体がなす集合とすると

$$g(\boldsymbol{x}_j) = y_j \quad (j = 1, 2, \ldots, n)$$

をみたす任意の関数 g は $\widehat{\mathsf{Err}}_{\mathfrak{t}}(g) = 0$ をみたす．しかしこの g の汎化誤差はよいことが期待できない．いわゆる**過学習**が起きていることになる．

予測の観点から問題を考える．(X, Y) は $(\boldsymbol{x}_j, y_j)$ を発生させて確率分布と同じ分布に従う確率ベクトルとする．任意の関数 $g : \mathbb{R}^d \longrightarrow \mathbb{R}$ に対して, 2 乗誤差

$$\mathsf{E}\bigl[\{Y - g(X)\}^2\bigr]$$

を考える．このとき 2 乗誤差を最小にする g は

$$g^*(\boldsymbol{x}) = \mathsf{E}\bigl[Y \bigm| X = \boldsymbol{x}\bigr]$$

で与えられた．関数族 $\mathcal{G}$ は解釈が容易であるような簡単なもので, かつ最適の g^* を含むような豊かなものがが望ましい．

関数族 $\mathcal{G}$ をよりよく理解するために $\mathcal{G}$ を Hilbert 空間 (完備な内積空間) とみなす．たとえば$\mathcal{G}$ を $\mathbb{R}^d$ 上の線型関数すべてのなす集合とすれば, これは Hilbert 空間とみなすことができる．$\mathcal{G} = \{g_\beta(\boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{\beta};\, \boldsymbol{\beta} \in \mathbb{R}^d\}$ の元 g_β と $\boldsymbol{\beta} \in \mathbb{R}^d$ を同一視し, $\mathcal{G}$ 上の内積を

$$\langle g_{\beta_1}, g_{\beta_2}\rangle = \boldsymbol{\beta}_1^\top\boldsymbol{\beta}_2,$$

で定義する．ただし $g_{\beta_1}(\boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{\beta}_1$, $g_{\beta_2}(\boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbb{R}^d$ である．すなわち $\langle\cdot, \cdot\rangle$ は $\mathbb{R}^d$ 上の Euclid 内積である．するとこの内積 $\langle\cdot, \cdot\rangle$ をもつ $\mathcal{G}$ は Hilbert 空間になること[1]がわかる．

[1] これは実数の完備性より直ちに確認できる．

ex:6-1 **例 6.1.** () 多項式回帰訓練データの実現値を

$$\mathfrak{t}_n = \{(u_1,\, y_1),\, (u_2,\, y_2),\, \ldots,\, (u_n,\, y_n)\}$$

とする．ただし $u_j,\, y_j \in \mathbb{R}\,(j = 1,\, 2,\, \ldots,\, n)$ である．$d \in \mathbb{N},\, d \geq 2$ とし，変換

$$\boldsymbol{\phi}(u) := \begin{pmatrix} 1 \\ u \\ u^2 \\ \vdots \\ u^{d-1} \end{pmatrix},\, \mathbb{R} \longrightarrow \mathbb{R}^d$$

を考える．すると $\boldsymbol{\beta} \in \mathbb{R}^d$ としたとき，関数

$$g_\beta \colon u \mapsto \boldsymbol{\phi}(u)^\top \boldsymbol{\beta}$$

全体のなす集合を $\mathcal{G}$ としたとき，対応

$$h_\beta \longleftrightarrow \boldsymbol{\beta}$$

と $\mathbb{R}^d$ 上の Euclid 内積により $\mathcal{G}$ は Hilbert 空間となる． □

回帰問題において線型空間では単純すぎるが，複雑すぎる関数の利用は避けたいところである．そのために，問題そのものを「代入が内積で表現される空間 (再生核 Hilbert 空間)」に変換しその変換先での線型関数を考える方法が**カーネル法**である．カーネル法では，半正定値行列を一般化した**カーネル関数**とよばれる 2 変数関数 k を考え，データ $\boldsymbol{x}_j$ を $k_{\boldsymbol{x}_j} = k(\,\cdot\,, \boldsymbol{x}_j)$ に変換することを**カーネルトリック**とよぶ．特に写像

$$\phi \colon \boldsymbol{x}_j \mapsto k_{\boldsymbol{x}_j}$$

は**特徴写像**とよばれる．データを $\mathbb{R}^d$ とは曲り具合の異なる高次元空間に埋め込むことに相当する．この高次元空間のことを**特徴空間**とよぶ．

6.2 正則化

sec:6-2

$d \in \mathbb{N},\, d \geq 2$ とする． 訓練データの実現値を $\mathfrak{t}_n = \{(\boldsymbol{x}_j,\, y_j);\, j = 1,\, 2,\, \ldots,\, n\}$ とする．ただし $\boldsymbol{x}_j \in \mathbb{R}^d,\, y_j \in \mathbb{R}$ とする．$\mathbb{R}^d$ 上の実数値関数の族 $\mathcal{G}$ に関してある損失関数 $\mathsf{Loss} : \mathbb{R}^2 \longrightarrow \mathbb{R}$ を用いた訓練誤差

$$\widehat{\mathsf{Err}}_{\mathfrak{t}}(g) := \frac{1}{n} \sum_{j=1}^{n} \mathsf{Loss}\bigl(g(\boldsymbol{x}_j),\, y_j\bigr) \quad (g \in \mathcal{G})$$

を最小にする $g_\mathfrak{t} \in \mathcal{G}$ をみつけたい. 関数族 $\mathcal{G}$ はある内積に関して Hilbert 空間になっているとする. $\mathcal{G}$ が豊かすぎると訓練誤差は零または零に近くなり, 過学習を起こす. 過学習をさけるためのひとつの方法はモデルの複雑さに罰則

$$J : \mathcal{G} \longrightarrow [0, \infty) \cup \{+\infty\}$$

を導入し, 罰則付きの訓練誤差の最小化問題を考えることである. すなわち

$$\min_{g \in \mathcal{G}} \big\{ \widehat{\mathsf{Err}}_\mathfrak{t}(g) + J(g) : \, g \in \mathcal{G} \big\}$$

である.

ex:6-2 **例 6.2.** (リッジ回帰) リッジ回帰は 2 上ノルムの罰則項をもった線型回帰である. $d, n \in \mathbb{N}$ とする. 訓練データの実現値を $\mathfrak{t}_n = \{(\boldsymbol{x}_j, y_j);\, j = 1, 2, \ldots, n\}$ とする. ただし $\boldsymbol{x}_j \in \mathbb{R}^d$, $y_j \in \mathbb{R}$ である. 関数族

$$\mathcal{G} = \{ g(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta} : \, \boldsymbol{\beta} \in \mathbb{R}^d \}$$

