3C-01

区分線形近似を用いた逐次ロジットモデルの変数選択

佐藤俊樹 † 高野祐一 ‡ 宮代隆平 ♮
† 筑波大学 ‡ 専修大学 ♮ 東京農工大学

1 はじめに

逐次ロジットモデルは，順序つき多クラス分類手法のひとつであり，ロジスティック回帰モデルを逐次的に適用してサンプルのクラスラベルを予測する．多くの説明変数を利用すれば学習データへのあてはまりはよくなるが，過剰適合により未知データに対する予測能力が損なわれる．そのため，予測モデル作成の際には説明変数を適切に選択することが望ましい．

説明変数の候補となる集合から，適切に部分集合を選択する問題を変数選択問題という．近年は変数選択問題に対して整数最適化手法によるアプローチが注目を集めている [1, 2]．変数選択に使う適合度指標として，赤池情報量規準 (AIC) やベイズ情報量規準 (BIC)，Mallows' C_p 規準などが知られている．

本研究では，逐次ロジットモデルの変数選択問題を扱う．この問題に対して田中・中川 [3] は，ロジスティック損失関数を二次近似し混合整数二次最適化問題に定式化して解く手法を提案した．しかし，この手法は近似の誤差が大きく，よい変数集合を選ぶことは難しいと考えられる．一方で，Sato, Takano, Miyashiro, & Yoshise [1] は，ロジスティック回帰モデルの変数選択問題に対し，ロジスティック損失関数を区分線形近似し混合整数線形最適化問題に定式化して解く手法を提案した．この手法は十分な数の接線を用意することで近似の誤差が小さくなり，良い変数集合を選択することが期待できる．

このような背景から，本研究では情報量規準に基づく逐次ロジットモデルの変数選択問題に対して，ロジスティック損失関数を区分線形近似し混合整数線形最適化問題として定式化する手法を提案する．また，二次近似を用いた変数選択手法との比較実験を行い，提案手法の有効性を検証する．

Piecewise-Linear Approximation for Feature Subset Selection in a Sequential Logit Model
Toshiki SATO†, Yuichi TAKANO‡ and Ryuhei MIYASHIRO♮
‡ University of Tsukuba
‡ Senshu University
♮ Tokyo University of Agriculture and Technology

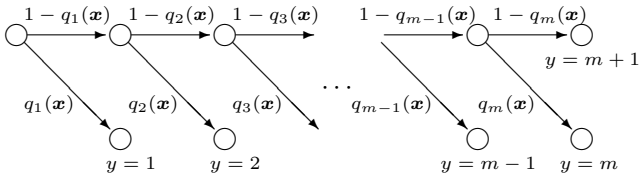

図 1: 逐次ロジットモデルのダイアグラム

2 逐次ロジットモデル

各サンプル $i=1,2,\ldots,n$ の説明変数ベクトル $\boldsymbol{x}_i \in \mathbb{R}^p$ と，順序つきクラスラベル $y_i \in \{1,2,\ldots,m+1\}$ のペア $(\boldsymbol{x}_i, y_i)$ が与えられたとき，逐次ロジットモデルは，クラス $k=1,2,\ldots,m$ に属する確率を次のようにモデル化する（図 1）：

$$q_k(\boldsymbol{x}_i) = \Pr(y_i = k \mid y_i \geq k, \boldsymbol{x}_i) = \frac{1}{1+\exp(-(\boldsymbol{w}_k^\top \boldsymbol{x}_i + b_k))} \quad (i=1,2,\ldots,n).$$

ここで b_k は切片，$\boldsymbol{w}_k = (w_{1k}, w_{2k}, \ldots, w_{pk})^\top$ は偏回帰係数である．逐次ロジットモデルの対数尤度関数は，記号 δ_{ik} と ψ_{ik} を

$$\delta_{ik} = \begin{cases} 1 & \text{if } y_i = k, \\ 0 & \text{otherwise,} \end{cases} \qquad \psi_{ik} = \begin{cases} -1 & \text{if } k < y_i, \\ +1 & \text{if } k = y_i, \\ 0 & \text{otherwise} \end{cases}$$

と定義し，$\boldsymbol{b} = (b_1, b_2, \ldots, b_m)^\top$ と $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m)$ を用いて次のように書ける：

$$L(\boldsymbol{b}, \boldsymbol{W}) = \log \prod_{i=1}^{n} \prod_{k=1}^{m} \Big(1 - q_k(\boldsymbol{x}_i)\Big)^{1-\sum_{j=1}^{k} \delta_{ij}} \Big(q_k(\boldsymbol{x}_i)\Big)^{\delta_{ik}} = -\sum_{i=1}^{n} \sum_{k=1}^{m} |\psi_{ik}| f(\psi_{ik}(\boldsymbol{w}_k^\top \boldsymbol{x}_i + b_k)).$$

ここで $f(v) = \log(1+\exp(-v))$ はロジスティック損失関数である．

3 変数選択問題の定式化

多くの情報量規準は説明変数集合の部分集合 $S \subseteq \{1,2,\ldots,p\}$ を用いて

$$-2\max\{L(\boldsymbol{b},\boldsymbol{W}) \mid w_{jk}=0\ (j\notin S; k=1,2,\ldots,m)\} + Fm(|S|+1)$$

と表すことができる．この F は選択する説明変数の個数に対するペナルティであり，$F=2$ のときは AIC, $F=\log(n)$ のときは BIC に対応する．この情報量規準の値の小さい S が望ましい説明変数集合とされる．この考え方に基づき変数選択問題を定式化する．

