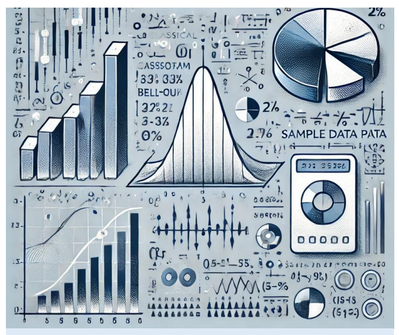
ニューラルネットと統計科学

奥野彰文[1,2,3]

[1]統計数理研究所　[2]総合研究大学院大学　[3]理化学研究所

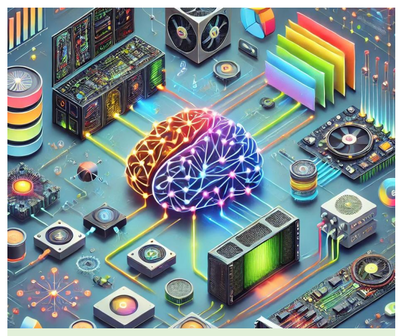
背景：統計科学と機械学習のターゲットの違い



統計科学

小サンプル
単純モデル
精密な理論保証

線形回帰モデル，
信頼度評価，漸近正規性，…



機械学習

ビッグデータ
巨大モデル
ラフな理論保証

大規模言語モデル (LLM)，
深層学習 (Deep Learning)，…

※単純化した統計科学・機械学習の分類です．例外もあります．

ニューラルネット（NN）は基本的に「ビッグデータ」「巨大モデル」での学習を想定するが，現実には学習データが十分に得られないことも多く，**統計科学との組み合わせも重要**である．

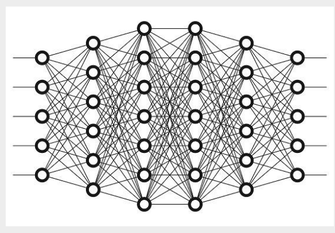
一方で既存の統計科学は線形性や正則性を仮定することが多く，**NNに適用するためには方法論の更新が必要**．特に，**表現能力の高い予測モデルをいかに小サンプルで学習するかは重要な課題**．

研究1．情報量規準を介したNNの効率的な汎化予測

Okuno* and Yano, JCGS2023, Jxiv:537

【解決したかった問題】
Cross-Validationは計算コストが高く，一方で理論的保証のつく赤池情報量規準（AIC）やWAICをはじめ，既存の情報量規準は予測誤差に正則性やパラメータの有限性を仮定することが多く，非正則・パラメータ過剰のNNへ直接適用できない．

【我々の貢献】
1．**特異モデルに対する情報量規準WAICは過剰パラメータモデルにも利用できることを示した．**
2．勾配のみ利用する，メモリ効率の良い計算法を提案した．



研究2．NNを利用した非比例順序回帰

Okuno* and Harada, JCGS2024, Jxiv:549

【解決したかった問題】
生存時間解析では主にハザード関数を推定するが，ここでは条件付きの累積確率分布関数 $\mathrm{logit}(P(H \le u \mid X = x))$ を線形回帰モデルで予測する問題を考える．回帰係数 β が u に依存するとき非比例順序回帰と呼び，$\beta(u)$を表現能力の高いNNでモデリングしたい．

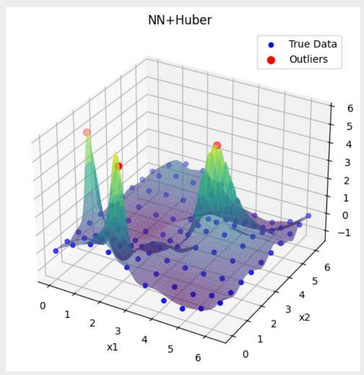
【我々の貢献】
1．予測された関数が常に（累積分布関数として）単調性を保つのは，回帰係数 $\beta(u)$ が定数の場合のみであることを初めて理論的に示した．つまり**制約なしでのNNの学習は破綻することを理論的に示した．**
2．共変量 x が有界範囲内にあるという制約を入れることで，$\beta(u)$ が定数関数でなく，単調性も保つ学習法を示した．

研究3．変動正則化を介したNNの外れ値ロバスト推定

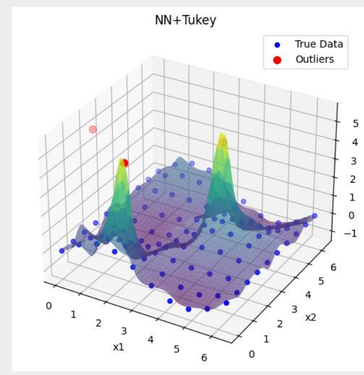
Okuno* and Yagishita, arXiv:2308.02293 in revision, Jxiv:928