を考え, $\gamma > 0$ に対して

$$\min_{g \in \mathcal{G}} \left\{ \frac{1}{n} \sum_{j=1}^{n} \big(y_j - g(\boldsymbol{x}_j) \big)^2 + \gamma \|g\|^2 \right\} \tag{6.1}$$

eq:6-1

を考える. ただし $\|\cdot\|$ は $\mathcal{G}$ の内積から定義される $\mathcal{G}$ 上のノルムである. $g \in \mathcal{G}$ に対して $\boldsymbol{\beta} \in \mathbb{R}^d$ を対応させると

$$\|g\|^2 = \langle \boldsymbol{\beta}, \boldsymbol{\beta} \rangle = \big| \boldsymbol{\beta} \big|_{2,\,d}^2$$

となる. ただし, $|\cdot|_{2,\,d}$ は Euclid $\mathbb{R}^d$ のノルムである. したがって, 問題は

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^d} \left\{ \frac{1}{n} \big| \boldsymbol{y} - \mathbf{X}\boldsymbol{\beta} \big|_{2,\,n}^2 + \gamma |\boldsymbol{\beta}|_{2,\,d}^2 \right\} = \min_{\boldsymbol{\beta} \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{j=1}^{n} \big(y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta} \big)^2 + \gamma |\boldsymbol{\beta}|_{2,\,d}^2 \right\} \tag{6.2}$$

eq:6-2

と書きなおすことができる. ただし

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \boldsymbol{x}_2^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

とした. 最小化問題 (6.1) の解は

$$\boldsymbol{x} \mapsto \boldsymbol{x}^\top \boldsymbol{\beta}$$

という形となり $\boldsymbol{\beta}^*$ は最小化問題 (6.1) の解である．$\gamma \to \infty$ のとき罰則項が支配的になり $g \equiv 0$ となる．

問題 (6.2) は凸問題 (対象の関数が凸関数) なので勾配を零とおく．

$$h(\boldsymbol{\beta}) := \frac{1}{n}\bigl|\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\bigr|_2^2 + \gamma|\boldsymbol{\beta}|_{2,\,d}^2$$

とおいたとき

$$\frac{\partial h}{\partial \boldsymbol{\beta}} = \frac{1}{n}\mathbf{X}^\top\bigl(\mathbf{X}\boldsymbol{\beta} - \boldsymbol{y}\bigr) + \gamma\boldsymbol{\beta} = \mathbf{0} \tag{6.3}$$ eq:6-3

とおく．$\gamma = 0$ のとき上の方程式は正規方程式になる．$\mathbf{X}^\top\mathbf{X} + \boldsymbol{I}_d$ が正則のとき

$$\widehat{\boldsymbol{\beta}}_\gamma := \bigl(\mathbf{X}^\top\mathbf{X} + \gamma\boldsymbol{I}_d\bigr)^{-1}\mathbf{X}^\top\boldsymbol{y}$$

が方程式 (6.3) の解となる．$\widehat{\boldsymbol{\beta}}_\gamma$ のことを**リッジ回帰推定量**という． □

ある Hilbert 空間 $\mathcal{G}$ に対して罰則項を用いるとき $\mathcal{G}$ を 2 つの直交部分空間 $\mathcal{H}$ と $\mathcal{C}$ に分解するとよい．すなわち $\forall g \in \mathcal{G}$ に対して, ある $h \in \mathcal{H}$ と $c \in \mathcal{C}$ が一意的に存在して

$$g = h + c, \quad \langle h,\, c\rangle_{\mathcal{G}} = 0$$

となる．ただし $\langle\cdot,\,\cdot\rangle_{\mathcal{G}}$ は Hilbert 空間 $\mathcal{G}$ の内積である．Hilbert 空間 $\mathcal{G}$ の任意の元がこのような分解をもつとき, $\mathcal{G}$ は $\mathcal{H}$ と $\mathcal{C}$ の直和であるといい, $\mathcal{G} = \mathcal{H} \oplus \mathcal{C}$ と書く．

ex:6-3 **例 6.3.** (例 6.2 の続き)

$$\widetilde{\boldsymbol{x}}_j = \bigl(1,\, \boldsymbol{x}_j^\top\bigr)^\top \quad (j = 1,\, 2,\, \ldots,\, n)$$

とする．第 1 成分に対応する回帰係数 (切片) には罰則を加えない．$\widetilde{\boldsymbol{x}} = \bigl(1,\, \boldsymbol{x}^\top\bigr)^\top$ とおいたとき

$$\mathcal{G} := \bigl\{g(\widetilde{\boldsymbol{x}}) = \beta_0 + \boldsymbol{x}^\top\boldsymbol{\beta};\, \beta_0 \in \mathbb{R},\, \boldsymbol{\beta} \in \mathbb{R}^d\bigr\}$$

とし

$$\widehat{\mathsf{Err}}(g) = \frac{1}{n}\bigl|\boldsymbol{y} - \widetilde{\mathbf{X}}\widetilde{\boldsymbol{\beta}}\bigr|_{2,\,n}^2 + \gamma|\boldsymbol{\beta}|_{2,\,d}^2$$

の最小化問題を考える．ただし $\gamma > 0$ で

$$\widetilde{\mathbf{X}} = \bigl(\mathbf{1},\, \mathbf{X}\bigr), \quad \widetilde{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta} \end{pmatrix}, \quad \mathbf{1} = \bigl(\underbrace{1,\, 1,\, \ldots,\, 1}_{n}\bigr)^\top$$

である．すると $\gamma \to \infty$ のとき $g(\widetilde{\boldsymbol{x}}) = \beta_0$ となる．$g \in \mathcal{G}$ は

$$\widetilde{\boldsymbol{x}} \mapsto \beta_0 + \boldsymbol{x}^\top\boldsymbol{\beta}$$

である．さらに関数族 $\mathcal{C}$ を

$$c : \widetilde{\boldsymbol{x}} \mapsto \beta_0$$

とし，$\mathcal{H}$ を

$$h : \widetilde{\boldsymbol{x}} \mapsto \boldsymbol{x}^\top \boldsymbol{\beta}$$

