

グループ正則化に基づく順序ロジットモデルにおける隣接クラスの統合

永沼 瑞穂¹・吉川 剛平²・川野 秀一²

(受付 2020 年 3 月 23 日；改訂 6 月 29 日；採択 6 月 29 日)

要 旨

本稿では、多クラス分類を目的とした順序ロジットモデルの枠組みにおいて、クラスを統合する方法を提案する。クラスを統合することによって、モデルの解釈が容易になるとともに、冗長なクラスが存在する場合はその冗長性を排除することができる。クラスの統合は隣接する各クラスの事後確率が等しくなるときに実行し、この目的のために構造的な正則化の一つであるグループ正則化を用いる。グループ正則化は微分不可能な点を含むため、モデルに含まれるパラメータの推定値は、交互方向乗数法に基づく推定アルゴリズムにより得る。正則化パラメータの値は、グループ正則化により推定されたモデルの自由度をもつベイズ型情報量規準により選択する。モンテカルロ・シミュレーションおよび実データへの適用を通して、提案手法の有効性を検証する。

キーワード：グループ lasso, 交互方向乗数法, 順序付きカテゴリカルデータ, 隣接カテゴリロジットモデル。

1. はじめに

多クラス分類は、自然科学・社会科学のさまざまな分野で用いられており、それを実行するための統計モデルは、クラスを表すカテゴリカルデータを目的変数、それに関係する共変量を説明変数としてモデル化される。多クラス分類においては、クラスの個数が多くなるときには注意が必要である。得られた解析結果の解釈が困難になる (Price et al., 2019), または冗長なクラスが生じる, といった現象が起きやすくなる。その例として、マーケティング分野における顧客市場のセグメンテーションを挙げることができる。顧客市場のセグメンテーションでは、大規模データ分析を行うことで顧客市場が過剰に細分化されてしまい、クラス毎の解釈が困難になるという現象が生じる。これはユーザーの分析コストを増大させ、マーケティング資源を浪費してしまうという恐れがある。このようなクラスの個数が多い状況においては、クラスを統合することによって、分析コストを削減することができ有用である。近年、そのクラスの統合を目的として、Price et al. (2019) は、多項ロジットモデルと構造的な正則化に基づいた統計手法 group fused multinomial logistic regression (GFMR) を提案している。

一方、カテゴリカルデータの中でも、順序関係を持つデータは順序付きカテゴリカルデータ

¹ 電気通信大学大学院 情報理工学研究科：〒182-8585 東京都調布市調布ヶ丘 1-5-1 (現 株式会社マクロミル：〒108-0075 東京都港区港南 2-16-1 品川イーストワンタワー 11F)

² 電気通信大学大学院 情報理工学研究科：〒182-8585 東京都調布市調布ヶ丘 1-5-1

と呼ばれている．例えば，企業の格付けや 5 段階で表される学生の成績，アンケートの選好度などがある．このような順序付きカテゴリカルデータと共変量との間をモデリングする統計手法として，順序ロジットモデル (Agresti, 2010) が広く用いられている．前述したクラス過多な状況から生じる問題は，順序付きカテゴリカルデータを目的変数としたときも同様な理由で生じる．したがって，この場合でもクラスを統合することは有用である．Price et al. (2019) では，順序付きカテゴリカルデータを目的変数に用いる場合でも GFMR によってクラス統合を行っている．しかし，GFMR は多項ロジットモデルに基づいており，順序付きカテゴリカルデータを目的変数とするならば，順序ロジットモデルを用いる方が望ましい．

本稿では，順序ロジットモデルの枠組みの下，クラスを統合するための統計手法を提案する．順序ロジットモデルの中でも，隣接カテゴリロジットモデル (Goodman, 1984) を用い，クラスの統合は，隣接する各クラスの事後確率が等しくなるときに実行する．この目的のために，構造的正則化の一つであるグループ正則化，特にグループ lasso (Yuan and Lin, 2006) を用いる．グループ lasso の正則化項は微分不可能な点を含むため，モデルに含まれるパラメータの推定値は，交互方向乗数法 (Boyd et al., 2011) に基づく推定アルゴリズムにより得る．構築したモデルには正則化パラメータが含まれるが，この値をグループ正則化により推定されたモデルの自由度をもつベイズ型情報量規準 (Schwarz, 1978) により選択する．

本稿の構成は次の通りである．2 節では順序ロジットモデル，特に隣接カテゴリロジットモデルを紹介する．3 節では提案手法とその推定アルゴリズム，正則化パラメータの選択方法，関連研究の GFMR について述べる．4 節ではモンテカルロ・シミュレーションおよび実データへの適用を通して，提案手法の有効性を検証する．5 節では本稿のまとめを行い，今後の課題について検討する．

2. 順序ロジットモデル

順序ロジットモデルは，各クラスの事後確率の対数オッズを，説明変数の線形結合で表すことによってモデル化される (Ananth and Kleinbaum, 1997; Agresti, 2010)．この対数オッズの設定方法によって，累積ロジットモデル (McCullagh, 1980; Armstrong and Sloan, 1989; Whitehead et al., 2001)，連続比ロジットモデル (Fienberg and Mason, 1979; Fienberg, 2007)，隣接カテゴリロジットモデル (Adjacent-Categories Logit model, ACL) (Anderson, 1984; Goodman, 1984; Agresti, 1992) などのさまざまなモデルが提案されている．本稿ではこの中でも，隣接カテゴリロジットモデルを用いる．

順序付きカテゴリカル変数 $G \in \{1, \dots, J\}$ と p 次元説明変数 $\mathbf{x}^\dagger = (x_1, \dots, x_p)^\top$ より， n 組のデータ $\{(g_i, \mathbf{x}_i^\dagger); i = 1, \dots, n\}$ が得られたとする．ここで，順序付きカテゴリカル変数とは，何らかの尺度についてのクラスの順序を示したものである．例えば，アンケート選好度を変数とした場合，その取り得る選択肢として，1: あてはまる，2: ややあてはまる，3: ややあてはまらない，4: あてはまらない，の 4 つのクラスは，順序付きカテゴリカル変数である．また，説明変数に関するデータは標準化されているとする．つまり， $\sum_{i=1}^n x_{ij}^\dagger = 0$ ， $\sum_{i=1}^n (x_{ij}^\dagger)^2 = n$ ($j = 1, \dots, p$) が成り立つものとする．

説明変数に関するデータ $\mathbf{x}_i^\dagger$ が与えられた下で， j 番目のクラスに属する事後確率を $\Pr(g_i = j | \mathbf{x}_i^\dagger) = \pi_j(\mathbf{x}_i^\dagger)$ ($j = 1, \dots, J$) とする．このとき，隣接カテゴリロジットモデルは

$$(2.1) \quad \log \frac{\pi_j(\mathbf{x}_i^\dagger)}{\pi_{j+1}(\mathbf{x}_i^\dagger)} = \beta_{0j} + \beta_{1j}x_{i1} + \dots + \beta_{pj}x_{ip} = \boldsymbol{\beta}_j^\top \mathbf{x}_i, \quad j = 1, \dots, J-1$$

となる．ここで， $\mathbf{x}_i = (1, \mathbf{x}_i^\dagger)^\top$ ， $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{pj})^\top$ は j 番目のクラスに対する $(p+1)$

次元の回帰係数ベクトルである．つまり， j 番目のクラスに属する確率と $(j+1)$ 番目のクラスに属する事後確率の対数オッズと，説明変数間をモデル化したものが隣接カテゴリロジットモデルである．

