[小川研究奨励賞受賞記念講演] 表現能力の高い予測モデルを用いた統計的推定手法の開発

奥野彰文[1,2]

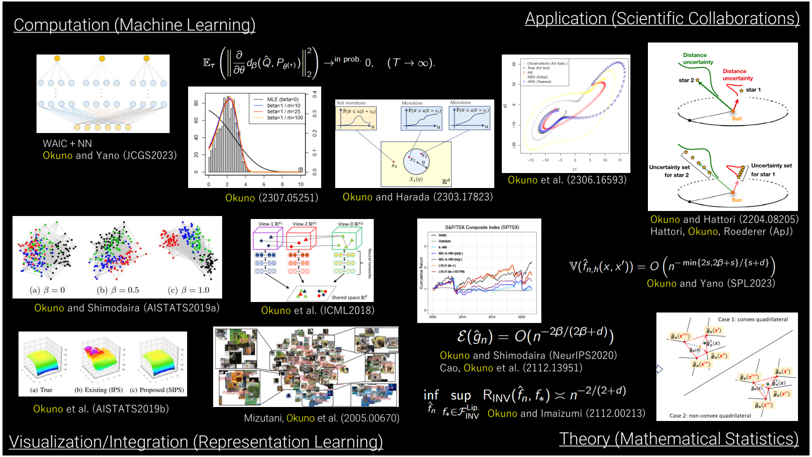
[1]統計数理研究所-助教　[2]理研AIP-客員研究員

経歴

- 2009年4月-2014年3月
  大阪大学基礎工学部　情報科学科
- 2014年4月-2017年3月
  大阪大学大学院　基礎工学研究科
- 2017年4月-2019年3月
  京都大学大学院　情報学研究科
- 2019年4月-2020年9月
  理化学研究所　革新知能統合研究センター@京大
- 2020年9月: 博士号 (情報学，京都大学)
- 2020年10月–
  統計数理研究所

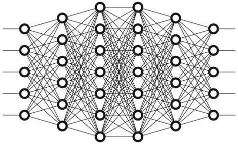
指導教員はずっと下平英寿先生でした.

基本的には何でもやります



▶ 核融合プラズマ企画セッション `https://statplasma.github.io/jjsm2024`

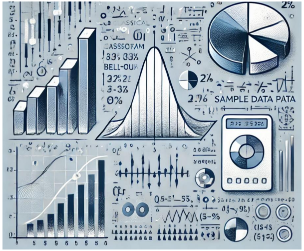
研究目標



PyTorch TensorFlow

Keras

ニューラルネット（深層学習）



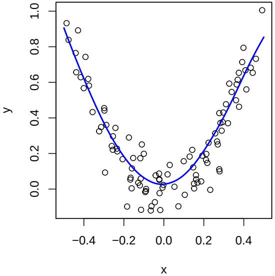
(DALL−E3によるイメージ図)

伝統的な統計学

▶ ニューラルネットで実装する小サンプル領域での統計学

なぜニューラルネットなのか?

- 汎用性が高く，ユーザが多く，プラットフォームが整備されている.
- 基底を学習できる.



- 基底学習による万能近似と，正則化によるモデル制約の両方が重要.
- ニューラルネットの自由度がどのくらいで，どうやって制約するかに興味がある.

今回の受賞対象論文

1. Akifumi Okuno and Keisuke Yano. “A generalization gap estimation for overparameterized models via the Langevin functional variance”, *Journal of Computational and Graphical Statistics*, 32(4):1287-1295, 2023.

2. Akifumi Okuno and Kazuharu Harada. “An interpretable neural network-based non-proportional odds model for ordinal regression”, *Journal of Computational and Graphical Statistics*, 2024. To appear.

3. Akifumi Okuno and Masaaki Imaizumi. “Minimax Analysis for Inverse Risk in Nonparametric Planer Invertible Regression”, *Electronic Journal of Statistics*, 18(1):355-394, 2024.