とする．このとき $\mathcal{G}$ の内積を $\bigl(1,\, \boldsymbol{\beta}^\top\bigr)^\top$ に対する $\mathbb{R}^{d+1}$ 上の Euclid 内積とすれば

$$\mathcal{G} = \mathcal{C} \oplus \mathcal{H}$$

となる．

　よって

$$\min_{g \in \mathcal{H} \oplus \mathcal{C}} \left\{ \frac{1}{n} \sum_{j=1}^{n} \bigl(y_j - g(\widetilde{\boldsymbol{x}})\bigr)^2 + \gamma \|g\|_{\mathcal{H}} \right\} \tag{6.4}$$ `eq:6-4`

となる．ただし $\|\cdot\|_{\mathcal{H}}$ は $\mathcal{H}$ のノルムである．問題 (6.4) は

$$\min_{\beta_0,\, \boldsymbol{\beta}} \left\{ \frac{1}{n} \bigl|\boldsymbol{y} - \beta_0 \mathbf{1} - \mathbf{X}\boldsymbol{\beta}\bigr|_{2,\,n}^2 + \gamma |\boldsymbol{\beta}|_{2,\,d}^2 \right\} =: \min_{\beta_0,\, \boldsymbol{\beta}} h\bigl(\beta_0,\, \boldsymbol{\beta}\bigr) \tag{6.5}$$ `eq:6-5`

と書きなおすことができる．$\gamma \to \infty$ のとき

$$g(\widetilde{\boldsymbol{x}}) \longrightarrow \overline{y}_n = \frac{1}{n} \sum_{j=1}^{n} y_j$$

となる．問題 (6.5) も凸なので

$$\frac{\partial h}{\partial \boldsymbol{\beta}}(\beta_0,\, \boldsymbol{\beta}) = \mathbf{X}^\top \bigl(\beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} - \boldsymbol{y}\bigr) + n\gamma\boldsymbol{\beta} = \mathbf{0} \tag{6.6}$$ `eq:6-6`

$$\frac{\partial h}{\partial \beta_0}(\beta_0,\, \boldsymbol{\beta}) = n\beta_0 - \mathbf{1}^\top \bigl(\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\bigr) = 0 \tag{6.7}$$ `eq:6-7`

を解けばよい．(6.7) の解を (6.6) に代入すれば $\boldsymbol{\beta}$ をみつけるためには方程式

$$\bigl(\mathbf{X}^\top\mathbf{X} - n^{-1}\mathbf{X}^\top\mathbf{1}\mathbf{1}^\top\mathbf{X} + n\gamma\boldsymbol{I}_d\bigr)\boldsymbol{\beta} = \bigl(\mathbf{X}^\top - n^{-1}\mathbf{X}^\top\mathbf{1}\mathbf{1}^\top\bigr)\boldsymbol{y} \tag{6.8}$$ `eq:6-8a`

を解けばよい．

　議論を簡単にするために $n \geq d$ かつ $\operatorname{rank}\mathbf{X} = d$ $\bigl(\Longleftrightarrow$ $\mathbf{X}^\top\mathbf{X}$ は正則$\bigr)$ を仮定する．すると

$$\begin{aligned}
(6.6) &\Longleftrightarrow \beta_0\mathbf{X}^\top\mathbf{1} + \mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} - \mathbf{X}^\top\boldsymbol{y} + n\gamma\boldsymbol{\beta} = \mathbf{0} \\
&\Longleftrightarrow \bigl(\mathbf{X}^\top\mathbf{X} + n\gamma\boldsymbol{I}_d\bigr)\boldsymbol{\beta} = \mathbf{X}^\top\bigl(\boldsymbol{y} - \beta_0\mathbf{1}\bigr)
\end{aligned}$$

なので $\boldsymbol{\beta}$ は $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ の一次結合で書ける．よって

$$\boldsymbol{\beta} = \mathbf{X}^\top \boldsymbol{\alpha}, \quad \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^\top \in \mathbb{R}^n$$

とおく．これを (6.8) に代入すれば eq:6-8a

$$\big(\mathbf{X}\mathbf{X}^\top - n^{-1}\mathbf{1}\mathbf{1}^\top \mathbf{X}\mathbf{X}^\top + n\gamma \boldsymbol{I}_n\big)\boldsymbol{\alpha} = \big(\boldsymbol{I}_n - n^{-1}\mathbf{1}\mathbf{1}^\top\big)\boldsymbol{y}$$

を得る．$\mathbf{X}\mathbf{X}^\top - n^{-1}\mathbf{1}\mathbf{1}^\top \mathbf{X}\mathbf{X}^\top + n\gamma \boldsymbol{I}_n$ は正則と仮定すると

$$\widehat{\boldsymbol{\alpha}} = \big(\mathbf{X}\mathbf{X}^\top - n^{-1}\mathbf{1}\mathbf{1}^\top \mathbf{X}\mathbf{X}^\top + n\gamma \boldsymbol{I}_n\big)^{-1}\big(\boldsymbol{I}_n - n^{-1}\mathbf{1}\mathbf{1}^\top\big)\boldsymbol{y} \tag{6.9}$$ eq:6-8

となる．ここで

$$\mathbf{X}\mathbf{X}^\top = \begin{pmatrix} \langle \boldsymbol{x}_1, \boldsymbol{x}_1\rangle_{2,d} & \langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle_{2,d} & \cdots & \langle \boldsymbol{x}_1, \boldsymbol{x}_n\rangle_{2,d} \\ \langle \boldsymbol{x}_2, \boldsymbol{x}_1\rangle_{2,d} & \langle \boldsymbol{x}_2, \boldsymbol{x}_2\rangle_{2,d} & \cdots & \langle \boldsymbol{x}_2, \boldsymbol{x}_n\rangle_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{x}_n, \boldsymbol{x}_1\rangle_{2,d} & \langle \boldsymbol{x}_n, \boldsymbol{x}_2\rangle_{2,d} & \cdots & \langle \boldsymbol{x}_n, \boldsymbol{x}_n\rangle_{2,d} \end{pmatrix}$$

となることに注意すると $\widehat{\boldsymbol{\alpha}}$ は $\{\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle_{2,d}; (i, j = 1, 2, \ldots, n)\}$ のみを通して $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ に依存することがわかる．すなわち 各 $\boldsymbol{x}_j\,(j = 1, 2, \ldots, n)$ の値がわからなくとも $\{\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle_{2,d}; (i, j = 1, 2, \ldots, n)\}$ の値がわかれば $\widehat{\boldsymbol{\alpha}}$ の値は計算できる．さらに (6.9) を (6.7) に代入すれば eq:6-8 eq:6-7

$$\widehat{\beta}_0 = \frac{1}{n}\mathbf{1}^\top\big(\boldsymbol{y} - \mathbf{X}\mathbf{X}^\top\widehat{\boldsymbol{\alpha}}\big)$$

