# Fused Lassoに基づくスパース順序ロジットモデリング

大阪府立大学大学院工学研究科　　加藤 駿典
電気通信大学大学院情報システム学研究科　川野 秀一

## 1. はじめに

順序あり離散データと共変量との関係性をモデル化する方法として，順序ロジットモデル (Agresti, 2010) が知られている．順序ロジットモデルでは，同じ変数にかかる順序間の係数パラメータを同じ値と仮定（平行性仮定）するか否かの選択が求められる．安川・椿 (2003) では，その選択に関して情報量規準 AIC を用いている．しかし，共変量の個数が増加するにつれて，変数選択を実行することも考慮に入れるとモデルの候補数が莫大になってしまい，情報量規準を用いたモデル選択は現実的には不可能になる．

この問題点を解決するため，本報告では，連続比ロジットモデルに Fused Lasso を組み合わせたモデリング手法を提案する．推定アルゴリズムには，交互方向乗数法を採用する．また，数値実験を通して提案手法の有効性を検証する．

## 2. スパース順序ロジットモデリング

いま，順序つき目的変数 $Y$ と $p$ 次元説明変数 $\boldsymbol{x} = (x_1, \cdots, x_p)^T$ に関して，$n$ 組のデータ $\{(Y_i, \boldsymbol{x}_i); i = 1, \cdots, n\}$ が得られたとする．順序つき目的変数 $Y$ は 1 から $c$ までのカテゴリー値をとるものとし，説明変数に関するデータは規準化されているとする．つまり，$\sum_{i=1}^n x_{ij} = 0$, $\sum_{i=1}^n x_{ij}^2 = 1$ $(j = 1, \cdots, p)$ が成り立つと仮定する．また，$\pi_k(\boldsymbol{x}_i) = Pr(Y_i = k)$ $(k = 1, \cdots, c)$ とする．

連続比ロジットモデルは，

$$\log\left(\frac{\pi_{k+1}(\boldsymbol{x}) + \cdots + \pi_c(\boldsymbol{x})}{\pi_k(\boldsymbol{x})}\right) = \alpha_k + \boldsymbol{x}^T\boldsymbol{\beta}_k, \quad k = 1, \cdots, c-1$$

で与えられ，ここで，$\alpha_1, \cdots, \alpha_{c-1}$ は切片パラメータ，$\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_{c-1}$ は各々$p$ 次元パラメータベクトルである．このとき，次の最大化問題を考える．

$$\max_{\alpha_1, \cdots, \alpha_{c-1}, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_{c-1}} \left\{ LL(\alpha_1, \cdots, \alpha_{c-1}, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_{c-1}) - \lambda_1 \sum_{k=1}^{c-1} ||\boldsymbol{\beta}_k||_1 - \lambda_2 \sum_{k \neq l} ||\boldsymbol{\beta}_k - \boldsymbol{\beta}_l||_1 \right\}.$$

ここで，$LL(\alpha_1, \cdots, \alpha_{c-1}, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_{c-1})$ は連続比ロジットモデルに対する対数尤度関数，$\lambda_1, \lambda_2$ は正の値を取る正則化パラメータである．二項目 $\sum_{k=1}^{c-1} ||\boldsymbol{\beta}_k||_1$ は変数選択のための罰則項，三項目 $\sum_{k \neq l} ||\boldsymbol{\beta}_k - \boldsymbol{\beta}_l||_1$ は平行性仮定のための罰則項である．推定アルゴリズムを Fused lasso に対する交互方向乗数法 (Ye and Xie, 2011) に基づき構成し，その詳細ならびに数値実験の結果は当日報告する．

## 参考文献

Agresti, A. (2010). *Analysis of Ordinal Categorical Data (Second Edition)*. Wiley.
安川武彦, 椿広計 (2003). 格付けデータの比例性と線形性の検証：隣接カテゴリーロジットモデルによる分析. 計算機統計学, **16**, 133–156.
Ye, G.-B. and Xie, X. (2011). Split Bregman method for large scale fused Lasso. *Computational Statistics and Data Analysis*, **55**, 1552–1569.