多変量解析入門
3 非線形回帰モデル

A.Maeda

2017 年 5 月 7 日

3.1 現象のモデル化

説明変数が 1 つのときは, データ $(y_i, x_i), \quad (i = 1, 2, \ldots, n)$ に対して

$$y_i = u(x_i; \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

という関係式が成り立つと仮定する。同様に説明変数が複数ある場合は

$$y_i = u(\boldsymbol{x}_i; \boldsymbol{\beta}), \quad i = 1, 2, \ldots, n \tag{2}$$

が成り立つと仮定する。モデルの例としては

$$y_i = \beta_0 + \beta_1 x_i + \ldots + \beta_n x_i^n + \varepsilon_i \qquad \text{(多項式回帰モデル)}$$
$$y_i = \beta_0 \exp(\beta_1 x_i) + \varepsilon_i \qquad \text{(成長曲線モデル)}$$

などがある。

3.2 基底関数に基づくモデル

スプライン

データ $(x_i, y_i)(i = 1, 2, \ldots, n)$ に対して説明変数を含むいくつかの区間に分割して、区分的な多項式モデルを当てはめることで多項式近似における次数の増加を抑えることができる。3 次スプラインとは区間 $[a, t_1], [t_1, t_2], \ldots, [t_m, b]$ に対して区分的に 3 次多項式を対応させ、各区間の境界における 2 次までの微分係数の連続性条件を課すことによって、なめらかに接続したものである。節点 $t_1, t_2, \ldots, t_m$ を持つ 3 次スプラインは次の式で与えられる。

$$u(x; \boldsymbol{\theta}) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{i=1}^{m} \theta_i (x - t_i)_+^3 \tag{3}$$

ここで

$$\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3, \theta_1, \theta_2, \ldots, \theta_m)^\top \tag{4}$$
$$(x - t_i)_+^3 := (\max\{0, x - t_i\})^3 \tag{5}$$

である。例として節点が t_1, t_2, t_3 の 3 つの場合で上の定式化が正しいことを確かめる。$m = 3$ の場合なので

$$u(x;\boldsymbol{\theta}) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{i=1}^{3} \theta_i (x - t_i)_+^3 \tag{6}$$

となる。場合分けをして考えると

$$u(x;\boldsymbol{\theta}) = \begin{cases} \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \theta_1 (x - t_1)_+^3 & t_1 \leq x_1 < t_2 \\ \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \theta_1 (x - t_1)_+^3 + \theta_2 (x - t_2)_+^3 & t_3 \leq x_1 < t_3 \end{cases} \tag{7}$$

であるから、3 次スプラインの条件

$$\lim_{x \uparrow t_2} u^{(i)}(x;\boldsymbol{\theta}) = \lim_{x \downarrow t_2} u^{(i)}(x;\boldsymbol{\theta}), \quad i = 0, 1, 2 \tag{8}$$

が満たされていることがわかる。ただし $u^{(0)}(x;\boldsymbol{\theta}) = u(x;\boldsymbol{\theta})$ である。

自然 3 次スプライン

$$u(x;\boldsymbol{\theta}) = \beta_0 + \beta_1 x + \sum_{i=1}^{m-2} \theta_i \{d_i(x) - d_{m-1}(x)\} \tag{9}$$

ただし,$\boldsymbol{\theta} = (\beta_0, \beta_1, \theta_1, \theta_2, \ldots, \theta_{m-2})^\top$ であり

$$d_i(x) = \frac{(x - t_i)_+^3 - (x - t_m)_+^3}{t_m - t_i} \tag{10}$$

である。

B-スプライン

はじめに与えられたデータに対して節点を次のようにとる

$$t_1 < t_2 < t_3 < t_4 = x_1 < \cdots < t_{m+1} = x_n < t_{m+2} < t_{m+3} < t_{m+4} \tag{11}$$

これによりデータは $m-3$ 個の区間 $\{[t_i, t_{i+1}]\}_{i=4}^{m-4}$ に分割される。この区間に対して de Boor のアルゴリズムにより $B-$ スプライン関数を定めていく。まずはじめに 0 次の $B-$ スプライン関数を以下で定義する

$$b_j(x;0) = \begin{cases} 1 & x \in [t_j, t_{j+1}] \\ 0 & x \notin [t_j, t_{j+1}] \end{cases}, \quad j = 1, 2, \ldots, m \tag{12}$$