となる．よって

$$g_{\mathrm{t}}\big(\widetilde{\boldsymbol{x}}\big) = \widehat{\beta}_0 + \boldsymbol{x}^\top \mathbf{X}^\top \widehat{\boldsymbol{\alpha}} = \widehat{\beta}_0 + \sum_{j=1}^n \widehat{\alpha}_j \langle \boldsymbol{x}, \boldsymbol{x}_j\rangle_{2,d}, \quad \widehat{\boldsymbol{\alpha}} = \big(\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_n\big)^\top$$

となる．$\widehat{\beta}_0$ と $\widehat{\alpha}_j\,(j = 1, 2, \ldots, n)$ は内積 $\{\langle \boldsymbol{x}_i, \boldsymbol{x}_j\rangle_{2,d}; i, j = 1, 2, \ldots, n\}$ の値がわかればよい．さらに $g_{\mathrm{t}}\big(\widetilde{\boldsymbol{x}}\big)$ の値を求めるためには, $\{\langle \boldsymbol{x}_j, \boldsymbol{x}\rangle_{2,d}; j = 1, 2, \ldots, n\}$ の値がわかればよい．内積だけの計算になっていることが肝である．

6.3 カーネル関数

sec:6-3
df:6-4

定義 6.4. $\mathbb{X}$ を空でない集合とし k を $\mathbb{X}$ 上の 2 変数関数とする. k が次の 2 条件をみたすとき, k は $\mathbb{X}$ 上の**カーネル関数**とよばれる.

(i) $\forall x, y \in \mathbb{X}$ に対し

$$k(x, y) = k(y, x). \qquad \text{(対称性)}$$

(ii) $\forall n \in \mathbb{N}$, $\{x_j\}_{j=1}^n \subset \mathbb{X}$, $\{c_j\}_{j=1}^n \subset \mathbb{R}$ に対し

$$\sum_{j,\,\ell=1}^{n} c_j c_\ell k(x_j,\, x_\ell) \geq 0. \qquad (\text{半正定値性})$$

定義 df:6-4 6.4(ii) は, $\forall n \in \mathbb{N}$ と $\{x_j\}_{j=1}^n \subset \mathbb{X}$ に対し

$$\begin{pmatrix} k(x_1,\, x_1) & k(x_1,\, x_2) & \cdots & k(x_1,\, x_n) \\ k(x_2,\, x_1) & k(x_2,\, x_2) & \cdots & k(x_2,\, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n,\, x_1) & k(x_n,\, x_2) & \cdots & k(x_n,\, x_n) \end{pmatrix} \text{は半正定値}$$

となる.

重要なカーネル関数の例をあげておく.

ex:6-5 **例 6.5.** f を $\mathbb{X}$ 上の関数とする. このとき

$$k(x,\, y) = f(x)f(y) \qquad (x,\, y \in \mathbb{X})$$

はカーネル関数となる. 実際, $\forall n \in \mathbb{N}$, $\{x_j\}_{j=1}^n \subset \mathbb{X}$, $\{c_j\}_{j=1}^n \subset \mathbb{R}$ に対し

$$\sum_{j,\,\ell=1}^{n} c_j c_\ell k(x_j,\, x_\ell) = \sum_{j,\,\ell=1}^{n} c_j c_\ell f(x_j) f(x_\ell) = \left(\sum_{j=1}^{n} c_j f(x_j)\right)^2 \geq 0$$

より半正定値性はわかる. 対称性は定義より明らか. □

ex:6-6 **例 6.6.** k_1, k_2 を $\mathbb{X}$ 上のカーネル関数とする. このとき

$$(k_1 + k_2)(x,\, y) := k_1(x,\, y) + k_2(x,\, y) \qquad (x,\, y \in \mathbb{X})$$

はカーネル関数. □

pro:6-1 **問 6.1.** 例 ex:6-6 6.6 の k の対称性と半正定値性を確認せよ.

ex:6-7 **例 6.7.** $d \in \mathbb{N}$, $\mathbb{X}$ を空でない集合とし

$$\phi\colon \mathbb{X} \longrightarrow \mathbb{R}^d$$

を $\mathbb{X}$ から $\mathbb{R}^d$ への写像とする. $\mathbb{R}^d$ 上の任意の内積 $\langle \cdot, \cdot \rangle$ を考える. このとき

$$k(x,\, y) = \langle \phi(x),\, \phi(y) \rangle \qquad (x,\, y \in \mathbb{X})$$

はカーネル関数である. 上式の右辺の ϕ のことを**特徴写像**, $\mathbb{R}^d$ を**特徴空間**という.

pro:6-1 **問 6.2.** 内積の性質を用いて例 ex:6-7 6.7 の k の対称性と半正定値性を確認せよ.

6.4 カーネル関数から内積空間の構成

sec:6-4

$\mathbb{X}$ を空でない集合とし k を $\mathbb{X}$ 上のカーネル関数とする. カーネル関数 k と $x \in \mathbb{X}$ に対し, $\mathbb{X}$ 上の関数 k_x を

$$k_x : \mathbb{X} \longrightarrow \mathbb{R}, \quad y \mapsto k(x, y)$$

で定める. $\{k_x\}_{x\in\mathbb{X}}$ で張られる部分ベクトル空間 $\mathcal{V}$ を

$$\mathcal{V} := \left\{ \sum_{j=1}^{n} c_j k_{x_j};\ \text{ある } n \in \mathbb{N} \text{ があって}, \{x_j\}_{j=1}^n \subset \mathbb{X}, \{c_j\}_{j=1}^n \subset \mathbb{R} \right\}$$

で定める. $f = \sum_{j=1}^n a_j x_j$, $f = \sum_{j=1}^m b_j y_j \in \mathcal{V}\,(a_j, b_j \in \mathbb{R}, x_j, y_j \in \mathbb{X}, m, n \in \mathbb{N})$ に対し, 写像 $\langle\cdot, \cdot\rangle_{\mathcal{V}} : \mathcal{V} \times \mathcal{V} \longrightarrow \mathbb{R}$ を

$$\langle f, g\rangle_{\mathcal{V}} := \Big\langle \sum_{j=1}^n a_j k_{x_j}, \sum_{j}^{m} b_j k_{y_j} \Big\rangle_{\mathcal{V}} = \sum_{j=1}^n \sum_{\ell=1}^m a_j b_\ell k\big(y_\ell, x_j\big)$$