(1) WAICによるニューラルネットの汎化予測

対象論文

- Akifumi Okuno and Keisuke Yano. “A generalization gap estimation for overparameterized models via the Langevin functional variance”, *Journal of Computational and Graphical Statistics*, 32(4):1287-1295, 2023. `https://doi.org/10.1080/10618600.2023.2197488`
- 日本語での解説: Jxiv.537
- WAICをニューラルネットで使えないか調べ，計算法を提案した.

予測モデルの汎化 = まだ見ぬデータへの当てはまり

Pos(α)はパラメータαでリッジ正則化した，訓練データ$(\mathbf{X}, \mathbf{y})$での事後分布とすると

$$\text{Gibbs汎化損失}\quad \mathcal{G}(\alpha) = \mathbb{E}_{\mathbf{y}^*, \mathbf{y}}\left(\mathbb{E}_{\boldsymbol{\theta}\sim\mathrm{Pos}}\left[\sum_{i=1}^{n}\{y_i^* - f_{\boldsymbol{\theta}}(\mathbf{x}_i)\}^2\right]\right).$$

ただし$\mathbf{y}^*$はテストデータ.

現実で得られる訓練データだけから汎化を予測したい.

- Cross-Validation：何度も最適化しないといけない．3-foldなどよく使われる.
- 情報量規準：最適化が一回で済む.

$$\mathbb{E}[\text{情報量規準}] = \text{汎化損失} + o(1).$$

ニューラルネットに使ってみるとどうなるか?

- AIC = -対数尤度 + p
  - ニューラルネットのパラメータ数が大きすぎる.
  - そもそも特異モデルに使えない.
- TIC = -対数尤度 + $\mathrm{tr}\{IJ^{-1}\}$.
  - Jが退化して計算できない.
  - そもそも特異モデルに使えない.

WAIC = -対数尤度 + 汎関数分散.
- 特異モデル (ただし$p \leq n$) に使える.
- 過剰パラメータモデル ($p > n$)に使えるかわからない.

本研究の貢献

- (理論) 過剰パラメータ線形回帰で: $|\mathbb{E}[\mathrm{WAIC}] - \text{汎化誤差}| \to^p 0$ を証明.
- (計算) 一般モデルで: 勾配情報のみを用いたWAICの効率的計算法を提案.

標準正規乱数$\mathbf{e}^{(t)}$について,

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \frac{1}{4}\frac{\delta}{\sigma_0^2}\frac{\partial}{\partial\boldsymbol{\theta}}\left\{\sum_{i=1}^n (y_i - f_{\boldsymbol{\theta}^{(t)}}(\mathbf{x}_i))^2 + n\alpha\|\boldsymbol{\theta}^{(t)}\|_2^2\right\} + \delta^{1/2}\mathbf{e}^{(t)},$$

$$\text{LWAIC} = \text{対数尤度} + \sum_{i=1}^n \mathbb{V}[\{\log p(y_i \mid \mathbf{x}_i, \boldsymbol{\theta}^{(t)})\}_t].$$

- ニューラルネット+二乗損失の勾配のみから計算可能.
- $p \times p$行列(=2次)を用意するTICと異なり，1次の情報しか使わない.
- 原理上，最適化と同時に計算できる.

ニューラルネットでの実験

- f_θ: 1-hidden-layer-perceptron.
- tanhで活性化，中間素子数$M \in \{50, 100, 150\}$.
- d: 入力次元，n: サンプルサイズ，T: ランジュバンの反復回数.
- $\tilde{\Delta}$: 汎化ギャップ.
- パラメータ数 $M(d+2)$ が n を超えている設定を灰色で示した.

Table: 50回の実験での平均±標準偏差.