次に r 次のスプライン関数を以下の漸化式で定めていく。

$$b_j(x;r) = \frac{x - t_j}{t_{j+r} - t_i} b_j(x;r-1) + \frac{t_{j+r+1} - x}{t_{j+r+1} - t_{j+1}} b_{j+1}(x;r-1) \tag{13}$$

動径基底関数

動径 $\|\boldsymbol{x}-\boldsymbol{\mu}\|$ のみに依存する関数 $\phi_{i}{}_{i=1}^{m}$ で展開する。

$$y_j = w_0 + \sum_{j=1}^{m} w_j \phi(\|\boldsymbol{x}_i - \boldsymbol{\mu}_j\|) + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{14}$$

動径基底関数の例としてガウス基底がある。

$$\phi_j = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{\mu}_j\|^2}{2h_j^2}\right), \quad j = 1, 2, \ldots, m \tag{15}$$

3.3 基底展開法

y と $\boldsymbol{x}$ にある関係式がなりたっていてそれが基底関数 $\{b(\boldsymbol{x})\}_{j=1}^{m}$ と 1 の線形結合とで表せることを仮定する。すなわち

$$y \sim u(\boldsymbol{x}; \boldsymbol{w}) = \sum_{j=0}^{m} w_j b_j(\boldsymbol{x}) \tag{16}$$

と書けることを期待する。実際には各観測データに関して誤差があるので

$$y_i = \sum_{j=0}^{m} w_j b_j(\boldsymbol{x}_i) + \varepsilon_i \tag{17}$$

となるが

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & b_1(\boldsymbol{x}_1) & b_2(\boldsymbol{x}_1) & \cdots & b_m(\boldsymbol{x}_1) \\ 1 & b_1(\boldsymbol{x}_2) & b_2(\boldsymbol{x}_2) & \cdots & b_m(\boldsymbol{x}_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & b_1(\boldsymbol{x}_n) & b_2(\boldsymbol{x}_n) & \cdots & b_m(\boldsymbol{x}_n) \end{pmatrix} \begin{pmatrix} w_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{18}$$

この式より目的変数の観測値ベクトルを $\boldsymbol{y}$、上の行列を B, 係数ベクトルを $\boldsymbol{w}$, 誤差ベクトルを $\boldsymbol{\varepsilon}$ と書くことにすると

$$\boldsymbol{y} = B\boldsymbol{w} + \boldsymbol{\varepsilon} \tag{19}$$

と書くことができる。

最小 2 乗法

$\boldsymbol{\varepsilon}$ の成分が互いに無相関でこの仮定の元で $S(\boldsymbol{w}) = (\boldsymbol{y} - B\boldsymbol{w})^\top(\boldsymbol{y} - B\boldsymbol{w}) = \boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}$ の最小化を考えるとこれは 2 章の線形回帰モデルの式と全く同じ形 ($X = B, \boldsymbol{\beta} = \boldsymbol{w}$ とする) なので $S_\gamma(\boldsymbol{w})$ の停留点を求めると

$$
\begin{aligned}
&\frac{\partial S(\boldsymbol{w})}{\partial \boldsymbol{w}} && (20)\\
=&\frac{\partial}{\partial \boldsymbol{w}}\left(-\boldsymbol{w}^\top B^\top \boldsymbol{y} - \boldsymbol{y}^\top B\boldsymbol{w} + \boldsymbol{w}B^\top B\boldsymbol{w}\right) && (21)\\
=&\frac{\partial}{\partial \boldsymbol{w}}\left(-2(B^\top \boldsymbol{y})^\top \boldsymbol{w} + \boldsymbol{w}B^\top B\boldsymbol{w}\right) && (22)\\
=&-2B^\top \boldsymbol{y} + 2B^\top B\boldsymbol{w} = 0 && (23)
\end{aligned}
$$