と定める. するとカーネル関数 k の半正定値性より $\langle\cdot, \cdot\rangle_{\mathcal{V}} \geq 0$ となる. さらに $f, g \in \mathcal{V}$ の表し方に依らず $\langle f, g\rangle_{\mathcal{V}}$ は一意的に定まる. 実際, $x \in \mathbb{X}$ に対し

$$\sum_{j=1}^n a_j k_{x_j}(x) = \sum_{j=1}^{m'} a'_j k_{x'_j}(x), \quad \sum_{j=1}^n b_j k_{y_j}(x), = \sum_{j=1}^{m'} b'_j k_{y'_j}(x)$$

とする. このとき $k(x, y) = k_y(x)$ と $k(x, y) = k(y, x)$ より

$$\begin{aligned}
\sum_{j,\ell} a_j b_\ell k(y_\ell, x_j) &= \sum_j a_j \sum_\ell b_\ell k(x_j, y_\ell) = \sum_j a_j \sum_\ell b_\ell k_{y_\ell}(x_j) \\
&= \sum_j a_j \sum_\ell b'_\ell k_{y'_\ell}(x_j) = \sum_j a_j \sum_\ell b'_\ell k\big(x_j, y'_\ell\big) \\
&= \sum_j a_j \sum_\ell b'_\ell k\big(y'_\ell, x_j\big) = \sum_\ell b'_\ell \sum_j a_j k\big(y'_\ell, x_j\big) \\
&= \sum_\ell b'_\ell \sum_j a_j k_{x_j}(y'_\ell) = \sum_\ell b'_\ell \sum_j a_j\prime k_{x'_j}(y'_\ell) \\
&= \sum_{j,\ell} b'_\ell a'_j k(y'_\ell, x'_j)
\end{aligned}$$

が成り立つ. よって $\langle\cdot, \cdot\rangle_{\mathcal{V}}$ は f, g の表現の仕方によらず f, g で一意的に定まる.

次に, $f, g, h \in \mathcal{V}$ と $\forall\alpha \in \mathbb{R}$ に対し,

(1) $\langle f, g + h\rangle_{\mathcal{V}} = \langle f, g\rangle_{\mathcal{V}} + \langle f, h\rangle_{\mathcal{V}}$,

(2) $\langle f, g\rangle_{\mathcal{V}} = \langle g, f\rangle_{\mathcal{V}}$,

(3) $\langle \alpha f,\, g\rangle_{\mathcal{V}} = \alpha\langle f,\, g\rangle_{\mathcal{V}}$

が成り立つ.
(2), (3) は明らか. (1) だけを示す.
(1) の証明: $f = \sum_{j=1}^{n} a_j k_{x_j}$, $g = \sum_{j=1}^{m} b_j k_{y_j}$, $\gamma k_z\,(a_j,\, b_j,\, z \in \mathbb{R},\, x_j,\, y_j,\, z \in \mathbb{X})$ に対し, $b_{m+1} = \gamma$, $y_{m+1} = z$ とおく. すると

$$\begin{aligned}\langle f,\, g+h\rangle_{\mathcal{V}} &= \sum_{j=1}^{n}\sum_{\ell=1}^{m} a_j b_\ell k\bigl(y_\ell,\, x_j\bigr) + \sum_{j=1}^{n} a_j b_{m+1} k\bigl(y_{m+1},\, x_j\bigr)\\ &= \sum_{j=1}^{n}\sum_{\ell=1}^{m} a_j b_\ell k\bigl(y_\ell,\, x_j\bigr) + \sum_{j=1}^{n} a_j \gamma k\bigl(z,\, x_j\bigr)\\ &= \langle f,\, g\rangle_{\mathcal{V}} + \langle f,\, \gamma k_z\rangle_{\mathcal{V}}\end{aligned}$$

が成り立つ. ここで示した等式を繰り返し用いれば, (1) が得られる.

(4) $\mathcal{V}$ に対し

$$f(x) = \langle f,\, k_x\rangle_{\mathcal{V}} \qquad (x \in \mathcal{V})$$

が成り立つ.

証明: $f = \sum_{j=1}^{m} a_j k_{x_j}$ とおく. このとき

$$\begin{aligned}\langle f,\, k_x\rangle_{\mathcal{V}} &= \Bigl\langle \sum_{j=1}^{m} a_j k_{x_j},\, k_x\Bigr\rangle_{\mathcal{V}} = \sum_{j=1}^{m} a_j \langle k_{x_j},\, k_x\rangle_{\mathcal{V}} = \sum_{j=1}^{m} a_j k(x,\, x_j)\\ &= \sum_{j=1}^{m} a_j k(x_j,\, x) = \sum_{j=1}^{m} a_j k_{x_j}(x) = f(x).\end{aligned}$$

(5) (Cauchy-Schwarz の不等式) $\forall f,\, g \in \mathcal{V}$ に対し

$$\bigl|\langle f,\, g\rangle_{\mathcal{V}}\bigr| \le \langle f,\, f\rangle_{\mathcal{V}}\, \langle g,\, g\rangle_{\mathcal{V}}. \tag{6.10}$$

eq:6-9

等号成立は $f = cg\,(c \in \mathbb{R})$ のとき.

証明: $\forall t \in \mathbb{R}$ に対し

$$\langle tf+g,\, tf+g\rangle_{\mathcal{V}} = t^2\langle f,\, f\rangle_{\mathcal{V}} + 2t\langle f,\, g\rangle_{\mathcal{V}} + \langle g,\, g\rangle_{\mathcal{V}} \ge 0$$

である. $\langle f,\, f\rangle_{\mathcal{V}} = 0$ のとき,

$$2t\langle f,\, g\rangle_{\mathcal{V}} + \langle g,\, g\rangle_{\mathcal{V}} \ge 0$$

となる. t は任意であったので, $\langle f,\, g\rangle_{\mathcal{V}} = 0$. よって, (6.10) は成立.
$\langle f,\, f\rangle_{\mathcal{V}} \ne 0$ のとき,

$$\langle f,\, f\rangle_{\mathcal{V}}\left\{t + \frac{\langle f,\, g\rangle_{\mathcal{V}}}{\langle f,\, f\rangle_{\mathcal{V}}}\right\}^2 + \frac{\langle f,\, f\rangle_{\mathcal{V}}\,\langle g,\, g\rangle_{\mathcal{V}} - \langle f,\, g\rangle_{\mathcal{V}}^2}{\langle f,\, f\rangle_{\mathcal{V}}} \ge 0$$

より

$$\frac{\langle f, f\rangle_{\mathcal{V}}\langle g, g\rangle_{\mathcal{V}} - \langle f, g\rangle_{\mathcal{V}}^2}{\langle f, f\rangle_{\mathcal{V}}} \geq 0 \Longleftrightarrow \left|\langle f, g\rangle_{\mathcal{V}}\right|^2 \leq \langle f, f\rangle_{\mathcal{V}}\langle g, g\rangle_{\mathcal{V}}$$

がわかる.