いま， i 番目のデータ $\mathbf{x}_i^\dagger$ が j 番目のクラスに属しているとする．このとき，指示変数 y_{ij} ($i = 1, \dots, n; j = 1, \dots, J$) を

$$(2.2) \quad y_{ij} = \begin{cases} 1 & \text{データ } i \text{ がクラス } j \text{ に属しているとき} \\ 0 & \text{その他} \end{cases}$$

と定義すると，データ $\mathbf{x}_i^\dagger$ に対する順序付きカテゴリカル変数 G_i は， J 次元のベクトル $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})^\top$ と表すことができる．ここで， $y_{iJ} = 1 - \sum_{j=1}^{J-1} y_{ij}$ が成り立つことに注意しておく．

ベクトル $\mathbf{y}_i$ は (2.2) 式の構成法から，確率 $\pi_1(\mathbf{x}_i^\dagger), \dots, \pi_J(\mathbf{x}_i^\dagger)$ を持つ多項分布にしたがうことがわかる．したがって，隣接カテゴリロジットモデルの対数尤度関数は，

$$(2.3) \quad \ell(\beta_1, \dots, \beta_{J-1}) = \sum_{i=1}^n \left\{ \sum_{j=1}^{J-1} y_{ij} \sum_{k=j}^{J-1} \beta_k^\top \mathbf{x}_i - \log \left[1 + \sum_{j=1}^{J-1} \exp \left(\sum_{k=j}^{J-1} \beta_k^\top \mathbf{x}_i \right) \right] \right\}$$

と表すことができる．(2.3) 式を最大にするパラメータは，解析的に陽に表現することが難しいので，非線形最適化手法，例えば，ニュートン・ラフソン法によって求める．なお，(2.3) 式の導出過程については Agresti (2010) を参照されたい．

3. 提案手法

本節では，隣接カテゴリロジットモデルの枠組みの下，クラスを統合するために，モデルに含まれるパラメータを構造的な正則化の一つであるグループ lasso (Yuan and Lin, 2006) に基づいて推定する方法を提案する．その上で，推定アルゴリズムならびに正則化パラメータの選択方法について述べる．また，関連研究として Price et al. (2019) の GFMR を紹介する．

3.1 クラス統合とグループ正則化対数尤度関数

Price et al. (2019) は，各クラスの事後確率が等しくなるときにクラスを統合することを提案した．本稿でもこの考えにしたがうものとする．しかし，本稿で扱っている隣接カテゴリロジットモデルには，クラスに順序が存在しているため，隣接していないクラスを統合することはせず，隣接しているクラスのみを統合する．すなわち，本稿では，隣接した各クラスに対して，それらの事後確率が等しくなる場合にクラスを統合することを考える．以降，計画行列 $(\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ の階数は $(p+1)$ ，すなわちフルランクであると仮定する．

隣接カテゴリロジットモデルにおいて，隣接したクラスの統合と回帰係数の関係について調べる．いま， j 番目のクラスと $(j+1)$ 番目のクラスを統合することを考える．すなわち， $\pi_j(\mathbf{x}_i^\dagger) = \pi_{j+1}(\mathbf{x}_i^\dagger)$ ($i = 1, \dots, n$) となる状況を考える．このとき，

$$\log \frac{\pi_j(\mathbf{x}_i^\dagger)}{\pi_{j+1}(\mathbf{x}_i^\dagger)} = 0, \quad i = 1, \dots, n,$$

となる．この式と (2.1) 式より， $\beta_j^\top \mathbf{x}_i = 0$ ($i = 1, \dots, n$) となり，計画行列がフルランクであるという仮定を用いると $\beta_j = \mathbf{0}$ となることがわかる．したがって，隣接カテゴリロジットモデルにおいては， $\pi_j(\mathbf{x}_i^\dagger) = \pi_{j+1}(\mathbf{x}_i^\dagger)$ ($i = 1, \dots, n$) ならば $\beta_j = \mathbf{0}$ であることがわかる．一方，上述の手順を逆に考えることにより， $\beta_j = \mathbf{0}$ ならば $\pi_j(\mathbf{x}_i^\dagger) = \pi_{j+1}(\mathbf{x}_i^\dagger)$ ($i = 1, \dots, n$) というこ

とがわかり、以上より、 $\pi_j(\mathbf{x}_i^\dagger) = \pi_{j+1}(\mathbf{x}_i^\dagger)$ ($i = 1, \dots, n$) となるための必要十分条件は $\beta_j = \mathbf{0}$ である。なお、必要条件のみならば、計画行列のフルランク性の仮定は必要ないことに注意されたい。

したがってクラス統合を行うために、 $\beta_j = \mathbf{0}$ と推定することが可能なグループ正則化項が付与された対数尤度関数

$$(3.1) \quad \max_{\beta_1, \dots, \beta_{J-1}} \left\{ \ell(\beta_1, \dots, \beta_{J-1}) - \lambda \sum_{j=1}^{J-1} \|\beta_j\|_2 \right\}$$

を最大化し、パラメータを推定する方法を提案する。ここで、 $\ell(\beta_1, \dots, \beta_{J-1})$ は (2.3) 式の隣接カテゴリーロジットモデルに対する対数尤度関数、 λ は非負の値をとる正則化パラメータ、 $\|\cdot\|_2$ はベクトルの L_2 ノルムである。(3.1) 式の第 2 項目は、隣接したクラスを統合するためのグループ正則化項であり、本稿ではグループ lasso (Yuan and Lin, 2006) を用いる。このグループ正則化項により、 β_j ($j = 1, \dots, J-1$) が 0 ベクトルか否かが判定され、もし、 $\beta_j = \mathbf{0}$ ならば j 番目のクラスと $(j+1)$ 番目のクラスを統合する。なお、 p_j を β_j の要素数と定義した場合、グループ lasso は通常 $\sum_{j=1}^{J-1} \sqrt{p_j} \|\beta_j\|_2$ と定式化されるが、(3.1) 式においては $p_1 = \dots = p_{J-1}$ のため、 $\sqrt{p_j}$ は含めずに定式化していることに注意されたい。

3.2 推定アルゴリズム

(3.1) 式に含まれるグループ正則化項は、微分不可能な点をもつため、ニュートン・ラフソン法等の非線形最適化手法を用いて推定値を求めることは困難である。本稿では、交互方向乗数法 (Boyd et al., 2011) に基づいた推定アルゴリズムにより (3.1) 式に対するパラメータ推定値を得る。

まず、(3.1) 式を最小化問題に書き換えると

$$\min_{\beta} \left\{ -\ell(\beta) + \lambda \sum_{j=1}^{J-1} \|\beta_j\|_2 \right\}$$

となる。ここで、 $\beta = (\beta_1^\top, \dots, \beta_{J-1}^\top)^\top$ である。次に、微分不可能な正則化項を分離するために $(p+1)$ 次元の変数 $\mathbf{z}_j$ ($j = 1, \dots, J-1$) を用いて、以下の等式制約付きの最小化問題に変形する。

$$\min_{\beta, \mathbf{z}} \left\{ -\ell(\beta) + \lambda \sum_{j=1}^{J-1} \|\mathbf{z}_j\|_2 \right\} \quad \text{subject to} \quad \beta = \mathbf{z}.$$

ここで、 $\mathbf{z} = (\mathbf{z}_1^\top, \dots, \mathbf{z}_{J-1}^\top)^\top$ である。したがって、この最小化問題に対する拡張ラグランジアンは

$$L_\rho(\beta, \mathbf{z}, \mathbf{u}) = -\ell(\beta) + \lambda \sum_{j=1}^{J-1} \|\mathbf{z}_j\|_2 + \sum_{j=1}^{J-1} \mathbf{u}_j^\top (\beta_j - \mathbf{z}_j) + \frac{\rho}{2} \sum_{j=1}^{J-1} \|\beta_j - \mathbf{z}_j\|_2^2$$