なので停留点 $\hat{\boldsymbol{w}}$ は

$$\hat{\boldsymbol{w}} = (B^\top B)^{-1}B^\top \boldsymbol{y} \tag{24}$$

となる。

最尤法

最尤法についても全く同じ形なのでガウスノイズを持つ非線形回帰モデルの対数尤度関数はガウスノイズを持つ非線形回帰モデルの対数尤度関数は

$$
\begin{aligned}
\ell(\boldsymbol{\theta}, \sigma^2) &= \sum_{i=1}^{n} \log f(y_i|\boldsymbol{x}_i; \boldsymbol{w}, \sigma^2) && (25)\\
&= \sum_{i=1}^{n} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_i - \sum_{j=0}^{m} w_j b_j(\boldsymbol{x}_i)\right)^2}{2\sigma^2}\right)\right] && (26)\\
&= \log\left[\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \prod_{i=1}^{n} \exp\left(-\frac{\left(y_i - \sum_{j=0}^{m} w_j b_j(\boldsymbol{x}_i)\right)^2}{2\sigma^2}\right)\right] && (27)\\
&= \log\left[\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\sum_{i=1}^{n} \frac{\left(y_i - \sum_{j=0}^{m} w_j b_j(\boldsymbol{x}_i)\right)^2}{2\sigma^2}\right)\right] && (28)\\
&= \log\left[\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{(\boldsymbol{y} - B\boldsymbol{w})^\top(\boldsymbol{y} - B\boldsymbol{w})}{2\sigma^2}\right)\right] && (29)\\
&= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}(\boldsymbol{y} - B\boldsymbol{w})^\top(\boldsymbol{y} - B\boldsymbol{w}) && (30)
\end{aligned}
$$

となる。ただし $\boldsymbol{\theta} = (\boldsymbol{w}, \sigma^2)$ である。$\ell(\boldsymbol{w}, \sigma^2)$ の停留点を求める。

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -\frac{1}{\sigma^2}(B^\top \boldsymbol{y} - B^\top B\boldsymbol{w}) = 0 && (31)\\
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{y} - B\boldsymbol{w})^\top(\boldsymbol{y} - B\boldsymbol{w}) = 0 && (32)
\end{aligned}
$$

より、停留点 $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{w}}, \hat{\sigma}^2)$ は

$$\hat{\boldsymbol{w}} = (B^\top B)^{-1} B^\top \boldsymbol{y}, \quad \hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - B\hat{\boldsymbol{w}})^\top (\boldsymbol{y} - B\hat{\boldsymbol{w}}) \tag{33}$$

が成立する。

モデルの選択と評価

モデルの選択基準として AIC を最小とする基準がある。AIC は

$$\mathrm{AIC} = -2(\text{モデルの最大対数尤度}) + 2(\text{モデルの自由パラメータ数}) \tag{34}$$

で与えられる。ガウス型非線形回帰モデルに対する AIC は次の式で与えられる。

$$\mathrm{AIC} = -2\big\{\log f(\boldsymbol{y}|X; \hat{\boldsymbol{w}}, \hat{\sigma}^2) - (m+2)\big\} \tag{35}$$

$$= n\log\big(2\pi\hat{\sigma}^2\big) + n + 2(m+2) \tag{36}$$

3.4 正則化法

基底関数の個数の増加に伴い、モデルのパラメータは増大し、複雑で安定性を欠くモデルになってしまう。そこで最小 2 乗誤差に加えて、パラメータの増大とともに増加するペナルティ項を付け加えた関数の最小化を考える。すなわち

$$S_\gamma(\boldsymbol{w}) = (\boldsymbol{y} - B\boldsymbol{w})^\top (\boldsymbol{y} - B\boldsymbol{w}) + \gamma R(\boldsymbol{w}) \tag{37}$$

の最小化問題を代わりに考える。

正則化最小二乗法

$R(\boldsymbol{w})$ としては非負定値行列 $K \in \mathbb{R}^{m\times m}$ を使った 2 次形式 $\boldsymbol{w}^\top K\boldsymbol{w}$ を用いることが多い。すなわち

$$S_\gamma(\boldsymbol{w}) = (\boldsymbol{y} - B\boldsymbol{w})^\top (\boldsymbol{y} - B\boldsymbol{w}) + \gamma \boldsymbol{w}^\top K\boldsymbol{w} \tag{38}$$