3.1 混合整数非線形最適化問題

0-1 決定変数 $\boldsymbol{z}=(z_1,z_2,\ldots,z_p)^\top$ を導入し，$j\in S$ のとき 1, $j\notin S$ のとき 0 とする．すると，情報量規準を最小化する変数選択問題は次のように書ける：

$$\underset{\boldsymbol{b},\boldsymbol{W},\boldsymbol{z}}{\text{minimize}}\ 2\sum_{i=1}^{n}\sum_{k=1}^{m}|\psi_{ik}|f(\psi_{ik}(\boldsymbol{w}_k^\top\boldsymbol{x}_i+b_k)) + Fm\left(\sum_{j=1}^{p}z_j+1\right) \tag{1}$$

$$\text{subject to } z_j=0\Rightarrow w_{jk}=0 \quad (j=1,2,\ldots,p; k=1,2,\ldots,m), \tag{2}$$

$$z_j\in\{0,1\}\ (j=1,2,\ldots,p). \tag{3}$$

制約 (2) はタイプ 1 の特殊順序集合 (SOS1) 制約による表現が可能である．しかしながら，この問題の目的関数は非線形となっており，このままでは標準的なソルバーを用いて求解することは難しい．したがって，ソルバーを利用するには，扱いやすい目的関数に変形する必要がある．

3.2 二次近似

先行研究 [3] は，ロジスティック損失関数を次のように二次近似する手法を提案した：

$$f(v)\approx\log 2-\frac{v}{2}+\frac{v^2}{8}.$$

これにより，問題 (1)–(3) は次の問題 (MIQO) に変形される：

$$\underset{\boldsymbol{b},\boldsymbol{W},\boldsymbol{z}}{\text{minimize}}\ 2\sum_{i=1}^{n}\sum_{k=1}^{m}|\psi_{ik}|\left(\log 2-\frac{(\psi_{ik}(\boldsymbol{w}_k^\top\boldsymbol{x}_i+b_k))}{2} + \frac{(\psi_{ik}(\boldsymbol{w}_k^\top\boldsymbol{x}_i+b_k))^2}{8}\right)+Fm\left(\sum_{j=1}^{p}z_j+1\right)$$

subject to 制約式 (2), (3).

この問題は混合整数二次最適化問題であり，標準的な整数最適化ソルバーで扱うことができる．

3.3 区分線形近似

先行研究 [1] は，ロジスティック損失関数を区分線形近似する方法を提案した．この近似手法に基づき問題 (1)–(3) を変形する．

点集合 $V=\{v_1,v_2,\ldots,v_h\}$ を与え，点 v_ℓ によるロジスティック損失関数の接線を

$$h(v;v_\ell)=f'(v_\ell)(v-v_\ell)+f(v_\ell)\ \ (\ell=1,2,\ldots,h)$$

のように表すと，ロジスティック損失関数は次のように近似できる：

$$\begin{aligned}f(v)&\approx\max\{h(v;v_\ell)\mid\ell=1,2,\ldots,h\}\\&=\min\{t\mid t\geq h(v;v_\ell)\ \ (\ell=1,2,\ldots,h)\}.\end{aligned}$$

よって，決定変数 $\boldsymbol{T}=(t_{ik}; i=1,2,\ldots,n; k=1,2,\ldots,m)$ を導入し，ロジスティック損失関数を区分線形近似すると，問題 (1)–(3) は以下の問題 (MILO) に変形される：

$$\underset{\boldsymbol{b},\boldsymbol{T},\boldsymbol{W},\boldsymbol{z}}{\text{minimize}}\ 2\sum_{i=1}^{n}\sum_{k=1}^{m}|\psi_{ik}|t_{ik}+Fm\left(\sum_{j=1}^{p}z_j+1\right)$$

$$\text{subject to } t_{ik}\geq f'(v_\ell)(\psi_{ik}(\boldsymbol{w}_k^\top\boldsymbol{x}_i+b_k)-v_\ell)+f(v_\ell) \quad (i=1,2,\ldots,n; k=1,2,\ldots,m; \ell=1,2,\ldots,h),$$

制約式 (2), (3).

この問題は混合整数線形最適化問題であり，標準的な整数最適化ソルバーで扱うことができる．

4 数値実験

二次近似による問題 (MIQO) と，区分線形近似による問題 (MILO) を，近似精度と計算時間の観点から比較する．実験には UCI Machine Learning Repository [`http://archive.ics.uci.edu/ml`] で公開されているデータセットを使用した．求解には最適化ソルバー Gurobi Optimizer 6.0.0 を用いた．区分線形近似のほうが，二次近似よりもよい変数集合を選択することを実験によって確認した．

参考文献

[1] T. Sato, Y. Takano, R. Miyashiro, & A. Yoshise, Feature subset selection for logistic regression via mixed integer optimization. Discussion Paper Series, No. 1324, Department of Policy and Planning Sciences, University of Tsukuba (2015).

[2] T. Sato, Y. Takano, & T. Nakahara, Using mixed integer optimisation to select variables for a store choice model. to apper in *International Journal of Knowledge Engineering and Soft Data Paradigms.*

[3] 田中克弘，中川秀敏，企業格付判別のための SVM 手法の提案および逐次ロジットモデルとの比較による有効性検証. 日本オペレーションズ・リサーチ学会和文論文誌, 57 (2014), 92–111.