となる。ここで、 $\mathbf{u}_j$ ($j = 1, \dots, J-1$) は $(p+1)$ 次元のラグランジュ乗数で、 $\mathbf{u} = (\mathbf{u}_1^\top, \dots, \mathbf{u}_{J-1}^\top)^\top$ である。また、 ρ は正の値を取る調整パラメータである。

交互方向乗数法では、まず各パラメータ $(\beta, \mathbf{z})$ について、拡張ラグランジアンを最小化するパラメータを推定する。次に、その最小となるパラメータを用いてラグランジュ乗数を勾配法 (最急上昇法) により更新する。つまり、 t ($t = 1, 2, \dots$) 回目の更新値をそれぞれ $\beta^t, \mathbf{z}^t, \mathbf{u}^t$ とすると、

$$(3.2) \quad \beta^{t+1} = \operatorname{argmin}_{\beta} L_{\rho}(\beta, \mathbf{z}^t, \mathbf{u}^t),$$

$$(3.3) \quad \begin{aligned} \mathbf{z}_j^{t+1} &= \operatorname{argmin}_{\mathbf{z}_j} L_{\rho}(\beta^{t+1}, \mathbf{z}_j, \mathbf{u}^t), \quad j = 1, \dots, J-1, \\ \mathbf{u}^{t+1} &= \mathbf{u}^t + \rho(\beta^{t+1} - \mathbf{z}^{t+1}) \end{aligned}$$

のように更新していく。

(3.2)式と(3.3)式の具体的な更新式を述べる。まず β の更新式(3.2)について、右辺の最小化は、

$$\min_{\beta} \left\{ -\ell(\beta) + \sum_{j=1}^{J-1} \mathbf{u}_j^{t\top} (\beta_j - \mathbf{z}_j^t) + \frac{\rho}{2} \sum_{j=1}^{J-1} \|\beta_j - \mathbf{z}_j^t\|_2^2 \right\}$$

と同値である。この最小化問題はパラメータ β に対して微分可能であるため、更新値 β^{t+1} はニュートン・ラフソン法により得ることができる。ニュートン・ラフソン法を実行するための1次導関数および2次導関数は順に、

$$\begin{aligned} \frac{\partial L_{\rho}}{\partial \beta} &= -\frac{\partial \ell}{\partial \beta} + \rho \beta - \rho \mathbf{z} + \mathbf{u}, \\ \frac{\partial^2 L_{\rho}}{\partial \beta \partial \beta^{\top}} &= -\frac{\partial^2 \ell}{\partial \beta \partial \beta^{\top}} + \rho I_{(J-1)(p+1)} \end{aligned}$$

となる。ここで、 $I_{(J-1)(p+1)}$ は $(J-1)(p+1) \times (J-1)(p+1)$ 型の単位行列であり、対数尤度関数の1次導関数と2次導関数の具体的な式は付録Aに載せている。

次に $\mathbf{z}$ の更新式について述べる。(3.3)式について、右辺の最小化は、

$$(3.4) \quad \min_{\mathbf{z}} \left\{ \lambda \sum_{j=1}^{J-1} \|\mathbf{z}_j\|_2 + \sum_{j=1}^{J-1} \mathbf{u}_j^{t\top} (\beta_j^{t+1} - \mathbf{z}_j) + \frac{\rho}{2} \sum_{j=1}^{J-1} \|\beta_j^{t+1} - \mathbf{z}_j\|_2^2 \right\}$$

と同値である。ここで、等式

$$\mathbf{u}_j^{t\top} (\beta_j^{t+1} - \mathbf{z}_j) + \frac{\rho}{2} \|\beta_j^{t+1} - \mathbf{z}_j\|_2^2 = \frac{\rho}{2} \left\| \beta_j^{t+1} - \mathbf{z}_j + \frac{1}{\rho} \mathbf{u}_j^t \right\|_2^2 - \frac{\rho}{2} \left\| \frac{1}{\rho} \mathbf{u}_j^t \right\|_2^2$$

に注意すると、問題(3.4)の $\mathbf{z}$ に関する最小化は、

$$\sum_{j=1}^{J-1} \rho \left[\frac{1}{2} \left\| \mathbf{z}_j - \left\{ \beta_j^{t+1} + \frac{1}{\rho} \mathbf{u}_j^t \right\} \right\|_2^2 + \frac{\lambda}{\rho} \|\mathbf{z}_j\|_2 \right]$$

の $\mathbf{z}$ に関する最小化と同値となる。またこれは、各 j ($j = 1, \dots, J-1$) 毎に問題が分離されているため、更新式は

$$(3.5) \quad \mathbf{z}_j^{t+1} = S_{\frac{\lambda}{\rho}} \left(\beta_j^{t+1} + \frac{1}{\rho} \mathbf{u}_j^t \right), \quad j = 1, \dots, J-1$$

となる。ここで、 $S_v(\mathbf{a})$ はベクトルに対する軟閾値作用素、

$$S_v(\mathbf{a}) = \left(1 - \frac{v}{\|\mathbf{a}\|_2} \right)_+ \mathbf{a}$$

であり、 $(\cdot)_+$ は任意の実数 x に対して、 $(x)_+ = \max(0, x)$ と定義される。なお、更新式(3.5)の導出については川野 他 (2018)を参照されたい。

3.3 正則化パラメータの選択

パラメータの推定値は、調整パラメータ ρ と正則化パラメータ λ を含んでいる。調整パラメータ ρ は Boyd et al. (2011) にしたがって $\rho = 1$ と設定する。一方、正則化パラメータ λ は情報量規準 BIC (Schwarz, 1978) に基づいて選択する。

BIC はその第 2 項目にモデルの自由度を必要とする。Yuan and Lin (2006) は、線形回帰モデルにおいて、グループ lasso により推定したときの推定値 $\hat{\xi}$ を持つモデルの自由度を

$$\sum_{j=1}^J I(\|\hat{\xi}_j\|_2 > 0) + \sum_{j=1}^J \frac{\|\hat{\xi}_j\|_2}{\|\hat{\xi}_j^{\text{LS}}\|_2} (p_j - 1)$$

と定義した。ここで $\hat{\xi}_j^{\text{LS}}$ は最小二乗推定値である。本稿ではこの考えに基づいて、提案手法のモデルの自由度を

$$(3.6) \quad \text{df} = \sum_{j=1}^{J-1} I(\|\hat{\beta}_j\|_2 > 0) + p \sum_{j=1}^{J-1} \frac{\|\hat{\beta}_j\|_2}{\|\hat{\beta}_j^{\text{ridge}}\|_2}$$

とする。ここで、 $\hat{\beta}_j$ は最大化問題 (3.1) の解、 $\hat{\beta}_j^{\text{ridge}}$ はリッジ推定値を表している。リッジ推定値に含まれる正則化パラメータの値は 10^{-5} に固定した。最尤推定値ではなくリッジ推定値を用いた理由は、4 節の数値実験において最尤推定値が求まらなかったことが多々あったためである。また、 $p_1 = \dots = p_{J-1} = p + 1$ となる関係性に注意されたい。