	$M=50$		$M=100$		$M=150$	
	LFV	$\tilde{\Delta}$	LFV	$\tilde{\Delta}$	LFV	$\tilde{\Delta}$
$d=5$	6.43 ± 0.96	4.26	6.49 ± 0.52	4.41	7.30 ± 0.80	4.60
$d=10$	11.03 ± 1.28	8.75	12.91 ± 1.54	9.18	13.56 ± 1.14	9.64
$d=15$	16.78 ± 1.39	17.64	18.93 ± 1.57	18.60	20.13 ± 2.07	19.46

(2) ニューラルネットを用いた整合的な非比例オッズモデル

対象論文

- Akifumi Okuno and Kazuharu Harada. "An interpretable neural network-based non-proportional odds model for ordinal regression", *Journal of Computational and Graphical Statistics*, 2024. To appear. `https://doi.org/10.1080/10618600.2024.2321208`
- 日本語での解説: Jxiv.549
- ニューラルネットを用いて整合的で解釈可能な非比例オッズモデルを提案

比例オッズモデル

応答の値そのものの予測でなく，閾値以上/以下である確率：

$$\mathbb{P}(H \leq u \mid X = \mathbf{x}) \text{ の予測がしたい.}$$

- 将棋で5段以上に上がるために重要な要素は? など.

離散値$G \in \{1, 2, \ldots, J\}$の場合：$j = 1, 2, \ldots, J$について

- 比例オッズモデル $\mathbb{P}_{\mathrm{POM}}(G \leq j \mid \mathbf{X} = \mathbf{x}) = \sigma(\alpha_j + \langle \mathbf{x}, \boldsymbol{\beta} \rangle)$.
  - jについて単調であってほしい
  - $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_J$ならOK.

閾値に依らず係数が常に一定 $\Rightarrow$ 否定的な例 (Long and Freese, 2006 など).

非比例オッズモデル

- 非比例オッズモデル $\mathbb{P}_{\mathrm{NPOM}}(G \leq j \mid \mathbf{X} = \mathbf{x}) = \sigma(\alpha_j + \langle \mathbf{x}, \boldsymbol{\beta}_j \rangle)$.
  - 👍 表現能力が高い.
  - 👎 $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_J$でも単調にならない場合がある.
  - 👎 単調性の条件がよく分かっておらず，アドホックにFused Lassoしたりする.

提案法

連続値 $H = [1, J]$について,

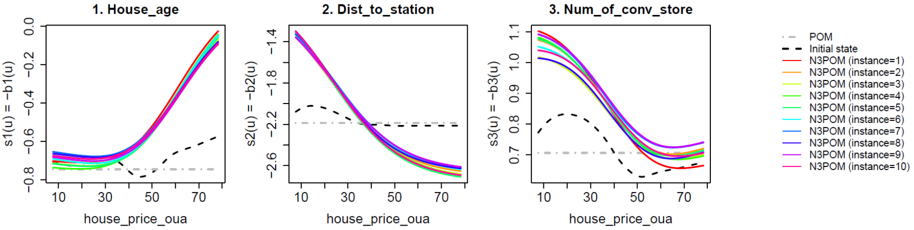
$$\mathbb{P}_{\mathrm{N^3POM}}(H \leq u \mid \mathbf{X} = \mathbf{x}) = \sigma(a(u) + \langle \mathbf{x}, \mathbf{b}(u) \rangle).$$

ただし$a(u)$は単調増加関数で$\mathbf{b}(u)$はニューラルネット

- 👍 微分値 ≥ 0 を解くことで単調性の十分条件が容易に導ける.
  - $(a(u)$の傾きの最小値$) \geq \eta \cdot$ $(\mathbf{b}(u)$の複雑度$)$
- 👍 $\mathbf{b}(u)$ を見れば共変量の応答への寄与が解釈できる.

動かしてみるといろいろ分かる

- Real-estate dataset.
- データは正規化済み.
- 築年数が大きい(=古い)と家の価格が上がる効果などが見つかった.
  - データを見てみると：古くて値段の高い屋敷の影響が捉えられていた.



