複雑な予測モデルの正則化

奥野彰文[1,2]

[1]統計数理研究所　[2]理化学研究所

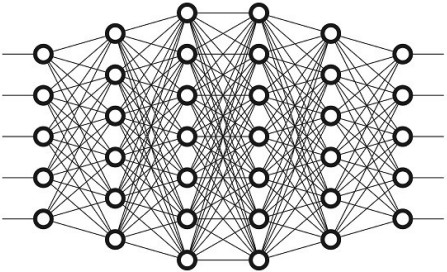
複雑な予測モデルの正則化

ニューラルネット

- 特にMulti-layer perceptron:

$$f_\theta(x) = (g^{(L)} \circ \sigma \circ g^{(L-1)} \circ \sigma \circ \cdots \sigma \circ g^{(1)})(x),$$

ただし$g^{(\ell)}(z) = A^{(\ell)}z + b^{(\ell)} : \mathbb{R}^{d^{(\ell-1)}} \to \mathbb{R}^{d^{(\ell)}}$は $A^{(\ell)} \in \mathbb{R}^{d^{(\ell)} \times d^{(\ell-1)}}$, $b^{(\ell)} \in \mathbb{R}^{d^{(\ell)}}$をパラメータに持つ線形モデル， σは活性化関数.



ニューラルネットのここがすごい

- 万能近似定理：任意の連続関数f_*について，$\inf_\theta \|f_* - f_\theta\|_\infty \approx 0$.
- 実装が簡単.

```python
# NN Definition
class NeuralNetwork(nn.Module):
    def __init__(self, input_dim=2, activation_func='sigmoid'):
        super(NeuralNetwork, self).__init__()
        self.first = nn.Linear(input_dim, 100)
        self.hidden1 = nn.Linear(100, 100)
        self.hidden2 = nn.Linear(100, 100)
        self.hidden3 = nn.Linear(100, 100)
        self.output = nn.Linear(100, 1)
        self.activation = nn.Sigmoid() if activation_func == 'sigmoid' else nn.ReLU()

    def forward(self, x):
        x = self.activation(self.first(x))
        x = self.activation(self.hidden1(x))
        x = self.activation(self.hidden2(x))
        x = self.activation(self.hidden3(x))
        return self.output(x)
```

図 1: PyTorchでの実装例.

私とニューラルネットの出会い

- （2012年度=学部3年）ビショップ本の輪読で5章（ニューラルネット）を担当.
- 当時気の利いたパッケージはなかった（のでC++フルスクラッチ）

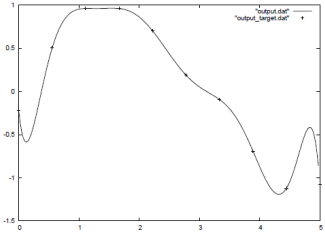


図 3 $f(x) = \sin x, (a, b, N, M, p, \lambda) = (0, 5, 10, 10, 0, 0)$

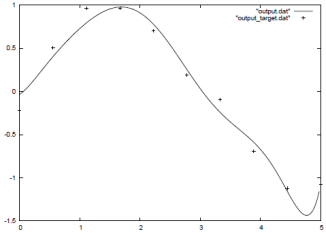


図 4 $f(x) = \sin x, (a, b, N, M, p, \lambda) = (0, 5, 10, 10, 2, 0.5)$

図 2: 当時の輪読で書いてみたコードによる当てはめ.