以上より、本稿で用いる情報量規準 BIC は (3.6) 式の自由度を用いて

$$\text{BIC} = -2\ell(\hat{\beta}) + \text{df} \log(n)$$

となる。BIC の値が最小となる正則化パラメータ λ の値を最適な値として採用する。

3.4 関連研究

Price et al. (2019) は、順序関係のないカテゴリカルデータを目的変数とした多項ロジットモデルに対し、クラス統合を実行する方法として GFMR を提案した。GFMR は以下の最小化問題として定式化される。

$$\min_{\beta_1, \dots, \beta_J} \left\{ -\ell_{\text{MLM}}(\beta_1, \dots, \beta_{J-1}) + \lambda \sum_{(j,m) \in \mathcal{L}} \|\beta_j - \beta_m\|_2 \right\} \quad \text{subject to} \quad \beta_J = \mathbf{0}.$$

ここで、 $\ell_{\text{MLM}}(\beta_1, \dots, \beta_{J-1})$ は多項ロジットモデルに対する対数尤度関数、 λ は正の値をとる正則化パラメータである。第 2 項目はクラス統合を行うための正則化項であり、候補集合 $\mathcal{L}$ は $\mathcal{S} = \{(a, b) \in \mathcal{J} \times \mathcal{J} : a < b\}$ からクラスの組み合わせを任意に設定する。ここで $\mathcal{J} = \{1, \dots, J\}$ である。例えば、1 番目のクラスと 3 番目のクラスが統合する可能性がある場合は、 $\mathcal{L} = \{(1, 3)\}$ とする。クラスの関係性に何も事前の情報がなければ、 $\mathcal{L} = \mathcal{S}$ とする。 $\beta_J = \mathbf{0}$ は多項ロジットモデルに関する制約条件である。なお、クラスを統合する方法として、提案手法ではグループ lasso, GFMR ではグループ連結 lasso (group fused lasso) (Bleakley and Vert, 2011; Alaíz et al., 2013) を用いていることに注意されたい。

推定アルゴリズムは、交互方向乗数法により得ている。更新式の詳細については Price et al. (2019) を参照されたい。正則化パラメータの値は、 K 分割交差検証法により選択している。具体的には、 K 分割したデータセットのうち k 番目のデータセットを $\mathcal{V}_k$ としたとき、評価関数

$$(3.7) \quad \sum_{k=1}^K \sum_{i \in \mathcal{V}_k} \sum_{l=1}^J y_{il} \log(\hat{\pi}_l^{(-\mathcal{V}_k)}(\mathbf{x}_i, \lambda))$$

を最大にする正則化パラメータの値を採用している．ここで， y_{il} は(2.2)と同様にデータ i がクラス l に属しているときに 1，その他のときに 0 をとる指示変数である． $\hat{\pi}_l^{(-V_k)}(\mathbf{x}_i, \lambda)$ は， k 番目のデータセットを除いたデータセットにより推定したモデルの事後確率を表す．

4. 数値実験

本節では，モンテカルロ・シミュレーションと実データへの適用を通して，提案手法の有効性を検証する．

4.1 モンテカルロ・シミュレーション

真のモデルを

$$(4.1) \quad \log \frac{\pi_j^*(\mathbf{x}_i^\dagger)}{\pi_{j+1}^*(\mathbf{x}_i^\dagger)} = \beta_j^{*\top} \mathbf{x}_i, \quad (i = 1, \dots, n, j = 1, \dots, J-1)$$

と設定した．ここで，回帰係数パラメータ $\beta_j^* = (\beta_{j0}^*, \beta_{j1}^*, \dots, \beta_{j10}^*)^\top$ の設定は次の 8 通りとした．

- 設定 1: $(\beta_1^*, \beta_2^*, \beta_3^*) = (\mathbf{0}, \delta, \mathbf{0})$
- 設定 2: $(\beta_1^*, \beta_2^*, \beta_3^*) = (\mathbf{0}, \delta, \delta)$
- 設定 3: $(\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*) = (\mathbf{0}, \delta, \delta, \delta)$
- 設定 4: $(\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*) = (\mathbf{0}, \delta, \delta, \mathbf{0})$
- 設定 5: $(\beta_1^*, \beta_2^*, \beta_3^*)^\top = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0.5 & 1 & 0.5 & 1 & 0.5 & 1 & 0.5 & 1 & 0.5 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$
- 設定 6: $(\beta_1^*, \beta_2^*, \beta_3^*)^\top = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0.5 & 0.5 & 0.5 \\ 1 & 1 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}$
- 設定 7: $(\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*)^\top = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0.5 & 0.5 & 0.5 \\ 1 & 1 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 1 & 0.5 & 1 & 0.5 & 1 & 0.5 & 1 & 0.5 & 1 & 0.5 & 1 & 1 \end{pmatrix}$
- 設定 8: $(\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*)^\top = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0.5 & 0.5 & 0.5 \\ 1 & 1 & 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

ただし， $\delta = (\underbrace{\delta, \dots, \delta}_{11})$ であり， $\delta = 0.5, 1$ の 2 通りを考えた．

ここで，設定したパラメータの意味について，設定 1 を例に取って説明する．設定 1 では $\beta_1^* = \beta_3^* = \mathbf{0}$ であることから，クラス 1 とクラス 2，クラス 3 とクラス 4 はそれぞれ事後確率が等しいことがわかる．すると，3.1 節の議論より，クラス 1 とクラス 2，クラス 3 とクラス 4 はそれぞれ同じクラスであり，実際には 2 クラスしかないことがわかる．つまり，設定 1 は，見かけ上は 4 クラスであるが，実際は 2 クラスしかないシミュレーション設定である．これより，すべての設定における実際のクラスの個数と見かけ上のクラスの個数は，表 1 のようにまとめることができる．

表 1. 実際のクラスの個数と見かけ上のクラスの個数.

	設定 1	設定 2	設定 3	設定 4	設定 5	設定 6	設定 7	設定 8
実際のクラスの個数	2	3	4	3	2	3	4	3
見かけ上のクラスの個数	4	4	5	5	4	4	5	5

説明変数に関するデータ $\mathbf{x}_i^\dagger$ は $N(\mathbf{0}, \Sigma)$ から発生させた. ここで, Σ は 10×10 型の分散共分散行列であり, その (i, j) 成分を $\zeta^{|i-j|}$ とし, $\zeta = 0, 0.5$ の 2 通りを考えた. (4.1) 式より計算される真の事後確率 $\pi_j^*(\mathbf{x}_i^\dagger)$ ($j = 1, \dots, J-1$) から, 乱数を用いて各データ $\mathbf{x}_i^\dagger$ ($i = 1, \dots, n$) が所属するクラスを決定した. サンプルサイズは $n = 200, 300, 500$ とし, シミュレーション回数は 100 回とした.

比較手法として, Price et al. (2019) による GFMR と 2 節で述べた ACL を用いた. ACL に含まれるパラメータはリッジ推定により求め, 正則化パラメータの値は 10^{-5} に固定した. 提案手法と GFMR に含まれる正則化パラメータの値は, 3.3 節の BIC により選択した. 正則化パラメータの値の候補はベイズ最適化 (Snoek et al., 2012) を用い, 提案手法では 0.1 から 7, GFMR では 0.1 から 100 の範囲で探索した. ベイズ最適化の計算には統計ソフトウェア R のパッケージ **rBayesianOptimization** を用いた. GFMR の候補集合 $\mathcal{L}$ については, 順序付きカテゴリカルデータを解析する際に Price et al. (2019) で推奨されている $\mathcal{L} = \{(1, 2), (2, 3), \dots, (J-1, J)\}$ を用いた.