の最小化問題を考える。K としては

$$K = I_m \quad K = D^\top D \tag{39}$$

$$D = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{n\times m} \tag{40}$$

などが用いられる。D は 2 次差分行列である。$S_\gamma(\boldsymbol{w})$ の停留点を求める。

$$\frac{\partial S_\gamma(\boldsymbol{w})}{\partial \boldsymbol{w}} \tag{41}$$

$$= \frac{\partial}{\partial \boldsymbol{w}}\big(-\boldsymbol{w}^\top B^\top \boldsymbol{y} - \boldsymbol{y}^\top B\boldsymbol{w} - \boldsymbol{w}B^\top B\boldsymbol{w} + \gamma\boldsymbol{w}^\top K\boldsymbol{w}\big) \tag{42}$$

$$= \frac{\partial}{\partial \boldsymbol{w}}\big(-2(B^\top \boldsymbol{y})^\top - \boldsymbol{w}(B^\top B + \gamma K)\boldsymbol{w}\big) \tag{43}$$

$$= -2B^\top \boldsymbol{y} - 2(B^\top B + \gamma K)\boldsymbol{w} \tag{44}$$

なので停留点 $\hat{\boldsymbol{w}}$ は

$$\hat{\boldsymbol{w}} = (B^\top B + \gamma K)^{-1} B^\top \boldsymbol{y} \tag{45}$$

となる。

正則化最尤法

ガウスノイズを持つ非線形回帰モデルの対数尤度関数は

$$\ell(\boldsymbol{\theta}, \sigma^2) = \sum_{i=1}^{n} \log f(y_i | \boldsymbol{x}_i; \boldsymbol{w}, \sigma^2) \tag{46}$$

$$= -\frac{n}{2} \log\bigl(2\pi\sigma^2\bigr) - \frac{1}{2\sigma^2} (\boldsymbol{y} - B\boldsymbol{w})^\top (\boldsymbol{y} - B\boldsymbol{w}) \tag{47}$$

であり、この式の対数尤度関数の停留点 (33) を代入する最大対数尤度が求められる。

$$-\frac{n}{2} \log\bigl(2\pi\hat{\sigma}^2\bigr) - \frac{n}{2} \tag{48}$$

データへの過度の当てはまりが起きるため $\hat{\sigma}^2$ が、残差を含む形であることから増大する。これを抑えるために対数尤度関数に次の正則化項を加えた形

$$\ell_\lambda(\boldsymbol{\theta}) = \ell(\boldsymbol{w}) - \frac{\lambda}{2} \boldsymbol{w}^\top K \boldsymbol{w} \tag{49}$$

を最大化する方法が考えられる。$\ell_\lambda(\boldsymbol{\theta})$ の停留点を求める。

$$\frac{\partial \ell_\lambda(\boldsymbol{\theta})}{\partial \boldsymbol{w}} = -\frac{1}{\sigma^2} (B^\top \boldsymbol{y} - B^\top B \boldsymbol{w}) - \lambda K \boldsymbol{w} = 0 \tag{50}$$

$$\frac{\partial \ell_\lambda(\boldsymbol{\theta})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (\boldsymbol{y} - B\boldsymbol{w})^\top (\boldsymbol{y} - B\boldsymbol{w}) = 0 \tag{51}$$

ただし $\boldsymbol{\theta} = (\boldsymbol{w}, \sigma^2)$ である。これより停留点 $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{w}}, \hat{\sigma}^2)$ は

$$\hat{\boldsymbol{w}} = (B^\top B + \lambda \hat{\sigma}^2 K)^{-1} B^\top \boldsymbol{y} \tag{52}$$

$$\hat{\sigma}^2 = \frac{1}{n} (\boldsymbol{y} - B\boldsymbol{w})^\top (\boldsymbol{y} - B\boldsymbol{w}) \tag{53}$$

となる。$\boldsymbol{w}$ の推定については正則化最尤法と正則化最小 2 乗法は同等である。実際 $\gamma = \lambda\sigma^2$ と置けば一致していることがわかる。

モデルの選択と評価

正則最尤法において $\ell_\lambda(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \lambda R(\boldsymbol{w})$ は λ の値によって $R(\boldsymbol{w})$ の影響力が変わるので、λ に依存したモデルの評価基準を考える必要がある。

モデルの有効自由度

モデルの評価基準として平滑化パラメータ λ も含めた AIC を最小化するようなパラメータを選ぶという方法が考えられるが、AIC はモデルの自由パラメータの数のみを評価するので平滑化パラメータの大小自体が