以上の議論により, k によって構成された, $\mathcal{V}$ は $\langle\cdot, \cdot\rangle_{\mathcal{V}}$ を内積としてもつ内積空間となることがわかった. あとは, $\mathcal{V}$ を完備化することになる. 完備化された空間を $(\mathcal{H}, \langle\cdot, \cdot\rangle_{\mathcal{H}})$ と書くことにする.

定理 6.8. (Moore-Aronszajn の定理) $\mathbb{X}$ 上のカーネル関数 k に対し, $\mathbb{X}$ の再生核 Hilbert 空間 $(\mathcal{H}, \langle\cdot, \cdot\rangle_{\mathcal{H}})$ で以下の 3 条件をみたすものが一意的に存在する.

(1) $\forall x \in \mathbb{X}$ に対し, $k(\cdot, x) \in \mathcal{H}_k$.

(2) $\mathrm{span}\{k(\cdot, x) \in x \in \mathbb{X}\}$ は $\mathcal{H}_k$ の稠密な部分集合. すなわち, $\forall\epsilon > 0$ と $\forall h \in \mathcal{H}$ に対し, ある $v \in \mathcal{V}$ が存在して,

$$\|f - v\|_{\mathcal{H}} < \epsilon$$

となる. ただし, $\|f\|_{\mathcal{H}} = \sqrt{\langle f, f\rangle_{\mathcal{H}}}$.

(3) k は $\mathcal{H}_k$ の再生核. すなわち,

$$\langle f, k(\cdot, x)\rangle_{\mathcal{H}} = f(x) \qquad (\forall x \in \mathbb{X}, \forall f \in \mathcal{H}).$$

Proof. 福水 (2010, pp.18-19). □

命題 6.9. k を空ではない位相空間 $\mathbb{X}$ 上の半正定値カーネル. $\mathcal{H}$ を対応する再生核 Hilbert 空間とする. 関数 $x \mapsto k(x, x)$ は連続で, 任意の $y \in \mathbb{X}$ に対し, $x \mapsto tok(x, y)$ が $x = y$ において連続であれば, $\mathcal{H}$ に属するすべての関数は $\mathbb{X}$ 上で連続である. とくに $\mathbb{X} \times \mathbb{X}$ 上で連続な半正定値カーネルにより定まる再生核 Hilbert 空間は連続関数からなる.

Proof. 福水 (2010, p.19). □

定理 6.10. (リプリゼンタ定理) k を $\mathbb{X}$ 上のカーネル関数とし, $\mathcal{H}$ を k に対応する再生核 Hilbert 空間とする. $\mathfrak{T} = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{X}$ をデータとする. Loss を n 変数の関数で, pen は $\mathbb{R}$ 上の非減少関数, $b \in \mathbb{R}$ とする. この設定で

$$\mathsf{Loss}\big(f(x_1) + b,\ f(x_2) + b,\ \ldots,\ f(x_n) + b\big) + \mathsf{pen}\big(\|f\|_{\mathcal{H}}^2\big)$$

を $\mathcal{H}$ の中で最小にするものは

$$f = \sum_{j=1}^{n} \xi_j k_{x_j} \qquad (\xi_j \in \mathbb{R},\, j = 1,\, 2, \ldots,\, n)$$

と仮定してよい.

Proof. $\mathbf{P}$ を $\mathrm{span}\{k_{x_1},\, k_{x_2},\, \ldots,\, k_{x_n}\} =: \mathcal{V}$ への直交射影とする. $\mathbf{P}$ は, $\langle \mathbf{P}f,\, g\rangle_{\mathcal{H}} = \langle f,\, \mathbf{P}f\rangle_{\mathcal{H}}\,(\forall f,\, g \in \mathcal{H})$ かつ $\mathbf{P}f = f\,(\forall f \in \mathcal{V})$ である. $f \in \mathcal{H}$ に対し

$$f(x_j) = \langle f,\, k_{x_j}\rangle_{\mathcal{H}} = \langle f,\, \mathbf{P}k_{x_j}\rangle_{\mathcal{H}} = \langle \mathbf{P}f,\, k_{x_j}\rangle_{\mathcal{H}} = (\mathbf{P}f)(x_j)$$

となる. したがって,

$$\begin{aligned}&\mathsf{Loss}\bigl(f(x_1)+b,\, f(x_2)+b,\, \ldots,\, f(x_n)+b\bigr) + \mathsf{pen}\bigl(\|f\|_{\mathcal{H}}^2\bigr)\\ &\quad= \mathsf{Loss}\bigl((\mathbf{P}f)(x_1)+b,\, (\mathbf{P}f)(x_2)+b,\, \ldots,\, (\mathbf{P}f)(x_n)+b\bigr) + \mathsf{pen}\bigl(\|f\|_{\mathcal{H}}^2\bigr)\end{aligned}$$

を得る. $\mathcal{V}^{\perp} := \{f \in \mathcal{H};\, \langle f,\, g\rangle = 0\,(\forall g \in \mathcal{V})$ とする. このとき, $\forall \in \mathcal{H}$ に対し,

$$f = \mathbf{P}f + f^{\perp}\,(f^{\perp} \in \mathcal{V}^{\perp})$$

と一意的に分解できるので,

$$\|\mathbf{P}f\|_{\mathcal{H}}^2 \le \|f\|_{\mathcal{H}}^2 \Longrightarrow \mathsf{pen}\bigl(\|\mathbf{P}f\|_{\mathcal{H}}^2\bigr) \le \mathsf{pen}\bigl(\|f\|_{\mathcal{H}}^2\bigr).$$

よって,

$$\begin{aligned}&\mathsf{Loss}\bigl(f(x_1)+b,\, f(x_2)+b,\, \ldots,\, f(x_n)+b\bigr) + \mathsf{pen}\bigl(\|f\|_{\mathcal{H}}^2\bigr) + \mathsf{pen}\bigl(\|f\|_{\mathcal{H}}^2\bigr)\\ &\quad\ge \mathsf{Loss}\bigl((\mathbf{P}f)(x_1)+b,\, (\mathbf{P}f)(x_2)+b,\, \ldots,\, (\mathbf{P}f)(x_n)+b\bigr)\\ &\qquad+ \mathsf{pen}\bigl(\|f\|_{\mathcal{H}}^2\bigr) + \mathsf{pen}\bigl(\|\mathbf{P}f\|_{\mathcal{H}}^2\bigr)\end{aligned}$$