クラス統合後のクラスの個数, 統合正答回数, 真の事後確率と推定した事後確率間のカルバック・ライブラー情報量 (KL) の 3 つの観点からそれぞれの手法を評価した. 100 回のシミュレーションにおいて, クラスが正しく統合された回数を, 統合正答回数とした. ここで, 各シミュレーションにおいてクラスが正しく統合されるとは, 見かけ上のクラスが実際のクラスに統合されることをいう. 例えば, 設定 1 においては, クラス 1 とクラス 2, クラス 3 とクラス 4 がそれぞれ統合されるときは正しく統合されているといえるが, クラス 1 とクラス 2, クラス 2 とクラス 3 がそれぞれ統合されるときは正しく統合されているとはいえない. また, KL は

$$\text{KL}(\hat{\pi}, \pi^*) = \sum_{j=1}^J \log \left(\frac{\hat{\pi}_j(\mathbf{x}_i^\dagger, \lambda)}{\pi_j^*(\mathbf{x}_i^\dagger)} \right) \hat{\pi}_j(\mathbf{x}_i^\dagger, \lambda)$$

と定義される. ただし, $\hat{\pi}_j(\mathbf{x}_i^\dagger, \lambda)$ は推定したモデルから求めた事後確率である. 100 回のシミュレーション毎にテストデータを 100 個発生させて KL を計算し, 得られた KL の平均と標準偏差を計算した.

表 2, 表 3, 表 4 はそれぞれ $\zeta = 0$ とした場合の設定 1 と設定 2, 設定 3 と設定 4, 設定 5 から設定 8 に対する計算結果である. また, 表 5, 表 6, 表 7 はそれぞれ $\zeta = 0.5$ とした場合の設定 1 と設定 2, 設定 3 と設定 4, 設定 5 から設定 8 に対する計算結果である. なお, 設定 3 の $n = 200, \delta = 1, \zeta = 0.5$ においては, リッジ推定値が発散して計算することができなかった.

まず, $\zeta = 0$ のときの結果を考察する. 表 2, 表 3, 表 4 から, ほぼすべての場合において, 提案手法は ACL よりも小さい KL の値を与えており, よりあてはまりの良いモデルを構築できていることがわかる. また, GFMR と比較しても, 同様に, 提案手法がよりあてはまりの良いモデルを構築できていることがわかる. 統合正答回数の観点から見ると, サンプルサイズが 200 の場合には GFMR が他の手法よりも多いが, サンプルサイズが大きくなるにつれ提案手法が多くなる.

次に, $\zeta = 0.5$ のときの結果を考察する. 表 5, 表 6, 表 7 から, $\zeta = 0$ のときに比べ GFMR

表 2. 設定 1 から設定 2 に対する結果($\zeta = 0$). 実験の各設定において, KL が一番小さい値, 統合正答回数が一番多い値を太字で表している.

	n	δ		提案手法			GFMR			ACL
				クラス個数	KL	統合正答回数	クラス個数	KL	統合正答回数	
設定 1	200	0.5	mean	1.58	34.50	31	1.97	10.86	97	17.72
			sd	0.74	23.27		0.17	8.80		4.66
		1	mean	2.69	11.51	51	2.08	11.76	92	18.84
			sd	0.83	15.76		0.27	5.07		4.90
	300	0.5	mean	1.96	21.53	71	1.93	21.37	76	17.79
			sd	0.57	27.91		0.57	27.98		3.68
		1	mean	2.16	9.35	87	2.09	12.55	92	19.00
			sd	0.44	4.03		0.32	5.59		4.34
	500	0.5	mean	2.02	8.77	99	2.15	24.67	73	17.39
			sd	0.20	3.74		0.78	42.08		4.47
		1	mean	2.00	8.94	100	2.04	11.33	96	17.67
			sd	0.00	4.13		0.20	4.95		4.66
設定 2	200	0.5	mean	3.30	16.17	58	2.66	18.95	66	18.78
			sd	0.67	21.99		0.48	8.24		4.32
		1	mean	3.11	16.14	89	3.02	22.69	96	18.90
			sd	0.31	4.14		0.20	10.00		4.69
	300	0.5	mean	3.09	11.42	91	2.87	16.91	87	18.10
			sd	0.29	3.94		0.34	8.60		4.39
		1	mean	3.01	15.60	99	3.02	23.65	98	19.08
			sd	0.10	4.40		0.14	11.66		4.69
	500	0.5	mean	3.01	12.28	99	2.81	40.43	80	17.79
			sd	0.10	3.97		0.66	84.28		4.19
		1	mean	3.00	14.66	100	3.00	25.67	100	17.32
			sd	0.00	3.51		0.00	15.40		3.80

はより良い結果を与えているが, 総合的に見て提案手法の方が KL の値は小さい. また, 統合正統回数もサンプルサイズが大きくなるにつれ提案手法が GFMR より大きい値をとっていることがわかる. 以上より, 提案手法は従来手法よりも KL を小さくし, さらにサンプルサイズが大きい場合は正確にクラス統合を実行することがわかる.

提案手法において BIC により選択されるモデルと K 分割交差検証法(CV)により選択されるモデルとを比較した. 交差検証法には(3.7)式を用い, 分割数は $K = 5$ とした. KL と統合正答回数の 2 つの観点から評価し, KL については BIC による結果を分子, CV による結果を分母とし, 統合正答回数については CV による結果を分子, BIC による結果を分母としてそれらの割合を計算した. これらより, KL と統合正答回数のどちらの指標においても, 1 より小さい値の場合は本稿で用いた BIC の方がより良いといえる.

設定 1 から設定 4 の結果を表 8, 設定 5 から設定 8 の結果を表 9 に載せている. KL については, ときおり 1 より大きい値をとっているが, 基本的には 1 より小さい. 統合正答回数については, ほとんどすべての場合において, 1 より小さい値をとっている. したがって, BIC の方が KL を小さくし, 統合正答回数を多くするモデルを選択することがわかる.

4.2 実データへの適用

赤ワインの品質ランクデータ (Cortez et al., 2009) に対し提案手法を適用した. 本データは,

表 3. 設定 3 から設定 4 に対する結果 ($\zeta = 0$). 実験の各設定において, KL が一番小さい値, 統合正答回数が一番多い値を太字で表している.

	n	δ		提案手法			GFMR			ACL
				クラス個数	KL	統合正答回数	クラス個数	KL	統合正答回数	
設定 3	200	0.5	mean	4.45	16.29	55	3.22	31.27	35	25.39
			sd	0.50	4.19		0.75	13.93		5.89
		1	mean	4.22	29.17	78	3.91	43.80	85	29.27
			sd	0.42	5.36		0.38	17.99		7.23
	300	0.5	mean	4.15	16.07	85	3.48	30.73	55	24.01
			sd	0.36	3.90		0.63	16.36		4.81
		1	mean	4.01	24.32	99	3.78	37.81	74	26.86
			sd	0.10	4.77		0.46	17.70		5.91
	500	0.5	mean	4.01	16.63	99	3.70	26.62	68	23.93
			sd	0.10	4.72		0.48	14.78		5.45
		1	mean	4.00	23.45	100	3.86	38.93	86	24.50
			sd	0.00	4.69		0.35	18.53		4.58
設定 4	200	0.5	mean	3.53	41.90	18	2.39	21.63	39	24.77
			sd	1.49	53.07		0.49	8.01		5.82
		1	mean	3.87	20.73	38	3.00	27.05	86	28.35
			sd	0.79	4.22		0.38	12.14		6.19
	300	0.5	mean	3.64	18.69	51	2.69	24.22	71	24.67
			sd	0.88	28.73		0.51	28.92		5.15
		1	mean	3.12	17.63	90	3.00	25.31	94	26.86
			sd	0.38	5.52		0.25	13.42		6.17
	500	0.5	mean	3.13	14.17	87	3.03	41.57	61	23.68
			sd	0.39	7.04		0.97	82.46		5.98
		1	mean	3.01	16.04	99	3.01	27.82	99	23.47
			sd	0.10	4.76		0.10	11.44		5.27

UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>) から取得できる. 赤ワインの品質は 6 段階で評価されており, 段階が上がるにつれて赤ワインの品質の評価は高くなる. つまり, 1 段階目の赤ワインの品質が最も評価が低く, 6 段階目に属する赤ワインの品質が最も評価が高い. この評価された段階を目的変数とし, 以降段階をクラスと呼ぶ. 説明変数には, 酒石酸濃度や残留糖分濃度などの赤ワインの成分に対する 11 種類の変数を用いた. 各クラスのサンプルサイズは表 10 の通りである.