- 意外と動かないと思った記憶が…（0で初期化して詰むなど）

統計学から機械学習へ

- 下平先生のところで卒論生をやることに.
- 情報量規準，スパース推定，相関分析，等々やったがどれも論文にできず.
- そしてニューラルネットへ…

表現学習で一発あてにいった

- 正準相関分析+グラフ = Shimodaira (2016)
- 正準相関分析+グラフ+ニューラルネット = Okuno et al. (ICML2018)

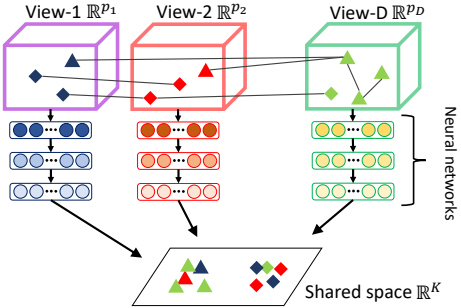


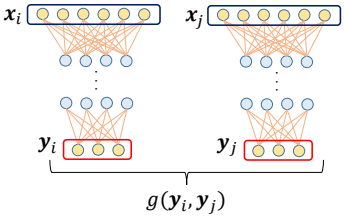
図 3: 線形モデルを…非線形ニューラルネットに…しましょう…

グラフ埋め込みの確率モデル

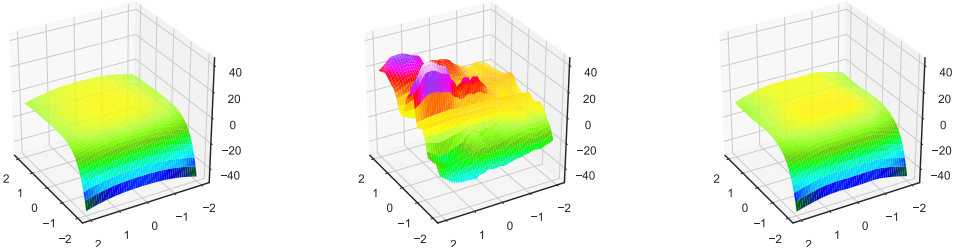
- データx_i, リンク$w_{ij} \in \{0, 1\}$, ニューラルネットf_θ.
  グラフ埋め込みの確率モデル：

$$w_{ij} \sim \mathrm{Bernoulli}(\mu_{ij}), \quad \mu_{ij} = \sigma(\underbrace{\langle f_\theta(x_i), f_\theta(x_j)\rangle}_{\mathrm{IPS}}).$$

- ロジスティック回帰の拡張になっている.
- Siamese型NNはカーネル法同様に正定値性の制限を受ける.



グラフ埋め込みの表現能力?



(a) 条件付き正定値　(b) 既存モデル(IPS)　(c) 提案モデル(SIPS)

図 4: 正定値カーネル以外も近似できる埋め込みモデルへ

- 次数の重みを入れることで，表現能力向上? (Okuno et al. AISTATS2019)
- このあたりの話で博論.

色々な形のニューラルネットが作られた

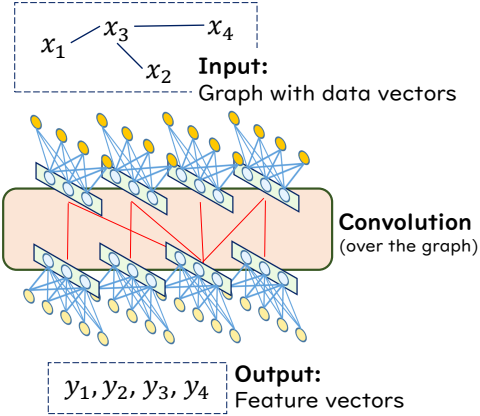
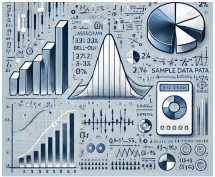


