Bi-level selection を用いた関数ロジスティック回帰モデルの推定

九州大学 大学院数理学研究院　松井 秀俊

1. 概要

複数の個体に対して経時的に観測，測定されたデータを関数化処理し，関数化データ集合に基づく解析を行う方法は関数データ解析とよばれている (Ramsay & Silverman, 2005)．本報告では，関数データに対するロジスティック回帰モデルにスパース正則化を適用することで，モデル推定と変数選択に加えて，各変数がどの判別に寄与しているかの選択，すなわち判別境界の選択を同時に行う方法について議論する．本研究では，関数データに対する判別問題について考察した Matsui (2014) の結果を拡張し，bi-level selection の一つである sparse group lasso (Simon *et al.*, 2013; Vincent & Hansen, 2014) 型の制約を適用する．本報告ではさらに，提案した制約を用いた推定アルゴリズムと，推定に伴う正則化パラメータを選択する方法について紹介する．そして，提案手法を実際のデータに適用した結果を報告する．

2. 関数ロジスティック回帰モデル

いま，関数データとして与えられた n 個の p 変量説明変数 $x_{ij}(t)$ $(i = 1, \ldots, n,\ j = 1, \ldots, p)$ と，L (≥ 3) 群のラベルを表す二値変数 y_{il} $(l = 1, \ldots, L-1)$ が得られたとする．このとき，関数ロジスティック回帰モデルは，データが観測された下での l 群への判別確率 $\pi_l(\boldsymbol{x}_i; \boldsymbol{\theta})$ を用いて次で与えられる．

$$\log\left\{\frac{\pi_l(\boldsymbol{x}_i;\boldsymbol{\theta})}{\pi_L(\boldsymbol{x}_i;\boldsymbol{\theta})}\right\} = \beta_{0l} + \sum_{j=1}^{p}\int x_{ij}(t)\beta_{jl}(t)dt. \quad (l = 1, \ldots, L-1) \tag{1}$$

ここで β_{0l} は定数項，$\beta_{jl}(t)$ は係数関数で，$\boldsymbol{\theta}$ はモデルに含まれる未知パラメータベクトルとする．説明変数および係数関数が基底関数展開によって表現できるという仮定を置くことで，モデル (1) は次式のように，ベクトルと行列に基づく回帰モデルの形で表現できる．

$$\log\left\{\frac{\pi_l(\boldsymbol{x}_i;\boldsymbol{\theta})}{\pi_L(\boldsymbol{x}_i;\boldsymbol{\theta})}\right\} = \sum_{j=1}^{p} Z_j \boldsymbol{b}_{jl}. \tag{2}$$

ただし Z_j は説明変数から構成される既知の行列，$\boldsymbol{b}_{jl}$ はパラメータベクトルである．

3. スパース正則化に基づく推定

パラメータ $\boldsymbol{b}_{jl}$ を，次の正則化対数尤度関数の最大化によって推定する．

$$\ell_\lambda(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - n(1-\alpha)\sum_{j=1}^{p}\lambda_j\left\{\sum_{l=1}^{L-1}\|\boldsymbol{b}_{jl}\|_2^2\right\}^{1/2} - n\alpha\sum_{j=1}^{p}\lambda_j\sum_{l=1}^{L-1}\|\boldsymbol{b}_{jl}\|_2. \tag{3}$$

ここで，$\ell(\boldsymbol{\theta})$ はモデル (2) から構成される対数尤度関数，$\|\cdot\|_2$ は L_2 ノルム，$\lambda_j > 0$，$0 \leq \alpha \leq 1$ はそれぞれ制約の強さとバランスを司る調整パラメータである．(3) 式の右辺第二項は各変数に対するグループ正則化，第三項は各変数と各群に対する正則化に対応している．

一般的に，(3) 式のようなスパース正則化に基づく推定では，解析的に推定量を導出することが困難である．本研究では，反復的に推定量を得る方法として，Simon *et al.* (2013) のアルゴリズムを拡張して適用した結果を紹介する．

参考文献

Matsui, H. (2014). Variable and boundary selection for functional data via multiclass logistic regression modeling. *Comput. Statist. Data Anal.* **78**, 176–185.

Ramsay, J., and Silverman, B. (2005), *Functional Data Analysis.* Springer, New York.

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2013). A sparse-group lasso. *J. Comput. Graph. Statist.* **22**, 231–245.

Vincent, M. and Hansen, N. R. (2014). Sparse group lasso and high dimensional multinomial classification. *Comput. Statist. Data Anal.* **71**, 771–786.