

TVACLMの動的計画法を用いた最適化



11/22 筑波大学, IBIS2022

伊従寛哉（東京大学情報理工学研究所）

武田優真（東京大学教養学部）

松島慎（東京大学総合文化研究科）

概要

順序回帰とは応答ラベルが順序尺度の問題のことであり、診断医学や社会科学などの分野で頻繁に扱われる。このような問題に対してはCLMやSVORなどのモデルが利用されてきたが、表現能力と解釈性の両方を持ち合わせているモデルは存在しなかった。そこで本研究では表現能力と解釈性を持ったTVACLMを提案する。実験では精度の比較によって提案手法が高い予測性能を持つことを示した。また提案アルゴリズムが最急降下法よりも優れていることを示した。

提案手法

TVACLM (全変動正則付き加法累積ロジットモデル)

- CLMに対して加法モデルを用いて拡張して、モデルの解釈性を維持したまま表現力を向上させる。
- 加法モデルに対して全変動正則化を用いてパラメトリックなモデルに落とし込む。

$$P(y \leq C_q | x) = \sigma(b_q - \sum_{j \in [d]} f_j(x_j))$$

$\gamma_{i,q} = b_q - \sum_{j \in [d]} f_j(\mathbf{X}_{ij})$ とおくと目的関数は、

$$J((f_j)_{j \in [d]}) = - \sum_{i=1}^N \sum_{q=1}^Q y_{i,q} \cdot \log(\sigma(\gamma_{i,q}) - \sigma(\gamma_{i,q-1})) + \lambda \sum_{j \in [d]} \text{TV}(f_j)$$

	CLM	RBF-SVOR	提案手法
解釈性	✓	×	✓
表現力	×	✓	✓

アルゴリズム

要素数 N のベクトル v_j を $v_j = f_j(\mathbf{X}_j)$ で定義する。 $J(v_j) = \|\mathbf{y} - v_j\|_2^2 + \lambda \text{TV}(v_j)$ という目的関数は動的計画法を用いて効率的に最小値をとる v_j を求めることができる[4]。これを利用してブロック座標降下法によって**TVACLM**を最適化する。

Input: データ行列 $\mathbf{X} \in \mathbb{R}^{N \times d}$, 応答行列 $\mathbf{y} \in [0, 1]^{N \times Q}$, 正則化係数 λ , 許容誤差 ϵ
初期化: $(v_j)_{j \in [d]} = 0$
while converge do
 Update $\mathbf{b}$
 for $j = 1, \dots, d$ **do**
 $\hat{\mathbf{y}} = \mathbf{v}_j - \frac{1}{L} \frac{\partial}{\partial \mathbf{v}_j} \ell(\sum_{j \in [d]} \mathbf{v}_j, \mathbf{y})$
 $\mathbf{v}_j = \text{FlsaDp}(\hat{\mathbf{y}}, \lambda)$ // $\mathbf{v}_j = \arg \min_{v_j} (\sum_{i \in [N]} \|\hat{\mathbf{y}} - v_j\|_2^2 + \lambda \text{TV}(v_j))$

背景と問題設定

順序回帰問題とは応答ラベルが順序尺度(e.g. 軽度、中度、重度)であるような問題のこと。

応答ラベル

$$\mathcal{Y} = \{C_q \mid q \in \mathbb{N}, 1 \leq q \leq Q, C_1 \prec C_2 \prec \dots \prec C_Q\}$$

を未知の入力ベクトル $\mathbf{x} \in \mathcal{X} = \mathbb{R}^d$ から予測する関数を教師データ $D = \{(\mathbf{X}_i, \mathbf{y}_i), i = 1, \dots, N\}$ を用いて発見する

順序回帰は、

- ラベル $\mathbf{y} \in \mathcal{Y}$ が計量的な情報を持たないため、**回帰問題とは異なる**。実数値にマッピングするための原理的手法も存在しない。
- ラベル $\mathbf{y} \in \mathcal{Y}$ が順序関係を持つため、**分類問題と異なる**。順序関係を無視すると予測性能が悪くなる。