比較手法として GFMR を考え, BIC より正則化パラメータの値を選択した. ここでは (3.6) 式の自由度に含まれるリッジ推定値が発散したため, 自由度 $df = (p+1) \left(\sum_{j=1}^{J-1} I(\|\hat{\beta}_j\|_2 > 0) \right)$ をもつ BIC を用いた. 正則化パラメータの値の候補は, 提案手法では 0.1 から 7, GFMR では 0.1 から 200 の範囲からベイズ最適化を用いて探索した. GFMR の候補集合 $\mathcal{L}$ については, 4.1 節と同様に $\mathcal{L} = \{(1, 2), (2, 3), \dots, (J-1, J)\}$ を用いた.

ブートストラップ法 (Efron and Tibshirani, 1993) を用いてクラス統合の安定性を検証した. まずクラス 1 をクラス 1a とクラス 1b に, サンプルサイズが同じになるようランダムに 2 分割し, このデータセットからブートストラップ標本を発生させて提案手法と GFMR を当てはめた. この操作を 100 回繰り返し, クラス 1a とクラス 1b が統合された回数 (統合正答回数) を数え上げた. なお, 統合正答回数の数え上げでは, 各ブートストラップ標本においてクラス 1a とクラス 1b だけを統合するときに限って数え上げることに注意しておく. クラス 2 からクラス

表 4. 設定 5 から設定 8 の結果($\zeta = 0$). 実験の各設定において, KL が一番小さい値, 統合正答回数が一番多い値を太字で表している.

n	提案手法				GFMR			ACL
		クラス個数	KL	統合正答回数	クラス個数	KL	統合正答回数	KL
設定 5	200	mean	2.72	10.95	2.03	12.05	97	18.82
		sd	0.87	13.29	0.17	4.743		5.12
	300	mean	2.11	8.77	2.05	11.19	95	18.20
		sd	0.37	3.76	0.22	4.87		4.56
	500	mean	2.00	8.72	2.04	11.52	97	17.55
		sd	0.00	3.99	0.24	4.27		4.71
設定 6	200	mean	3.14	13.66	2.72	22.59	70	20.10
		sd	0.35	3.84	0.47	9.76		5.53
	300	mean	3.00	12.65	2.78	22.76	78	18.53
		sd	0.00	3.62	0.42	10.75		4.57
	500	mean	3.00	12.83	2.96	18.38	96	17.43
		sd	0.00	4.22	0.20	10.11		4.75
設定 7	200	mean	4.24	20.49	3.65	35.30	67	26.07
		sd	0.43	4.69	0.63	15.60		6.86
	300	mean	4.00	18.44	3.79	32.95	79	24.52
		sd	0.00	3.95	0.41	15.53		5.54
	500	mean	4.00	19.09	3.83	32.26	83	23.85
		sd	0.00	4.10	0.38	13.32		5.89
設定 8	200	mean	4.09	15.73	2.60	24.82	56	26.02
		sd	0.82	4.30	0.53	9.57		5.69
	300	mean	3.22	14.54	2.78	21.80	76	24.85
		sd	0.46	4.36	0.44	9.14		5.35
	500	mean	3.05	14.65	2.93	18.63	93	23.63
		sd	0.22	4.67	0.26	10.24		5.94

6 に対しても同様の手順でブートストラップ標本を発生させ, 統合正答回数を数え上げた.

結果を表 11 にまとめている. クラス 2 とクラス 5 以外の結果については, 提案手法は統合正答回数が 80 回以上であり, GFMR より多くなっている. 特に, クラス 1 とクラス 6 の結果においては, 提案手法の統合正答回数は GFMR の 2 倍以上である. 一方, クラス 2 やクラス 5 では, そもそも統合正答回数が少ない. したがって, このときの結果についてはあまり議論すべきではないかも知れないが, クラス 2 においては提案手法の統合正答回数は GFMR より多くなっている. 以上より, 統合正答回数の観点から, 提案手法は GFMR よりも多くの場合において安定してクラス統合が行われていることがわかる.

5. まとめ

隣接カテゴリロジットモデルとグループ lasso により, 順序付きカテゴリカルデータを対象とするモデルにおけるクラス統合の方法を提案した. 推定アルゴリズムを交互方向乗数法により構成し, グループ lasso により推定されたモデルの自由度をもつ情報量規準 BIC により正則化パラメータの値を選択した. 数値実験の結果をいくつかの観点から考察することにより, 従来手法に比べて提案手法は多くの点で有用であることがわかった.

4.1 節のモンテカルロ・シミュレーションにおいて, サンプルサイズが小さいときに, 提案

表 5. 設定 1 から設定 2 の結果($\zeta = 0.5$). 実験の各設定において, KL が一番小さい値, 統合正答回数が一番多い値を太字で表している.

	n	δ		提案手法			GFMR			ACL
				クラス個数	KL	統合正答回数	クラス個数	KL	統合正答回数	
設定 1	200	0.5	mean	2.95	16.83	33	2.01	8.00	99	18.91
			sd	1.02	26.89		0.10	4.21		5.13
		1	mean	2.47	8.42	68	2.08	11.45	93	20.18
			sd	0.74	3.55		0.31	5.35		5.20
	300	0.5	mean	2.67	9.45	59	2.12	10.36	88	18.43
			sd	0.87	5.05		0.41	16.67		4.62
		1	mean	2.06	7.19	95	2.08	10.32	92	18.57
			sd	0.28	3.62		0.27	4.78		4.83
	500	0.5	mean	2.03	7.24	98	2.36	28.47	69	17.07
			sd	0.22	3.50		0.87	70.34		4.28
		1	mean	2.00	7.65	100	2.05	10.11	95	17.84
			sd	0.00	3.62		0.22	4.56		5.12
設定 2	200	0.5	mean	3.51	10.76	49	2.99	14.95	95	19.69
			sd	0.50	3.46		0.22	9.81		5.30
		1	mean	3.18	16.73	82	3.02	23.00	98	21.77
			sd	0.39	3.93		0.14	10.56		6.26
	300	0.5	mean	3.21	10.94	79	2.98	14.50	96	18.94
			sd	0.41	4.10		0.20	10.91		4.98
		1	mean	3.03	13.72	97	3.03	20.50	97	19.99
			sd	0.17	3.47		0.17	10.40		5.68
	500	0.5	mean	3.03	11.00	97	3.02	28.11	92	18.35
			sd	0.17	4.09		0.38	84.41		5.01
		1	mean	3.00	13.60	100	3.00	23.85	100	18.19
			sd	0.00	3.49		0.00	12.82		4.67

手法が GFMR に統合正答回数で劣ることがあった. この理由としては, 次の 2 点が考えられる. 1 点目は, サンプルサイズとクラス間の順序との関係性である. 一般に, サンプルサイズが大きい場合には各クラスの性質が明確になるため, クラス間の順序が強調されることが知られている. 一方, サンプルサイズが小さい場合には各クラスの性質が明確にならないため, クラス間の順序が強調されず, 提案手法の統合正答回数は少なくなったと考えられる. 2 点目は, 正則化パラメータの選択方法である. サンプルサイズが小さい場合には, 統合後のクラスの個数が真のクラスの個数よりも大きくなり, その結果統合正答回数が少なくなる状況が見受けられた. これは, 正則化パラメータの値が小さめに選択されていることを表しており, BIC を用いた選択方法がうまく機能していないことを示唆している.