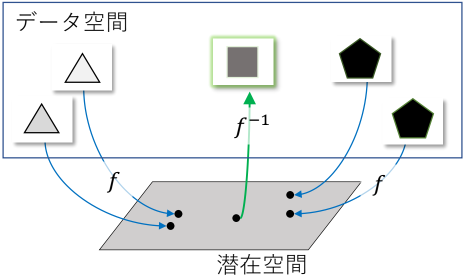
(3) 可逆関数推定のミニマックス解析

対象論文

- Akifumi Okuno and Masaaki Imaizumi. “Minimax Analysis for Inverse Risk in Nonparametric Planer Invertible Regression”, *Electronic Journal of Statistics*, 18(1):355-394, 2024.
  `https://doi.org/10.1214/23-EJS2202`
- 日本語での解説: Jxiv.616
- 生成モデルなどにも関係する，可逆関数推定の難しさを評価

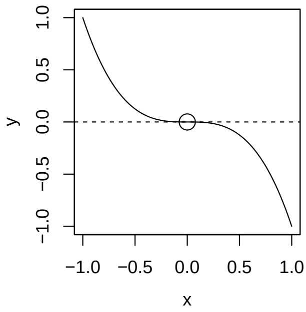
関数の可逆性は重要な制約

- **f** :データ空間→潜在空間は可逆とする.

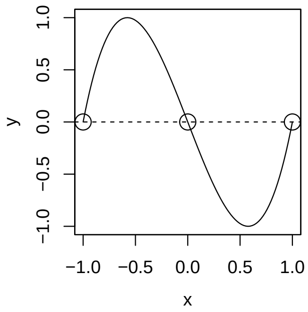


- $f^{-1}\left(\{f(\text{三角形})+f(\text{五角形})\}\right)\approx$ 四角形.
- 密度推定などでも出てくる.

一次元で連続な可逆関数 = 単調関数



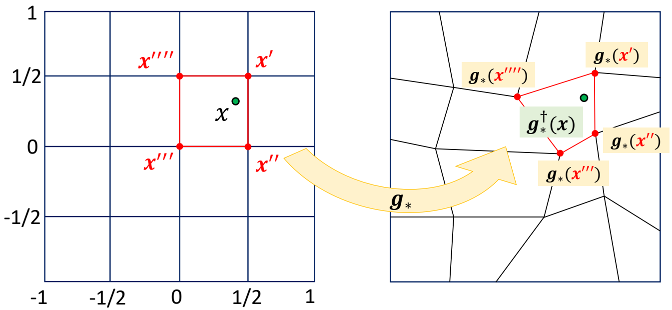
(a) 可逆



(b) 可逆でない

- 2次元以上の場合，可逆性を特徴付けるのが難しい.

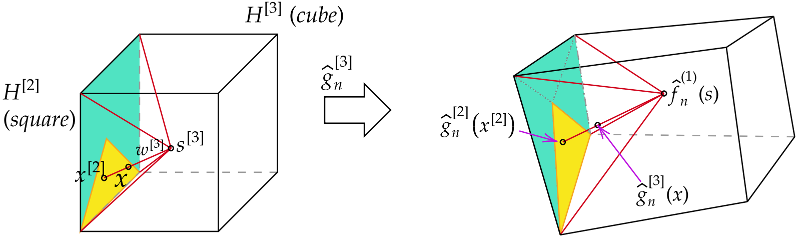
連続な可逆関数 ⇔ 同相写像



- 可逆関数 ⇔ 2つの空間のグリッドとグリッドを対応付けられる関数
- ミニマックスレートを導出した.
  - 可逆性はリプシッツ連続性より弱い制約だとわかった.

3次元以上の場合はどうなるのか?

- f_*に「リプシッツ連続性」しか課さなかったのが間違いだった.
  - 病的な例が出てくる. 特に$d = 4$次元で特異な例がある(らしい).
- f_*に「C^2級」を課す方針でどうにかできそう(ongoing).



まとめ

1. Akifumi Okuno and Keisuke Yano. "A generalization gap estimation for overparameterized models via the Langevin functional variance", *Journal of Computational and Graphical Statistics*, 32(4):1287-1295, 2023.

2. Akifumi Okuno and Kazuharu Harada. "An interpretable neural network-based non-proportional odds model for ordinal regression", *Journal of Computational and Graphical Statistics*, 2024. To appear.

3. Akifumi Okuno and Masaaki Imaizumi. "Minimax Analysis for Inverse Risk in Nonparametric Planer Invertible Regression", *Electronic Journal of Statistics*, 18(1):355-394, 2024.

質問やコメントは `okuno@ism.ac.jp` まで.