が成り立つので,

$$f = \mathbf{P}f = \sum_{j=1}^{n} \xi_j k_{x_j} \quad (\xi_j \in \mathbb{R},\, j = 1,\, 2,\, \ldots,\, n)$$

としてよい. □

6.5 例

6.5.1 連立方程式

連立方程式

$$\begin{cases} x_1 + x_3 = 0 \\ \qquad x_2 = 1 \end{cases} \tag{6.11}$$

eq:6-10

を考える. 行列とベクトルで表現すると

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Longleftrightarrow \mathbf{R}^\top \boldsymbol{x} = \boldsymbol{\eta},$$

$$\mathbf{R} = (\boldsymbol{r}_1, \boldsymbol{r}_2), \ \boldsymbol{r}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \ \boldsymbol{r}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \ \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \ \boldsymbol{\eta} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

は $\mathbf{R}\boldsymbol{x} = \boldsymbol{\eta}$ とし
ばいい.

である.

すると

$$\boldsymbol{x}_* = a \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} =: a \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + \boldsymbol{x}_0 \quad (a \in \mathbb{R})$$

は方程式 eq:6-10 6.11 の解となる.

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \in \operatorname{span}(\mathbf{R}) = \{a\boldsymbol{r}_1 + b\boldsymbol{r}_2;\ a,\, b \in \mathbb{R}\},$$

$$\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \in \operatorname{span}(\mathbf{R})^\perp := \{\boldsymbol{x} \in \mathbb{R}^3;\ \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0 \ (\forall \boldsymbol{y} \in \operatorname{span}(\mathbf{R}))\}$$

であることに注意せよ. ベクトル $\boldsymbol{x}_0 = (0, 1, 0)^\top$ は

$$\|\boldsymbol{x}\|^2 := \langle \boldsymbol{x}, \boldsymbol{x} \rangle = x_1^2 + x_2^2 + x_3^2$$

を最小にする解である. すなわち, $(0, 1, 0)^\top$ はノルムが最小となる解 (唯一存在) で, 任意の解 $\boldsymbol{x}_*$ に対して

$$\boldsymbol{x}_0^\top \boldsymbol{x}_0 \le \boldsymbol{x}_*^\top \boldsymbol{x}_*$$

をみたしている.

以上の考察より一般的に考える．$d, n \in \mathbb{N}$, $n \geq d$ とし, $\mathbf{R}$ を $n \times d$ 実行列, $\boldsymbol{\eta}$ を $d \times 1$ の縦ベクトルで, 方程式

$$\mathbf{R}^\top \boldsymbol{x} = \boldsymbol{\eta} \tag{6.12}$$

eq:6-11

を解くことを考える．

$$\boldsymbol{\eta} \in \operatorname{span}\left(\mathbf{R}^\top\right) \in \mathbb{R}^d$$

ならば, 解 $\boldsymbol{x}_*$ は存在する．ただし, $\operatorname{span}\left(\mathbf{R}^\top\right)$ は $\mathbf{R}^\top$ の列ベクトルで張られた $\mathbb{R}^d$ の部分空間である．1 つの解 $\boldsymbol{x}_*$ が与えられたとき, 他の解を $\boldsymbol{x}$ と書けば

$$\mathbf{R}^\top \boldsymbol{x}_* = \mathbf{R}^\top \boldsymbol{x} \iff \mathbf{R}^\top\left(\boldsymbol{x}_* - \boldsymbol{x}\right) = \mathbf{0} \iff \boldsymbol{x}_* - \boldsymbol{x} \in \operatorname{span}\left(\mathbf{R}\right)^\perp$$

となる．$\boldsymbol{x}_1 := \boldsymbol{x}_* - \boldsymbol{x} \in \operatorname{span}\left(\mathbf{R}\right)^\perp$ とおけば,

$$\mathbb{R}^n = \operatorname{span}\left(\mathbf{R}\right) \otimes \operatorname{span}\left(\mathbf{R}\right)^\perp$$

なので, 解 $\boldsymbol{x}_*$ は

$$\boldsymbol{x}_* = \boldsymbol{x}_0 + \boldsymbol{x}_1 \quad \left(\boldsymbol{x}_0 \in \operatorname{span}\left(\mathbf{R}\right), \boldsymbol{x}_1 \in \operatorname{span}\left(\mathbf{R}\right)^\perp\right)$$

と一意的に書ける．さらに,

$$\boldsymbol{x}_0^\top \boldsymbol{x}_0 \leq \boldsymbol{x}_0^\top \boldsymbol{x}_0 + \boldsymbol{x}_1^\top \boldsymbol{x}_1$$

なので, $\boldsymbol{x}_0$ は最ノルム小解である．よって, 最小ノルム解は

$$\boldsymbol{x}_0 = \mathbf{R}\boldsymbol{\xi} \quad \left(\boldsymbol{\xi} \in \mathbb{R}^d\right)$$

と書ける．$\boldsymbol{x}_0 = \mathbf{R}\boldsymbol{\xi}$ を方程式 (6.12) に代入すると

$$\mathbf{R}^\top \mathbf{R} \boldsymbol{\xi} = \boldsymbol{\eta}$$

となる．$\boldsymbol{\xi}$ とは異なるベクトル $\widetilde{\boldsymbol{\xi}} \in \mathbb{R}^d$ が

$$\mathbf{R}^\top \mathbf{R} \widetilde{\boldsymbol{\xi}} = \boldsymbol{\eta}$$

をみたすとすると

$$\mathbf{R}\boldsymbol{\xi} = \mathbf{R}\widetilde{\boldsymbol{\xi}}$$

となる．なぜならば, $\mathbf{R}^\top \boldsymbol{x} = \boldsymbol{\eta}$ をみたし, $\boldsymbol{x} \in \operatorname{span}\left(\mathbf{R}\right)$ となるものは唯一だからである．

例にもどれば

$$\mathbf{R}^\top \mathbf{R} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

より

$$\mathbf{R}^\top \mathbf{R}\boldsymbol{\xi} = \boldsymbol{\eta} \iff \begin{cases} 2\xi_1 = 0 \\ \xi_2 = 1 \end{cases}$$

より, $(\xi_1, \xi_2) = (0, 1)$ となる. よって,

$$\boldsymbol{x}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

となる.