図 5: グラフを入力にもつGraph Neural Networkなど

素朴な気持ちに戻ってくると

- 巨大モデルに巨大データを投入して学習（頑張る）．よくわからない．
- 個人的にあんまり興味がなくなってきた．

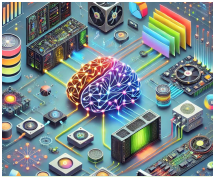


統計学

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，…

実は溝がある



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)，…

図 6: 最近よく使っている気持ちの図

巨大なものにばかり目がいきがちですが・・・

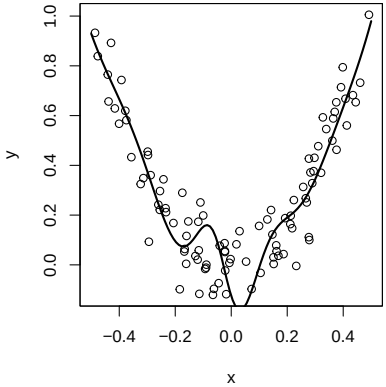
- 基本的には大サンプルでの気合の入った学習が念頭にある.
- 統計学っぽい小サンプルの話にもNNは有用では？
  カーネル法を含めた，基底関数モデル：

$$f_\theta(x) = \theta_0 + \sum_j \theta_j \underbrace{\phi_j(x)}_{\text{基底}}$$

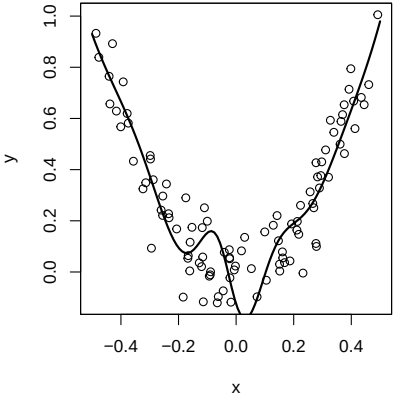
- 基底はどう定めるのか? $\Rightarrow$ NNは基底まで含めた学習になる.

Q. NNを小サンプルで学習するとどうなるか?

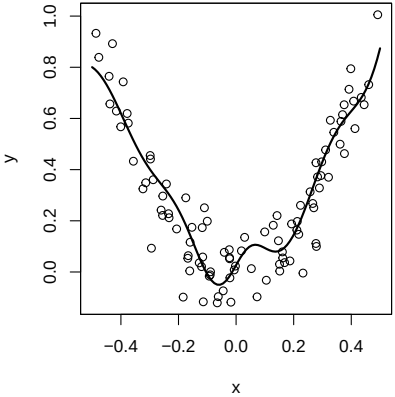
- A. グニャグニャになる.



- 単純なパラメータ正則化は実はあんまり効果なし.



(a) 正則化なし



(b) リッジ正則化 $\|\theta\|_2^2$

「変動の正則化」

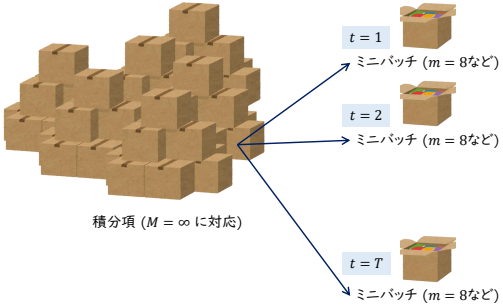
- グニャグニャを明示的に抑えればよい：はるか古のスプライン…?
- Higher-Order Variation Regularization (HOVR)

$$C(f_\theta) = \int \left\| \frac{\partial^k f_\theta(x)}{\partial x^k} \right\|_2^2 \mathrm{d}x$$

- HOVR込みでの正則化を考えればよろしい，が…
  - 基底関数モデルでは：$C(f_\theta) \asymp \|\theta\|_2^2$.
  - ニューラルネットでは：$C(f_\theta)$が明示的に求まらない.

