3.5 オーバーラップグループ Lasso

説明変数に関して $\{1,2,3\},\{3,4,5\}$ のように，グループに含まれる変数が重複する場合にもグループ Lasso の処理を構成したい．グループ $k=1,\cdots,K$ において，変数に対する係数 $\beta_1,\cdots,\beta_p$ のうち，用いない変数 β_k については 0 とおく．また $\mathbb{R}^p \ni \beta = \sum_{k=1}^{K}\theta_k$ を満たす $\theta_1,\cdots,\theta_k \in \mathbb{R}^p$ を用意する．そして，与えられる $X \in \mathbb{R}^{N\times p}, y \in \mathbb{R}^N$ から

$$L(\theta) = \frac{1}{2}\left\| y - X\sum_{k=1}^{K}\theta_k \right\|_2^2 + \lambda\sum_{k=1}^{K}\|\theta_k\|_2$$

を最小にする θ を求めることを考える．

例として，$p=5, K=2$ で β_3 のみがオーバーラップしている場合を考える．この場合，θ_1,θ_2 は以下のようになる

$$\theta_1 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_{3,1} \\ 0 \\ 0 \end{bmatrix}, \quad \theta_2 = \begin{bmatrix} 0 \\ 0 \\ \beta_{3,2} \\ \beta_4 \\ \beta_5 \end{bmatrix} \quad (\beta_3 = \beta_{3,1} + \beta_{3,2}).$$

ここで，$X \in \mathbb{R}^{N\times 5}$ の最初の 3 列を $X_1 \in \mathbb{R}^{N\times 3}$, 最後の 3 列を $X_2 \in \mathbb{R}^{N\times 3}$ と書き，L について θ_1,θ_2 の最初の 3 成分 γ_1,γ_2 で劣微分をとると

$$\frac{\partial L}{\partial \gamma_1} = \begin{bmatrix} \frac{\partial L}{\partial \beta_1} \\ \frac{\partial L}{\partial \beta_2} \\ \frac{\partial L}{\partial \beta_{3,1}} \end{bmatrix} = -X_1^T(y - X_1\gamma_1) + \lambda\partial\|\gamma_1\|_2$$
$$\frac{\partial L}{\partial \gamma_2} = \begin{bmatrix} \frac{\partial L}{\partial \beta_{3,2}} \\ \frac{\partial L}{\partial \beta_4} \\ \frac{\partial L}{\partial \beta_5} \end{bmatrix} = -X_2^T(y - X_2\gamma_2) + \lambda\partial\|\gamma_2\|_2$$

となることがわかるので，グループ数が 1 の場合での全体の微分が 0 を含むとした (3.8) 式は

$$-X_1^T(y - X\theta_1) + \lambda\partial\|\theta_1\|_2 = 0$$
$$-X_2^T(y - X\theta_2) + \lambda\partial\|\theta_2\|_2 = 0$$

に相当することがわかる．したがって $\theta_j = 0 (j=1,2)$ で最小となるとき $\|X_j^T y\|_2 \leq \lambda$ が成立することから，結局オーバーラップがあった場合でも通常のグループ Lasso と同じ手法

で最適化ができることがわかる．

3.6 目的変数が複数個ある場合のグループ Lasso

観測データ $X \in \mathbb{R}^{N\times p}, \beta \in \mathbb{R}^{p\times K}, y \in \mathbb{R}^{N\times K}$ から

$$L_0(\beta) := \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{K}\left(y_{i,k} - \sum_{j=1}^{p} x_{i,j}\beta_{j,k}\right)^2 \tag{1}$$

として

$$L(\beta) := L_0(\beta) + \lambda\sum_{j=1}^{p}\|\beta_j\|_2$$

を最小にする $\beta \in \mathbb{R}^{p\times K}$ を求めたい．上式で $K=1$ とすれば，これは第 1 章で扱った Lasso に相当する．$y_i = [y_{i,1}, \cdots, y_{i,K}]$ というように目的変数を K 個に拡張した場合，各 j について $\beta_j = [\beta_{j,1}, \cdots, \beta_{j,K}]$ の K 個のアクティブ・非アクティブのタイミングが同じであることを仮定している．

$L(\beta)$ の最小化について考える．(1) 式を $\beta_{j,k}$ で偏微分すると

$$\frac{\partial L_0}{\partial \beta_{j,k}} = \sum_{i=1}^{N}\{-x_{i,j}(r_{i,k}^{(j)} - x_{i,j}\beta_{j,k})\}$$

$$(\text{ただし } r_{i,k}^{(j)} := y_{i,k} - \sum_{h\neq j} x_{i,h}\beta_{h,k})$$

となることから $L(\beta)$ の β_j による劣微分は

$$\beta_j\sum_{i=1}^{N}x_{i,j}^2 - \sum_{i=1}^{N}x_{i,j}r_i^{(j)} + \lambda\partial\|\beta_j\|_2 \tag{2}$$

となることがわかる．もし $\hat{\beta}_j = 0$ が最適解ならば，(2) が 0 となるのは

$$\lambda\partial\|\beta_j\|_2 \ni \sum_{i=1}^{N}x_{i,j}r_i^{(j)}$$

$$\therefore \left\|\sum_{i=1}^{N}x_{i,j}r_i^{(j)}\right\|_2 \leq \lambda$$

のときであり，$\hat{\beta}_j \neq 0$ が最適解であるならば，グループ数が 1 の場合のグループ Lasso と同様の論法により

$$\hat{\beta}_j = \frac{1}{\sum_{i=1}^{N}x_{i,j}^2}\left(1 - \frac{\lambda}{\|\sum_{i=1}^{N}x_{i,j}r_i^{(j)}\|_2}\right)\sum_{i=1}^{N}x_{i,j}r_i^{(j)}$$

となることがわかる．以上より (2) 式を 0 にする $\hat{\beta}_j$ は

$$\hat{\beta}_j = \frac{1}{\sum_{i=1}^{N} x_{i,j}^2} \left(1 - \frac{\lambda}{\| \sum_{i=1}^{N} x_{i,j} r_i^{(j)} \|_2} \right)_+ \sum_{i=1}^{N} x_{i,j} r_i^{(j)}$$

とかけることがわかる．