FY2023/2023年度

High-Dimensional Causal Analysis Team

Masaaki Imaizumi

高次元因果解析チーム　今泉 允聡

RIKEN　AIP Center for Advanced Intelligence Project

チーム概要：

大自由度・高次元な
データ構造を数学的に理解

データ構造（因果構造など）
を解析する理論の構築

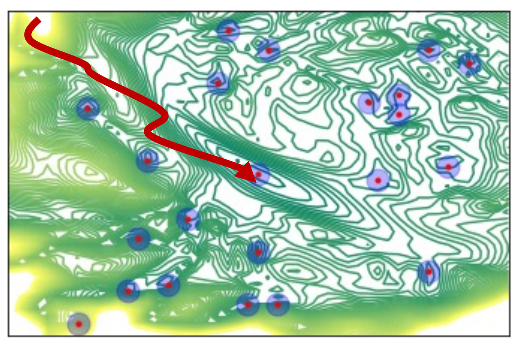
基礎的研究
- 高次元統計学
- 過剰パラメータ理論
- 深層学習の数理

応用的研究
- 因果推論
- 意志決定論
- 大規模データ解析

基礎的研究：高次元統計・深層モデルの解析

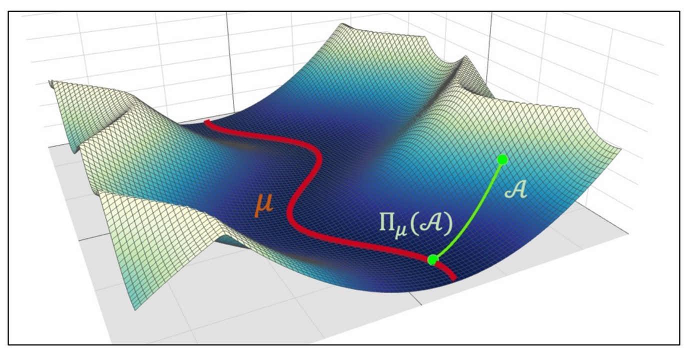
問題：深層ニューラルネットの予測精度の説明
　→パラメタ数が多くても過適合しない原理記述

アプローチ：エネルギー曲面を用いた汎化誤差解析
　→パラメータがエネルギー曲面の一部に滞留



エネルギー曲面 (等高線)

・局所解
●滞留集合
→学習過程



パラメータを制約する滞留集合と近傍の形状

結果：曲面上の滞留で過学習を防ぐ様相を記述

定理1 (滞留確率の保証)
減衰する学習率と、T回のパラメタ更新の元で：
$$p^* \geq 1 - c\Lambda^{-1}\exp(-T) - T\exp(-RT).$$
滞留点に制約される確率　学習が進む($T \to \infty$)と1に収束

定理2 (汎化誤差の評価)
確率 p^* 以上で以下が成立：
$$R(\hat{\theta}) - R_n(\hat{\theta}) = O\left(\frac{\sqrt{d\log K} + SLR\log^{0.5}\bar{h}}{\sqrt{n}}\right)$$
汎化誤差ギャップ　パラメータ数に依存しない誤差

n: データ数
θ: パラメタ
$\hat{\theta}$: 学習したパラメタ
$R(\theta)$: 汎化誤差
$R_n(\theta)$: 経験誤差
Λ, R: 滞留集合の体積/半径
d, K:滞留集合の次元と孤立領域の個数
$\bar{h}$: NNの最大幅
$S := \sup_{\theta\in\mathcal{B}} \prod_{\ell=1}^{L} \|A_\ell\|_S$: パラメタ行列$A_\ell$のスペクトル積

問題：構造入り高次元線形回帰
　パラメタ数が無限まで増加 + ノイズが非独立

X: 入力変数, Y: 出力変数, ε: ノイズ, $\theta \in \mathbb{R}^p$
$$Y = \langle X, \theta_* \rangle + \xi, \omega = E[X\xi] \neq 0, E[\xi|Z] = 0$$
Z: 潜在変数, Π: 変換行列, u: 構造ノイズ
$$X = \Pi Z + u$$

ノイズの相関構造を
潜在変数を用いて表現
$$\hat{L}(\theta) = \Sigma_{i=1}^n (Y_i - \langle X_i, \theta\rangle)^2$$
$$\hat{\theta} \in \mathrm{argmin}_\theta \hat{L}(\theta)$$

