## 1.3 の補足

最初に $\lambda$ の値を十分大きくして全ての $\beta_j$ を 0 に設定した後，$\lambda$ の値を下げながら，座標降下法を実行することを考える．簡単のため，各 $j = 1, \cdots, p$ について $\sum_{i=1}^{N} x_{i,j}^2 = 1$ であって，$\sum_{i=1}^{N} x_{i,j} y_i$ の値が全て異なると仮定する．このとき，全ての $j$ に対して $\beta_j = 0$ であるような $\lambda$ の値は $\lambda = \max_{1 \le j \le p} \left| \frac{1}{N} \sum_{i=1}^{N} x_{i,j} y_i \right|$ で与えられる；

$\beta_j$ を 1 つ選び，そのほかの $\beta_k$ は 0 として固定する．このとき，(1.10) 式から $L$ の劣微分を 0 にするような $\beta_j$ を求めると

$$
\begin{aligned}
&-\frac{1}{N} \sum_{i=1}^{N} x_{i,j} \left( y_i - \sum_{k=1}^{p} x_{i,k} \beta_k \right) + \lambda \begin{cases} 1 & \beta_j > 0 \\ [-1, 1] & \beta_j = 0 \\ -1 & \beta_j < 0 \end{cases} \\
&= -\frac{1}{N} \sum_{i=1}^{N} x_{i,j} y_i + \frac{1}{p} \sum_{i=1}^{N} x_{i,j}^2 \beta_j + \lambda \begin{cases} 1 & \beta_j > 0 \\ [-1, 1] & \beta_j = 0 \\ -1 & \beta_j < 0 \end{cases} \quad (\because \beta_k = 0 \ \ (k \neq j)) \\
&= -\frac{1}{N} \sum_{i=1}^{N} x_{i,j} y_i + \frac{1}{N} \beta_j + \lambda \begin{cases} 1 & \beta_j > 0 \\ [-1, 1] & \beta_j = 0 \ni 0 \\ -1 & \beta_j < 0 \end{cases} \quad \left( \because \sum_{i=1}^{N} x_{i,j}^2 = 1 \right) \\
&\therefore \beta_j = N \mathcal{S}_\lambda \left( \frac{1}{N} \sum_{i=1}^{N} x_{i,j} y_i \right) = 0 \quad \left( \because \lambda \ge \max_{1 \le j \le p} \left| \frac{1}{N} \sum_{i=1}^{N} x_{i,j} y_i \right| \right)
\end{aligned}
$$

となるからである．もし $\lambda$ をその値より小さくすると，ある 1 つの $j$ で $p_j = \frac{1}{N} \sum_{i=1}^{N} x_{i,j} y_i$ としたときに

$$
\beta_j = N \mathcal{S}_\lambda (p) = N \begin{cases} p - \lambda & (p > \lambda) \\ p + \lambda & (p < \lambda) \end{cases}
$$

となるので

$$
\begin{aligned}
\frac{1}{N} \left| \sum_{i=1}^{N} x_{i,j} \left( y_i - \sum_{k=1}^{p} x_{i,k} \beta_k \right) \right| &= \frac{1}{N} \left| \sum_{i=1}^{N} x_{i,j} y_i - \sum_{k=1}^{N} x_{i,j}^2 \beta_j \right| \\
&= \frac{1}{N} |N \lambda| = \lambda
\end{aligned}
$$

が成立する．

## Ridge

1.1 節において，行列 $X^TX$ が正則であるという仮定の下で二乗誤差 $\|y - X\beta\|$ を最小にする $\beta$ が $\hat{\beta} = (X^TX)^{-1}X^Ty$ となることを導いた．
その後 $N < p$ の場合には $X^TX$ が正則でないことを示したが，$N \geq p$ であって $X^TX$ が正則であっても，行列式が小さければ信頼区間が大きくなるなど不都合が生じる．このような問題を避けるため，定数 $\lambda \geq 0$ を用いて二乗誤差に $\beta$ のノルムの $\lambda$ 倍を加えた

$$L := \frac{1}{N}\|y - X\beta\|^2 + \lambda\|\beta\|^2$$

を最小にする方法がよく用いられる．この方法を Ridge と呼ぶ．上式を最小にする $\beta$ を求めるために，$L$ を $\beta$ で微分すると

$$\frac{\partial L}{\partial \beta} = -\frac{2}{N}X^T(y - X\beta) + 2\lambda\beta$$