【解決したかった問題】
データ中に含まれる外れ値や異常値に影響を受けないようNNを学習したい．統計学では，同様の問題設定で「ロバスト推定」が研究されてきたが，多くの既存法は予測モデルの自由度の低さに依拠しており，NNの学習にそのまま流用すると「外れ値を識別できない」問題に直面する．

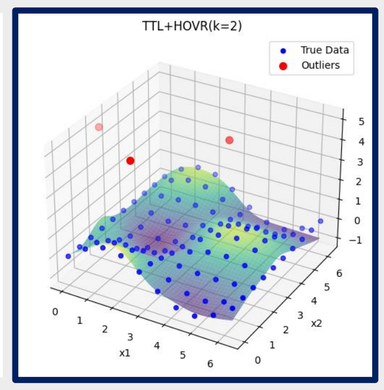
【我々の貢献】
1．NNの自由度を制御する「変動正則化」を提案した．
2．既存のロバストトリム損失とNNの変動正則化を組み合わせ，**ロバストにNNを学習できる効率的な最適化法を提案**した．
3．提案法はロバスト性の指標（Breakdown Point）が高いことを示した．



既存ロバスト法A ＋ NN



既存ロバスト法B ＋ NN



提案法 ＋ NN

研究4．計算代数を利用したNNの局所最適解全列挙

Fukasaku, Kabata, and **Okuno***, arXiv:2508.17783, Jxiv:2475

【解決したかった問題】
NNの学習において，損失関数は非常に複雑な非凸関数である．
勾配法などの数値的な解法により最適化を行うが，既存の数値的な方法では
（1）局所最適解はいくつぐらいあるのか
（2）到達した局所最適解は孤立点なのか（情報行列が退化しないのか），
（3）有限反復で打ちとめた場合，局所最適解にどのくらい近づけているか，
などを解明できない．

【我々の貢献】
1．回帰問題において，ReLU活性化を利用するパーセプトロン型NNを2乗損失で学習する場合，その損失関数は区分多項式になる．これを利用し，**計算代数によって停留点が満たすべき方程式の解を全列挙した．**
2．提案法を利用した実験により，特に，
A) 孤立型の局所最適解が活性化境界に集中していたこと，
B) リッジ正則化を入れていても，（境界を除いた内部領域に）1次元以上の連なった解空間を持つこと，
をそれぞれ観察できた．**これらの結果は，NNの不確実性予測等の統計科学理論研究にも将来的な応用が見込まれる．**

$$
\begin{aligned}
&c_1 - \frac{17b_{11}}{100} > 0, \quad c_2 - \frac{17b_{21}}{100} > 0, \quad \frac{11b_{11}}{25} + c_1 > 0, \quad \frac{11b_{21}}{25} + c_2 > 0,\\
&c_1 - B_{11} < 0, \quad c_2 - B_{21} < 0, \quad c_1 - \frac{2b_{11}}{5} < 0, \quad c_2 - \frac{2b_{21}}{5} < 0,\\
&c_1 - \frac{71b_{11}}{100} < 0 \quad c_2 - \frac{71b_{21}}{100} < 0,\\
0 &= b_{11} + R_1 c_1^7 + R_2 c_1^5 c_2^2 + R_3 c_1^5 + R_4 c_1^3 c_2^4 + R_5 c_1^3 c_2^2 + R_6 c_1^3 + R_7 c_1 c_2^6 \cdots - R_{10} c_1,\\
0 &= b_{21} + R_{11} c_1^6 c_2 + R_{12} c_1^4 c_2^3 + R_{13} c_1^4 c_2 + R_{14} c_1^2 c_2^5 + R_{15} c_1^2 c_2^3 + R_{16} c_1^2 c_2 \cdots - R_{20} c_2,\\
0 &= c_1^8 + 4c_1^6 c_2^2 + R_{21} c_1^6 + 6c_1^4 c_2^4 + R_{22} c_1^4 c_2^2 + R_{23} c_1^4 + 4c_1^2 c_2^6 + R_{24} c_1^2 c_2^4 \cdots - R_{30},
\end{aligned}
$$

ある問題に対して検出された連続した解空間の代数表示．
数値的な方法（例えば勾配法）ではこのような表現は得られない．

https://okuno.net/lab