影響する正則化法のモデルに対しては複雑さの程度を捉えきれているとは言えない。基底展開に基づくガウスノイズを持つ非線形回帰モデル

$$\boldsymbol{y} = B\boldsymbol{w} + \boldsymbol{\varepsilon} \tag{54}$$

に対して正則化項 $R(\boldsymbol{w}) = \boldsymbol{w}^\top K \boldsymbol{w}$ を加えた正則化最尤法を考えると

$$\hat{\boldsymbol{w}} = (B^\top B + \lambda\hat{\sigma}^2 K)^{-1} B^\top \boldsymbol{y} \tag{55}$$

であった。よって $\boldsymbol{y}$ の推定値 $\hat{\boldsymbol{y}}$ は $\hat{\boldsymbol{y}} = B\hat{\boldsymbol{w}} = B(B^\top B + \lambda\hat{\sigma}^2 K)^{-1} B^\top \boldsymbol{y}$ である。ここで

$$H(\lambda, m) = B(B^\top B + \lambda\hat{\sigma}^2 K)^{-1} B^\top \tag{56}$$

と置く。$H(\lambda, m)$ を平滑化行列という。$\operatorname{tr} H(\lambda = 0, m) = m$ となり、自由パラメータの個数を表すことから

$$\operatorname{tr} H(\lambda, m) = \operatorname{tr}\bigl[B(B^\top B + \lambda\hat{\sigma}^2 K)^{-1} B^\top\bigr] \tag{57}$$

によってモデルの複雑さの程度を評価する方法がある。$\operatorname{tr} H(\lambda, m)$ を有効自由度という。AIC の自由パラメータの個数の項をこれに置き換えると基準

$$\mathrm{AIC_M} = n(\log 2\pi + 1) + n\log\hat{\sigma}^2 + 2\operatorname{tr} H(\lambda, m) \tag{58}$$

が得られる。AIC_M を最小にする λ, m に対応するモデルを最適なモデルとして選択する。

クロスバリデーション (交差検証法)

回帰モデルを $u(\boldsymbol{x}, \boldsymbol{w})$ と置く。平滑パラメータ λ と基底関数の個数 m を事前に設定する。n 個のデータからひとつを取り除いた $n-1$ 個のデータから $\boldsymbol{w}$ を推定したものを $\hat{\boldsymbol{w}}^{(-i)}$ と置く。次に推定した $u(\boldsymbol{x}; \hat{\boldsymbol{w}}^{(-i)})$ と y_i との残差を考え、すべての $i = 1, 2, \ldots, n$ に対して 2 乗和の平均を考える。

$$\mathrm{CV}(\lambda, m) = \frac{1}{n}\sum_{i=1}^{n}\left\{y_i - u(\boldsymbol{x}_i; \hat{\boldsymbol{w}}^{(-i)})\right\}^2 \tag{59}$$

平滑行列 $H(\lambda, m)$ が $\boldsymbol{y}$ に依存しない形でかける場合は近似ができて

$$\mathrm{CV}(\lambda, m) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{y_i - u(\boldsymbol{x}_i; \hat{\boldsymbol{w}})}{1 - h_{ii}(\lambda, m)}\right\}^2 \tag{60}$$

ここで $h_{ii}(\lambda, m)$ を $n^{-1}\operatorname{tr} H(\lambda, m)$ で近似すると

$$\mathrm{GCV}(\lambda, m) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{y_i - u(\boldsymbol{x}_i; \hat{\boldsymbol{w}})}{1 - n^{-1}\operatorname{tr} H(\lambda, m)}\right\} \tag{61}$$

対数をとって、Taylor の定理を適用することで

$$n\log\mathrm{GCV}(\lambda, m) = n\log\left[\frac{1}{n}\sum_{i=1}^{n} y_i - u(\boldsymbol{x}_i; \hat{\boldsymbol{w}})^2\right] - 2n\log\left\{1 - \frac{1}{n}\operatorname{tr} H(\lambda, m)\right\} \tag{62}$$

$$\approx n\log\hat{\sigma}^2 + 2\operatorname{tr} H(\lambda, m) \tag{63}$$

が導かれる。(37) で $K = I_m$ とすると

$$S_\gamma(\boldsymbol{w}) = \sum_{i=1}^{n} \left\{ y_i - \sum_{j=1} w_j b_j(\boldsymbol{x}_i) \right\}^2 + \gamma \sum_{j=1}^{m} |w_j|^q \tag{64}$$