4.2 節の実データへの適用において, クラス 2 とクラス 5 の統合正答回数が非常に少なかった. これは次の 2 つの理由が重なったため生じたと考えられる. 1 つ目は, クラス 2 とクラス 5 はそれぞれ 2 つのクラスに挟まれており, 統合する隣接クラスの候補が多いということ. 2 つ目は, クラス 3 とクラス 4 に比べて, サンプルサイズが小さいということである. このような状況において, 統合正答回数を改善する統計モデルの構築に取り組むことは大変意義がある.

本稿で用いた BIC におけるモデルの自由度は Yuan and Lin (2006) の考えに基づいて導出したが, Yuan and Lin (2006) で導出されている自由度は線形回帰モデルかつ計画行列の各列が直交するという仮定が課されている. この点において提案モデルと乖離しており, 提案モデルの

表 6. 設定 3 から設定 4 の結果($\zeta = 0.5$). 実験の各設定において, KL が一番小さい値, 統合正答回数が一番多い値を太字で表している.

	n	δ		提案手法			GFMR			ACL
				クラス個数	KL	統合正答回数	クラス個数	KL	統合正答回数	KL
設定 3	200	0.5	mean	4.54	15.70	46	3.65	29.13	66	27.11
			sd	0.50	4.17		0.50	17.49		7.33
		1	mean	—	—	—	—	—	—	—
			sd	—	—	—	—	—	—	—
	300	0.5	mean	4.29	14.51	71	3.49	23.84	49	26.21
			sd	0.46	4.28		0.50	12.94		6.48
		1	mean	4.09	26.10	91	3.72	48.24	64	28.23
			sd	0.29	5.37		0.53	26.82		5.44
	500	0.5	mean	4.06	15.10	94	3.65	27.45	65	24.32
			sd	0.24	4.33		0.48	15.35		5.46
		1	mean	4.00	21.39	100	3.50	44.59	48	25.22
			sd	0.00	4.06		0.52	18.69		6.07
設定 4	200	0.5	mean	4.57	14.02	11	2.77	20.74	75	27.09
			sd	0.69	4.03		0.45	14.32		6.75
		1	mean	3.75	20.85	53	3.07	27.91	93	29.66
			sd	0.87	4.90		0.26	9.68		6.95
	300	0.5	mean	3.67	12.56	52	3.00	14.96	94	25.32
			sd	0.78	4.49		0.25	9.13		5.58
		1	mean	3.19	14.83	84	3.03	23.54	97	26.43
			sd	0.46	4.68		0.17	12.98		5.79
	500	0.5	mean	3.14	12.16	91	3.12	20.96	92	23.56
			sd	0.47	4.34		0.56	67.52		4.77
		1	mean	3.00	14.37	100	3.04	28.18	96	24.47
			sd	0.00	4.04		0.20	13.90		5.64

自由度についてはまだ多くの問題が残っている. 以上は今後の研究課題としたい.

謝 辞

本稿の改訂に当たって, 編集委員および査読者の方から大変貴重なご指摘と適切なお意見をいただきました. ここに記して御礼申し上げます. 本研究は日本学術振興会科学研究費補助金(19K11854)の助成を受けたものです.

付 録

A. 隣接カテゴリーロジットモデルの対数尤度関数の 1 次導関数および 2 次導関数

(2.3) 式の隣接カテゴリーロジットモデルの対数尤度関数は,

$$(A.1) \quad \ell(\beta_1, \dots, \beta_{J-1}) = \sum_{i=1}^n \left\{ \mathbf{1}_{J-1}^\top Y_i^{**} X_i^* \beta - \log \left[1 + \sum_{j=1}^{J-1} \exp \left(\mathbf{I}_j^\top X_i^* \beta \right) \right] \right\}$$

と書き換えることができる. ここで, $\mathbf{1}_{J-1}$ は要素がすべて 1 の $(J-1)$ 次元ベクトル, $\beta = (\beta_1^\top, \dots, \beta_{J-1}^\top)^\top$ は $(J-1)(p+1)$ 次元ベクトル, $\mathbf{I}_j = (I(1 > j-1), \dots, I(J-1 > j-1))^\top$

表 7. 設定 5 から設定 8 の結果 ($\zeta = 0.5$). 実験の各設定において, KL が一番小さい値, 統合正答回数が一番多い値を太字で表している.

	n		提案手法			GFMR			ACL
			クラス個数	KL	統合正答回数	クラス個数	KL	統合正答回数	
設定 5	200	mean	2.71	8.82	54	2.11	11.41	90	19.66
		sd	0.84	3.98		0.35	4.76		5.70
	300	mean	2.12	7.76	90	2.03	10.76	97	18.74
		sd	0.38	4.01		0.17	5.20		5.34
	500	mean	2.00	7.10	100	2.06	10.24	94	17.51
		sd	0.00	3.00		0.24	3.81		4.41
設定 6	200	mean	3.27	11.78	73	3.00	18.48	96	19.91
		sd	0.45	3.33		0.20	8.68		4.68
	300	mean	3.08	12.03	92	3.01	18.19	99	19.38
		sd	0.27	4.29		0.10	6.26		5.56
	500	mean	3.00	11.81	100	3.00	17.84	100	18.20
		sd	0.00	3.58		0.00	7.04		4.35
設定 7	200	mean	4.47	21.99	53	3.80	33.77	72	30.34
		sd	0.50	4.84		0.49	13.62		7.04
	300	mean	4.08	17.13	92	3.55	31.79	55	26.21
		sd	0.27	4.32		0.50	12.12		5.70
	500	mean	4.01	17.41	99	3.63	34.64	61	25.76
		sd	0.10	3.97		0.51	13.59		5.71
設定 8	200	mean	3.87	13.88	43	2.90	22.75	73	26.54
		sd	0.85	4.13		0.54	12.83		5.53
	300	mean	3.30	13.00	77	2.98	20.17	92	25.58
		sd	0.59	4.59		0.28	11.79		6.37
	500	mean	3.01	12.36	99	2.99	17.32	99	23.87
		sd	0.10	3.75		0.10	9.91		5.18

は $(J-1)$ 次元ベクトル, Y_i^{**} は

$$Y_i^{**} = \begin{pmatrix} y_{i1} & y_{i1} & \cdots & y_{i1} \\ 0 & y_{i2} & \cdots & y_{i2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & y_{iJ-1} \end{pmatrix}$$

となる $(J-1) \times (J-1)$ 型の正方行列, X_i^* は

$$X_i^* = \begin{pmatrix} \mathbf{x}_i^\top & 0 \cdots 0 & \cdots & 0 \\ 0 \cdots 0 & \mathbf{x}_i^\top & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 \cdots 0 & 0 \cdots 0 & \cdots & \mathbf{x}_i^\top \end{pmatrix}$$

となる $(J-1) \times (J-1)(p+1)$ 型の行列である. このとき, (A.1) 式の数尤度関数の 1 次導関数および 2 次導関数は, 順に

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^n \left\{ (\mathbf{1}_n^\top Y_i^{**} X_i^*)^\top - \frac{1}{1 + \sum_{j=1}^{J-1} \eta_{ij}} \sum_{j=1}^{J-1} \eta_{ij} \gamma_{ij}^\top \right\},$$

表 8. 設定 1 から設定 4 における正則化パラメータ選択法の比較. 実験の各設定において, 1 より小さい値を太字で表している.