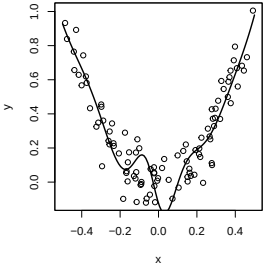
確率的勾配法

$$\|\mathbf{y} - f_\theta(\mathbf{x})\|_2^2 + \lambda C(f_\theta) = \frac{1}{n}\sum_{i=1}^{n}\{y_i - f_\theta(x_i)\}^2 + \lambda \cdot \operatorname*{p\,lim}_{M\to\infty} \frac{1}{M}\sum_{j=1}^{M}\left|\frac{\partial^k f_\theta(z_j)}{\partial z_j^k}\right|^2$$

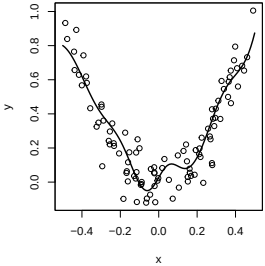


- 勾配をランダム化すれば積分型の項も最適化できる (Robbins and Monro, 1951)

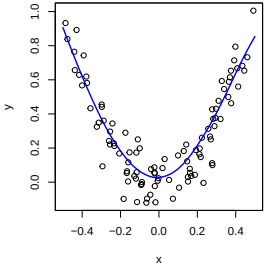
そして滑らかな曲線へ



(a) 正則化なし



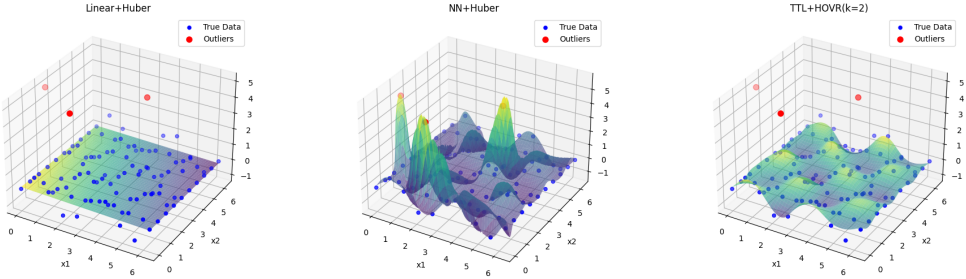
(b) リッジ正則化 $\|\theta\|_2^2$



(c) HOVR $C(f_\theta)$

- 好きなニューラルネットで使える．スプラインより実装が容易！
- NNで標準的な，自動微分とミニバッチ最適化しか使わない．
- カーネル法や一般化線形モデル，etc.. 何にでも使える．

変動正則化は外れ値への感度を抑える

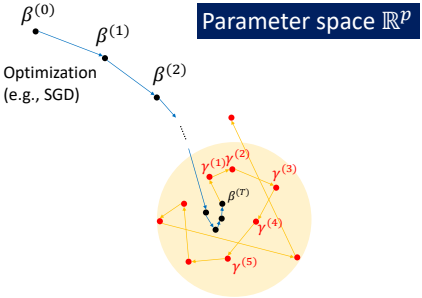


(a) 線形+Huber　(b) NN+Huber　(c) NN+Trim+HOVR

- 表現能力が高いが，過剰にならない予測モデルの学習.
- Okuno and Yagishita (arXiv:2308.02293, submitted)

ハイパーパラメータの選択

- Cross-validationを回せばよい，が…
- やっぱり情報量規準が使いたい（統数研的願望）.
- Widely-Applicable Information Criterion (WAIC) が実は効率的に計算できる…！



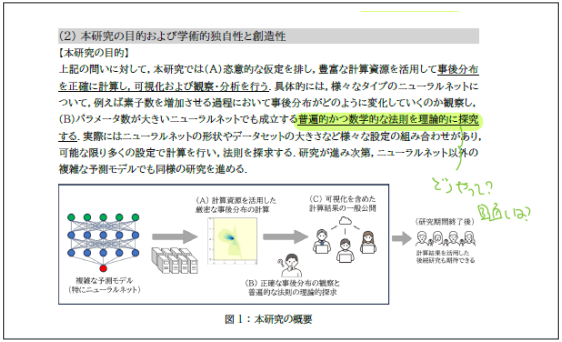
Posterior ≈ Langevin Distribution

- Okuno and Yano (JCGS2023)

これからの研究

研究をご一緒したことはありませんでしたが

- 科研費の申請書添削で大変お世話になりました.
  - 若手 (2020年) ⇒ 川崎先生（入所時の思考院長=上司),
  - 基盤B (2024年) ⇒ 栗木先生，二宮先生，日野先生.