となる．(ベクトル微分の公式 $\partial/\partial \boldsymbol{x}(A\boldsymbol{x} - \boldsymbol{b})^T(A\boldsymbol{x} - \boldsymbol{b}) = 2A^T(A\boldsymbol{x} - \boldsymbol{b})$ を用いた．) さらに $X^TX + N\lambda I$ が正則であれば，$\frac{\partial L}{\partial \beta} = 0$ となる $\hat{\beta}$ は

$$\begin{aligned} 0 &= -\frac{2}{N}X^T(y - X\beta) + 2\lambda\beta \\ &= \frac{2}{N}X^Ty - \frac{2}{N}(X^TX + N\lambda)\hat{\beta} \\ (X^TX - N\lambda)\hat{\beta} &= X^Ty \\ \hat{\beta} &= (X^TX + N\lambda)^{-1}X^Ty \end{aligned}$$

となることがわかる．ここで，$\lambda > 0$ ならば $X^TX + N\lambda$ が正則になることがわかる．証明は以下の通り

*Proof.* まず，$(X^TX)^T = X^TX$ が成立するので $X^TX$ は対称行列となる．さらに任意の $\boldsymbol{x} \in \mathbb{R}^p$ に対して

$$\begin{aligned} \boldsymbol{x}^T(X^TX)\boldsymbol{x} &= (\boldsymbol{x}^TX^T)(X\boldsymbol{x}) \\ &= (\boldsymbol{x}X)^T(\boldsymbol{x}X) \end{aligned}$$

となり，最右辺は $\boldsymbol{x}X$ 自身の内積を示しているので $\boldsymbol{x}^T(X^TX)\boldsymbol{x} \geq 0$ であること，つまり $X^TX$ が非負定値行列であることがわかる．さらに $X^TX$ は対称行列であるから，ある直交行列 $P$ と対角行列 $\Lambda$ を用いて

$$X^TX = P^{-1}\Lambda P$$

と表すことができる．今 $e_i$ を $\mathbb{R}^p$ の標準基底とすると，

$$\begin{aligned}(P^{-1}e_i)^T(X^TX)(P^{-1}e_i) &= ((P^{-1}e_i)^TP^{-1})\Lambda(P(P^{-1}e_i)) \\ &= (P(P^{-1}e_i)\Lambda(e_i) \quad (\because P^{-1} = P^T) \\ &= e_i^T\Lambda e_i = \mu_i \geq 0\end{aligned}$$

となることがわかる．ただし $\mu_i$ は $\Lambda$ の $(i,i)$ 成分．したがって任意の $\lambda$ に対して $\mu_i \geq 0$ であることがわかり，各 $\mu_i$ は $X^TX$ の固有値であることから $X^TX$ の全ての固有値が非負であることがわかる．さらに，$X^TX - n\lambda$ の固有値を $t$ とすると

$$\begin{aligned}\det|(X^TX + N\lambda) - tI| &= \det|X^TX - (t - N\lambda)I| = 0 \\ &\Rightarrow t - N\lambda = \mu_i \geq 0 \quad \forall i \\ &\Leftrightarrow t = N\lambda + \mu_i > 0 \quad \forall i\end{aligned}$$

が成立するので $X^TX - n\lambda$ の固有値が全て正であることがわかり，これより $X^TX$ の行列式が 0 でないことがわかる．したがって $X^TX$ が正則であることがわかる．

□