次元 $p \to \infty$, データ数 $n \to \infty$, $p/n \to \kappa \in (0, \infty)$

アプローチ：ガウス不等式による双対問題解析

評価尺度: 射影MSE $E\left[\left(E\left[\langle\hat{\theta}, X\rangle - \langle\theta_*, X\rangle \middle| Z\right]\right)^2\right]$
訓練データへの完全適合下での一様上界
$$\leq \sup_{\theta:\hat{L}(\theta)=0} E[(E[\langle\theta, X\rangle - \langle\theta_*, X\rangle|Z])^2]$$

高次元空間
$\hat{\theta}$
補間量の集合 $\{\theta: \hat{L}(\theta) = 0\}$

凸ガウスミニマックス定理（CGMT）
W: ガウスランダム行列, g, h: ガウスベクトル
$$\min_a \min_b \langle a, Wb\rangle + \psi(a, b) \approx \min_a \min_b \|a\|\langle g, b\rangle + \|b\|\langle h, a\rangle + \psi(a, b)$$

結果：一致推定可能性を証明

ω: 相関パラメタ
$\Xi := \Pi E[ZZ^\top]\Pi^\top$: 潜在共分散行列
$\Sigma_u := E[uu^\top]$: 構造ノイズ共分散

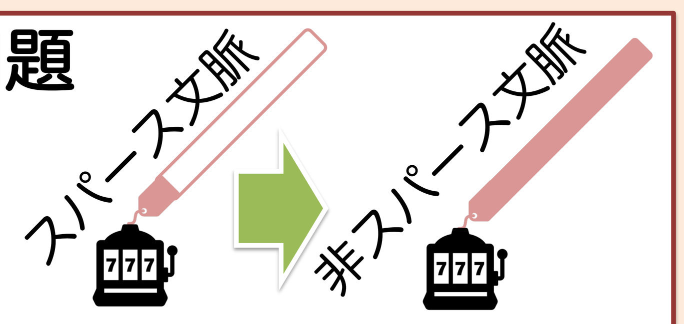
定理1 (誤差収束の上界)
Xの分布制約と直交条件のもとで
$$\|\hat{\theta} - \theta_*\|_\Xi^2 = O\left(\psi(\|\Sigma_u^+\omega\| + \|\theta_*\|)\sqrt{\frac{\mathrm{tr}(\Xi)}{n}}\right)$$

高次元・非独立ノイズでも
固有ベクトルの構造のもとで
一致推定が可能に

応用的研究：機械学習の諸問題への貢献

問題：高次元文脈バンディット問題

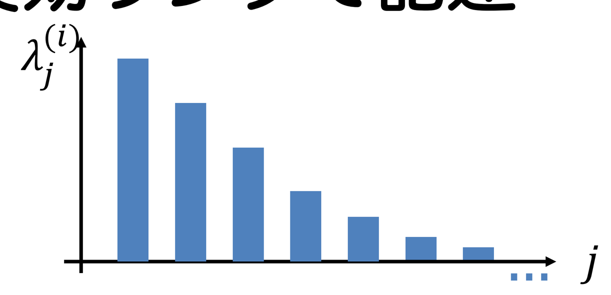
$X^{(i)}(t)$: 腕iの文脈ベクトル（時刻t）
（平均ゼロ、共分散 $\Sigma^{(i)} = E\left[X^{(i)}(X^{(i)})^\top\right]$）
選択$I(t)$の報酬($\theta^{(k)} \in \mathbb{R}^p$: 次元 $p \to \infty$)
$$Y^{I(t)}(t) = \langle X^{I(t)}(t), \theta^{I(t)}\rangle + \xi(t)$$



非スパース文脈に拡張

アプローチ：共分散$\Sigma^{(i)}$の構造を実効ランクで記述

数学的ツール：実効ランク
$\lambda_j^{(i)}$: $\Sigma^{(i)}$のj番目に大きい固有値
$$r_k(\Sigma^{(i)}) := \frac{\Sigma_{j>k}\lambda_j^{(i)}}{\lambda_{k+1}^{(i)}}, R_k(\Sigma^{(i)}) = \frac{\left(\Sigma_{j>k}\lambda_j^{(i)}\right)^2}{\Sigma_{j>k}\left(\lambda_j^{(i)}\right)^2}$$