順序回帰は診断医学や社会科学などで扱われ、解釈性が重要となる。

任意の入力 $\mathbf{x}$ に対して単変量 x_j についての出力への寄与を表す関数が同一な単変数関数 f_j である時、**解釈性を持つ**と定義する。

既存手法

CLM[1] (Cumulative Logit Model, 累積ロジットモデル)

ロジスティック回帰を順序回帰へ拡張

$$P(y \leq C_q | x) = \sigma(b_q - \mathbf{w}^T \mathbf{x})$$

σ は標準シグモイド関数

SVOR[2] (Support Vector Ordinal Regression)

SVMを順序回帰へ拡張

RBFカーネルSVORは順序回帰で最も予測性能が高いモデルの一つ[3]。

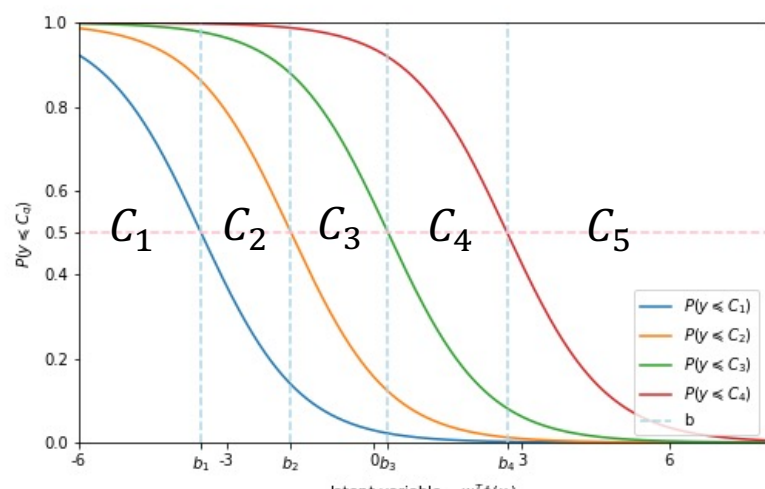


図1. CLMで潜在変数 $\mathbf{w}^T \mathbf{x}$ と各クラスに分類される確率の関係

CLMや線形カーネルSVORなど：モデルの**表現力**に乏しい
RBFカーネルSVOR、NNなど：**解釈性**を持たない

実験

実験1. 提案手法と既存手法の予測精度を比較する。

用いたデータセット: ORCA [5] 上のデータセットの内、 $N \geq 300$ のもの14 個

実験2. 提案アルゴリズムと最急降下法を比較する。

用いたデータセット: ランダムに生成したトイデータ

表1. 実験1の結果。提案手法の予測精度はCLMより優れており解釈性を持たないRBF SVOREXに迫っている。

	解釈性○	解釈性×
手法	提案手法 TVACLM	CLM RBF SVOREX
平均	0.517	0.488
$N \geq 300$ 平均	0.634	0.548

表2. 実験2の結果。 λ が小さくなくてもブロック座標降下法では比較的少ない反復回数で収束する。

手法	ブロック座標降下	最急降下
$\lambda = 0.01$	5099	244280
$\lambda = 0.1$	769	663220
$\lambda = 1$	69	56

参考文献

- Peter McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 42, No. 2, pp. 109–127, 1980.
- Wei Chu and S Sathya Keerthi. Support vector ordinal regression. *Neural computation*, Vol. 19, No. 3, pp. 792–815, 2007.
- Pedro Antonio Gutiérrez et al. Ordinal regression methods: survey and experimental study. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 1, 2015.
- Johnson, Nicholas A. "A dynamic programming algorithm for the fused lasso and l0-segmentation." *Journal of Computational and Graphical Statistics* 22.2 (2013): 246-260.
- Javier Sánchez-Monedero et al. Orca: A matlab/octave toolbox for ordinal regression. *Journal of Machine Learning Research*, Vol. 20, No. 125, pp. 1–5, 2019.