不等式制約を持つラグランジュの未定乗数法を用いると次と同値である。(ラグランジュの未定乗数法については付録を参照)

$$S_\gamma(\boldsymbol{w}) = \sum_{i=1}^{n} \left\{ y_i - \sum_{j=1} w_j b_j(\boldsymbol{x}_i) \right\}^2, \quad \sum_{j=1}^{m} |w_j|^q \leq \eta \tag{65}$$

リッジ回帰

線形回帰モデル

$$y_i = \beta_0 + \beta_1 x_i 1 + \ldots + \beta_p x_i p + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{66}$$

を考える。各説明変数の平均値を $\bar{x}_i = n^{-1} \sum_{j=1}^{n} x_{ij}$ と置く。

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \ldots, n && (67) \\ &= \beta_0 + \beta_1 \bar{x}_{i1} + \beta_2 \bar{x}_{i2} + \ldots + \beta_p \bar{x}_{ip} \\ &\quad + \beta_1 (x_{i1} - \bar{x}_1) + \ldots + \beta_2 (x_{i2} - \bar{x}_2) + \ldots + \beta_p (x_{ip} - \bar{x}_p) + \varepsilon_i && (68) \\ &= \beta^* + \beta_1 z_{i1} + \ldots + \beta_p z_{ip} + \varepsilon_i && (69) \end{aligned}$$

ただし $\beta_0^* = \beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \ldots + \beta_p \bar{x}_p$ である。線形回帰モデルは

$$\boldsymbol{y} = \beta_0^* \mathbf{1} + Z \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \tag{70}$$

である。ただし $Z = z_{ij} \in \mathbb{R}^{n \times p}$ として

$$z_{ij} = x_{ij} - \bar{x}_j \tag{71}$$

である。リッジ推定量は、誤差の 2 乗和に切片を除く回帰係数ベクトル $\boldsymbol{\beta}_1$ の L_2 ノルム正則化項を付け足した

$$S_\gamma(\beta_0^*, \boldsymbol{\beta}_1) = (\boldsymbol{y} - \beta_0^* \mathbf{1} - Z \boldsymbol{\beta}_1)^\top (\boldsymbol{y} - \beta_0^* \mathbf{1} - Z \boldsymbol{\beta}_1) + \gamma \boldsymbol{\beta}_1^\top \boldsymbol{\beta}_1 \tag{72}$$

の最小化によって与えられる。$S_\gamma(\beta_0^*, \boldsymbol{\beta}_1)$ の停留点を求めることにより

$$\hat{beta}_0^* = \bar{y}, \quad \hat{\boldsymbol{\beta}}_1 = (Z^\top Z + \gamma I_p)^{-1} Z^\top \boldsymbol{y} \tag{73}$$

が導かれる。切片は β_0^* の式から $\beta_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 + \ldots + \hat{\beta}_p x_p$ と求められる。このようにデータを中心化することによって切片と係数 $\beta_1, \ldots, \beta_p$ を切り離して推定を行うことができる。すなわち $\boldsymbol{\beta}_1$ の推定については切片の推定とは独立に

$$S_\gamma(\boldsymbol{\beta}_1) = (\boldsymbol{y} - Z \boldsymbol{\beta}_1)^\top (\boldsymbol{y} - Z \boldsymbol{\beta}_1) + \gamma \sum_{j=1}^{p} \beta_j^2 \tag{74}$$

最小化する問題を解けばよいことがわかる。

lasso 回帰

リッジ回帰の場合と同様、データは中心化されていて、切片と係数 $\beta_1, \ldots, \beta_p$ を切り離して考えられるものとする。リッジ回帰との違いは正則化項として L_1 ノルムを用いている点であり、

$$S_\gamma(\beta_1, \ldots, \beta_p) = \sum_{i=1}^{n} \left\{ y_i - \sum_{j=1}^{p} \beta_j (x_{ij} - \bar{x}_j)^2 \right\} + \sum_{j=1} |\beta_j| \tag{75}$$

という最小化問題を解く。ただ L_1 ノルムは微分不可能なので、解析解を求めることはできないが、数値計算法として shooting アルゴリズムや LARS などいくつかの手法が提案されている。