以上をまとめると,

$$\boldsymbol{r}_k \in \mathbb{R}^d \, (k = 1, 2, \ldots, d), \quad \boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)^\top \in \mathbb{R}^n$$

とする. 方程式

$$\langle \boldsymbol{r}_k, \boldsymbol{x} \rangle = \eta_k \qquad (k = 1, 2, \ldots, n)$$

をみたす, $\boldsymbol{x} \in \mathbb{R}^d$ でノルム

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$$

を最小とするものをみつける問題の解は

$$\boldsymbol{x}_s = \sum_{k=1}^{n} \xi_k \boldsymbol{r}_k$$

で $(\xi_1, \xi_2, \ldots, \xi_n)$ は

$$\sum_{k=1}^{n} \langle \boldsymbol{r}_i, \boldsymbol{r}_k \rangle \xi_i = \eta_i \qquad (k = 1, 2, \ldots, n)$$

をみたすものである.

この点は関数解析の観点からまとめなおすこと. 参考文献は, 堤正義『逆問題』(2012) の 6 章. 小川英光『工学系の関数解析』(2010) の 7 章. Saitoh, S. and Sawano, Y (2016) の 3 章.

6.5.2 補間スプライン

点 $(t_j,\, y_j)$, $j = 0,\ \ldots,\, n$, $y_0 = 0$, $0 = t_0 < t_1 < \cdots < t_n = 1$ が与えられたとする. 関数族

$$W_0^1 := \left\{ f \in L^2[0,\, 1];\, ,\, f(0) = 0,\ \int_0^1 \{\dot{f}(t)\}^2\, dt < \infty \right\}, \quad \dot{f}(t) = \frac{df}{dt}$$

の中で,

$$f(t_j) = y_j \qquad (j = 1,\, 2,\, \ldots,\, n)$$

をみたしたものの中で

$$\int_0^1 \{\dot{f}(t)\}^2\, dt$$

を最小とするものをみつけたい.

W_0^1 の内積として, $f,\, g \in W_0^1$ に対し

$$\langle f,\, g\rangle := \int_0^1 \dot{f}(t)\dot{g}(t)\, dt, \quad \|f\| = \sqrt{\langle f,\, f\rangle}$$

とおく. $t \in [0,\, 1]$ に対し

$$k_j(t) = \min\{t,\, t_j\} \qquad (j = 1,\, 2,\, \ldots,\, n)$$

とおく. このとき, $k_j(t) \in W_0^2$ であり,

$$\langle f,\, k_j\rangle = \int_0^1 \dot{f}(s)\dot{k}_j(s)\, ds = \int_0^{t_j} \dot{f}(s)\, ds = f(t_j) - f(0) = f(t_j).$$

したがって, $y_j = f(t_j)$ をみたす関数は

$$f(t_j) = \langle k_j,\, f\rangle = y_j \qquad (j = 1,\, 2, \ldots,\, n) \tag{6.13}$$ eq:6-12

条件 (eq:6-12 6.13) をみたす最小ノルムの関数は, リプリゼンタ定理から

$$\widehat{f}(s) = \sum_{j=1}^n \xi_j k_j(s), \qquad s \in [0,\, 1],\, \xi 1,\, \xi_2,\, \ldots,\, \xi_n \in \mathbb{R} \tag{6.14}$$ eq:6-13

の形でかける. (eq:6-13 6.14) を (eq:6-12 6.13) に代入すれば

$$\langle k_j,\, ,\sum_{\ell=1}^n \xi_\ell k_\ell\rangle = \sum_{\ell=1}^n \langle k_j,\, k_\ell\rangle \xi_\ell = y_j \qquad (j = 1,\, 2,\, \ldots,\, n)$$

となる.

$$\boldsymbol{K} := \begin{pmatrix} \langle k_1,\, k_1\rangle & \langle k_1,\, k_2\rangle & \cdots & \langle k_1,\, k_n\rangle \\ \langle k_2,\, k_1\rangle & \langle k_2,\, k_2\rangle & \cdots & \langle k_2,\, k_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle k_n,\, k_1\rangle & \langle k_n,\, k_2\rangle & \cdots & \langle k_n,\, k_n\rangle \end{pmatrix},\, \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},\, \boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}$$

とおけば,

$$\boldsymbol{K}\boldsymbol{\xi} = \boldsymbol{y}$$

となる．したがって, $0 < t_1 < t_2 < \cdots < t_n = 1$ なので, $\boldsymbol{K}$ は正定値になる．実際

$$\mathrm{Det}\left(\boldsymbol{K}\right) = t_1 \prod_{j=2}^{n} (t_j - t_2)$$

と書けること[2]からわかる．さらに, $\boldsymbol{K}_i$ を 1 行目から j 行目に対する小行列とすると, すべての $i = 1,\, 2,\, \ldots,\, n$ に対し $\mathrm{Det}\, \boldsymbol{K}_i \succsim \mathbf{0}$ なので, $\boldsymbol{K} \succsim \mathbf{0}$ がわかる．

たとえば, $n = 5$ とし,

$$f(0) = 0,\ f(0.1) = 0.1,\ f(0.25)1,\ f(0.5) = 2,\ f(0.75) = 1.5,\ f(1) = 1.75$$

とすれば,

$$\begin{pmatrix} 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.1 & 0.25 & 0.5 & 0.5 & 0.5 \\ 0.1 & 0.25 & 0.5 & 0.75 & 0.75 \\ 0.1 & 0.25 & 0.5 & 0.75 & 1 \end{pmatrix} \begin{pmatrix} \xi_1 \\ x_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 1 \\ 2 \\ 1.5 \\ 1.75 \end{pmatrix}$$

の解は

$$\begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{pmatrix} = \begin{pmatrix} -5 \\ -2 \\ 6 \\ -3 \\ 1 \end{pmatrix}$$

より

$$\widehat{f}(t) = -5k_1(t) + 2k_2(t) + 6k_3(t) - 3k_4(t) + k_5(t) = \begin{cases} t & (0 \le t \le 0.1) \\ 6t - 0.5 & (0.1 \le t \le 0.25) \\ 4t & (0.25 \le t \le 0.5) \\ -2t + 3 & (0.5 \le \le 0.75) \\ t + 0.75 & (0.75 \le t \le 1) \end{cases}$$

となる．

[2] $g(t_2,\, t_3,\, \ldots,\, t_n) = \mathrm{Det}\left(\boldsymbol{K}\right)$ とおく． $t_j = t_1\, (j = 2,\, 3,\, \ldots,\, n)$ とおけば, 1 行目と j 行目が等しくなるので, 行列式の性質より $g(t_2,\, t_3,\, \ldots,\, t_n) = 0$. よって, g は $(t_j - t_1)\, (j = 2,\, \ldots,\, n)$ という因子をもつことからわかる．