高次元で有効なランク概念
→戦略の誤差記述に有効

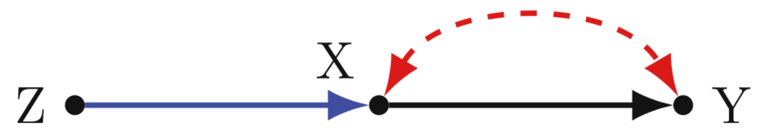
結果：高次元設定で有効戦略の開発とその誤差保証

Algorithm 1 Explore-then-Commit (EtC)
Require: Exploration duration T_0.
for $t = 1, .., T_0$ **do**
　Observe $X^{(i)}(t)$ for all $i \in [K]$.
　Choose $I(t) = t - K\lfloor t/K \rfloor$.
　Receive a reward $Y^{(I(t))}(t)$.
end for
for $i \in [K]$ **do**
　$\widehat{\theta}^{(i)} \leftarrow (\mathbf{X}^{(i)})^\top(\mathbf{X}^{(i)}(\mathbf{X}^{(i)})^\top)^{-1}\mathbf{Y}^{(i)}$.
end for
for $t = T_0 + 1, \dots, T$ **do**
　Observe $X^{(i)}(t)$ for all $i \in [K]$.
　Choose $I(t) = \mathrm{argmax}_{i\in[K]}\langle X^{(i)}(t), \widehat{\theta}^{(i)}\rangle$.
　Receive a reward $Y^{(I(t))}(t)$.
end for

定理 (期待リグレット評価)　$\alpha > 0$
Σがある共分散構造を持つとき、T回の試行のもとで
$$R(T) = \tilde{O}(T^\alpha K^{2/3}) \text{ かつ } R(T) = \tilde{\Omega}(T^\alpha)$$
期待リグレット上界　期待リグレット下界

問題：非線形な操作変数回帰モデル

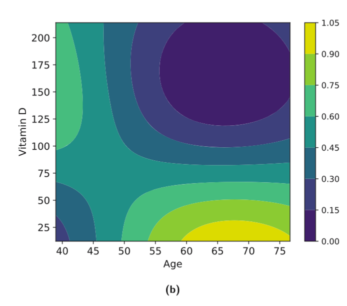
X: 入力変数, Y: 出力変数, ε: ノイズ
$$Y = f(X) + \varepsilon$$
Z: 操作変数, ν: 構造ノイズ
$$X = t(Z) + g(\varepsilon) + \nu$$



入力変数とノイズが相関
→操作変数を用いて対処

アプローチ：カーネル関数によるU統計量変換

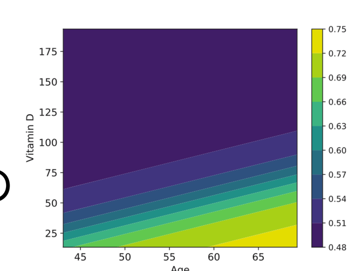
最大モーメント制約による損失関数（$\mathcal{H}_k$: 再生核ヒルベルト空間）
$$R_k(f) = \sup_{f\in\mathcal{H}_k}\left(E[(Y - f(X))h(Z)]\right)^2$$
再生性を用いた書き換え（(Y', X', Z'): iid copy）
$$R_k(f) = \sup_{f\in\mathcal{H}_k} E[(Y - f(X))(Y' - f(X'))k(Z, Z')]$$
U統計量で推定→経験平均



結果：予測精度と解釈性（検定）の両立

定理 (漸近分布への収束)
$$\sqrt{n}(\hat{f} - f_0^*) \to_d \nabla S(\mathbb{G}^*)$$
推定誤差　ガウス過程の線形変換

漸近分布の導出→統計的検定を開発



年齢と
栄養価の
死亡率への
影響

非線形性による精度向上

Selected Papers

M.Imaizumi, J.Schmidt-Hieber (2023), "On Generalization Bounds for Deep Networks Based on Loss Surface Implicit Regularization", IEEE Transaction on Information Theory, 69(2).
T.Tsuda, M.Imaizumi (2024+), "Benign Overfitting of Non-Sparse High-Dimensional Linear Regression with Correlated Noise", arXiv, R&R.
R.Zhang, M.Imaizumi, B.Schölkopf, K.Muandet (2023), "Instrumental Variable Regression via Kernel Maximum Moment Loss". Journal of Causal Inference, 11(1).
J.Komiyama, M.Imaizumi (2023), "High-dimensional Contextual Bandit Problem without Sparsity", Advances in Neural Information Processing Systems.
A. Okuno, M.Imaizumi (2024), "Minimax Analysis for Inverse Risk in Nonparametric Planer Invertible Regression", Electronic Journal of Statistics.