ζ	n	δ	設定 1		設定 2		設定 3		設定 4	
			KL	統合正答回数	KL	統合正答回数	KL	統合正答回数	KL	統合正答回数
0	200	0.5	2.629	0.677	1.183	0.828	0.802	0.855	2.526	0.778
		1	1.104	0.824	0.991	0.685	1.048	0.603	1.006	0.579
	300	0.5	1.990	0.817	0.730	0.846	0.711	0.918	1.199	0.765
		1	0.946	0.851	1.062	0.828	1.098	0.818	1.006	0.800
	500	0.5	0.854	0.788	0.782	0.919	0.965	0.909	0.916	0.874
		1	0.999	0.950	1.148	0.980	1.168	0.910	1.103	0.899
0.5	200	0.5	1.579	0.545	0.811	0.878	1.004	0.739	0.724	1.000
		1	0.894	0.691	0.996	0.720	—	—	—	—
	300	0.5	0.928	0.831	0.735	0.911	0.537	0.859	0.720	0.788
		1	0.946	0.895	1.026	0.804	1.297	0.802	0.992	0.798
	500	0.5	0.923	0.939	0.953	0.959	0.330	0.894	0.951	0.945
		1	0.995	0.960	1.105	0.940	1.134	0.940	1.081	0.940

表 9. 設定 5 から設定 8 における正則化パラメータ選択法の比較. 実験の各設定において, 1 より小さい値を太字で表している.

ζ	n	設定 5		設定 6		設定 7		設定 8	
		KL	統合正答回数	KL	統合正答回数	KL	統合正答回数	KL	統合正答回数
0	200	1.044	0.692	0.989	0.814	0.980	0.803	0.993	1.138
	300	0.975	0.901	0.990	0.900	1.020	0.880	0.989	0.800
	500	0.955	0.930	1.006	0.970	1.012	0.910	0.986	0.895
0.5	200	0.940	0.815	0.954	0.712	1.064	0.811	0.743	0.698
	300	0.920	0.878	0.977	0.870	1.032	0.859	0.938	0.727
	500	0.948	0.950	0.994	0.960	1.011	0.919	0.993	0.949

表 10. 各クラスのサンプルサイズ.

	クラス 1	クラス 2	クラス 3	クラス 4	クラス 5	クラス 6
サンプルサイズ	10	53	681	638	199	18

表 11. 各クラスのブートストラップ選択回数. 各クラスにおいて統合正答回数が多い方を太字にしている.

	クラス 1	クラス 2	クラス 3	クラス 4	クラス 5	クラス 6
提案手法	80	5	100	98	14	81
GFMR	38	2	87	79	20	37

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta^\top} = \sum_{i=1}^n \left[\frac{1}{\left(1 + \sum_{j=1}^{J-1} \eta_{ij}\right)^2} \left(\sum_{j=1}^{J-1} \eta_{ij} \gamma_{ij} \right) \left(\sum_{j=1}^{J-1} \eta_{ij} \gamma_{ij}^\top \right) - \frac{1}{1 + \sum_{j=1}^{J-1} \eta_{ij}} \sum_{j=1}^{J-1} \eta_{ij} \gamma_{ij} \gamma_{ij}^\top \right]$$

となる. ここで, $\eta_{ij} = \exp(\mathbf{I}_j^\top \mathbf{X}_i^* \beta)$, $\gamma_{ij} = \mathbf{I}_j^\top \mathbf{X}_i^*$ と置いている.

参 考 文 献

- Agresti, A. (1992). Analysis of ordinal paired comparison data, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **41**(2), 287–297.
- Agresti, A. (2010). *Analysis of Ordinal Categorical Data (Second Edition)*, John Wiley & Sons, New York.
- Alaíz, C. M., Barbero, A. and Dorronsoro, J. R. (2013). Group fused lasso, *International Conference on Artificial Neural Networks*, 66–73, Springer, Berlin, Heidelberg.
- Ananth, C. V. and Kleinbaum, D. G. (1997). Regression models for ordinal responses: A review of methods and applications, *International Journal of Epidemiology*, **26**(6), 1323–1333.
- Anderson, J. A. (1984). Regression and ordered categorical variables, *Journal of the Royal Statistical Society: Series B*, **46**(1), 1–22.
- Armstrong, B. G. and Sloan, M. (1989). Ordinal regression models for epidemiologic data, *American Journal of Epidemiology*, **129**(1), 191–204.
- Bleakley, K. and Vert, J.-P. (2011). The group fused lasso for multiple change-point detection, arXiv preprint, arXiv:1106.4199.
- Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends® in Machine Learning*, **3**(1), 1–122.
- Cortez, P., Cerdeira, A., Almeida, F., Matos, T. and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties, *Decision Support Systems*, **47**(4), 547–553.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall/CRC, New York.
- Fienberg, S. E. (2007). *The Analysis of Cross-classified Categorical Data*, Springer, New York.
- Fienberg, S. E. and Mason, W. M. (1979). Identification and estimation of age-period-cohort models in the analysis of discrete archival data, *Sociological Methodology*, **10**, 1–67.
- Goodman, L. A. (1984). *The Analysis of Cross-classified Data Having Ordered Categories*, Harvard University Press, Cambridge, Massachusetts.
- 川野秀一, 松井秀俊, 廣瀬慧 (2018). 『スパース推定法による統計モデリング』, 共立出版, 東京.
- McCullagh, P. (1980). Regression models for ordinal data, *Journal of the Royal Statistical Society: Series B*, **42**(2), 109–127.
- Price, B. S., Geyer, C. J. and Rothman, A. J. (2019). Automatic response category combination in multinomial logistic regression, *Journal of Computational and Graphical Statistics*, **28**(3), 758–766.
- Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**(2), 461–464.
- Snoek, J., Larochelle, H. and Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms, *Advances in Neural Information Processing Systems* **25**, 2951–2959.
- Whitehead, A., Omar, R. Z., Higgins, J. P., Savaluny, E., Turner, R. M. and Thompson, S. G. (2001). Meta-analysis of ordinal outcomes using individual patient data, *Statistics in Medicine*, **20**(15), 2243–2260.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B*, **68**(1), 49–67.

Fusing Adjacent Classes in an Ordinal Logistic Model via Group Regularization

Mizuho Naganuma¹, Kohei Yoshikawa² and Shuichi Kawano²

¹Graduate School of Informatics and Engineering, The University of Electro-Communications;
Now at Macromill, Inc.

²Graduate School of Informatics and Engineering, The University of Electro-Communications

This paper aims to fuse adjacent classes in an ordinal logistic model in light of the multi-class classification problem. Fusing the classes enables us to easily interpret the constructed model and remove irrelevant classes. Fusion of classes is performed when two adjacent classes have the same posterior probability. To this end, we developed an ordinal logistic model with group regularization for fusing adjacent classes. We established an estimation algorithm based on the alternating direction method of multipliers, and used Monte Carlo simulations and real data analysis to investigate the usefulness of our proposed method.