(2) 本研究の目的および学術的独自性と創造性

【本研究の目的】

上記の問いに対して，本研究では(A)恣意的な仮定を排し，豊富な計算資源を活用して事後分布を正確に計算し，可視化および観察・分析を行う．具体的には，様々なタイプのニューラルネットについて，例えば素子数を増加させる過程において事後分布がどのように変化していくのか観察し，(B)パラメータ数が大きいニューラルネットでも成立する普遍的かつ数学的な法則を理論的に探究する．実際にはニューラルネットの形状やデータセットの大きさなど様々な設定の組み合わせがあり，可能な限り多くの設定で計算を行い，法則を探求する．研究が進み次第，ニューラルネット以外の複雑な予測モデルでも同様の研究を進める．

どうやって？ 見通しは？

図 1：本研究の概要

図 14: 添削の後でほぼ全文書き直しました…

2025年度から基盤Bのプロジェクトを始めます

- 代表：奥野（統計的機械学習），
- 分担：深作（計算代数），高畠（確率過程+統計）.

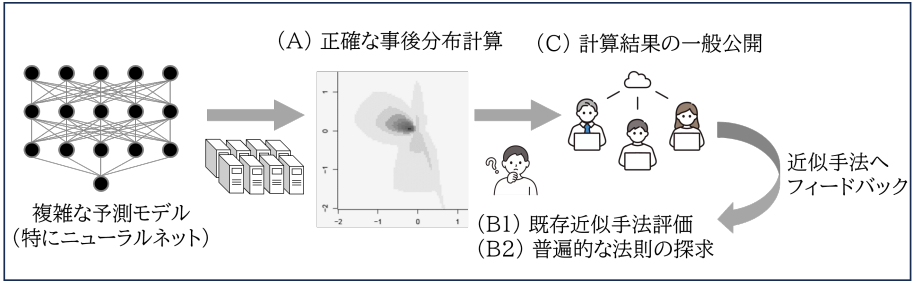


図 15: 課題名「複雑な予測モデルの厳密な事後分布計算手法の確立とその応用」

やることはざっくり：

- 複雑な予測モデル（ニューラル，時系列，...）の事後分布の精密計算.
- 計算代数で頑張って計算する（という話を含んでいる）.

```
f[x_, a_, b_, c_, d_] = a + b * ReLU[c * x + d];
Solve[Sum[(y[i] - f[x[i], a, b, c, d])^2, {i, 1, n}] == Min, {a, b, c, d}]
```

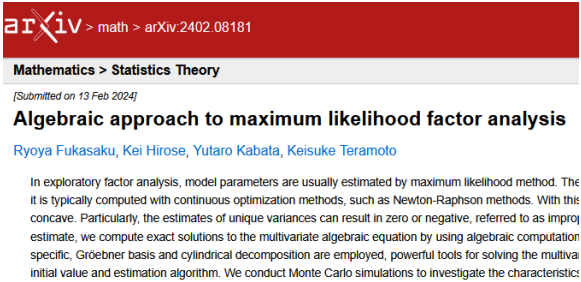
- 深作さんがすべてを計算してくれる.
- 私は少し違う計算をする…

いろいろなコミュニティとの交流があった

- 深作さん（九大），加葉田さん（長崎大），…
- 横山先生（立教大），神戸さん（三菱電機），グレブナー基底大好きbotさん，…
- 横山先生（都立大），学生の方々…

図 16: グレブナー基底を高速に計算する計良さん@思考院セミナー

代数統計に最近興味を持っている



arXiv > math > arXiv:2402.08181

Mathematics > Statistics Theory

[Submitted on 13 Feb 2024]

Algebraic approach to maximum likelihood factor analysis

Ryoya Fukasaku, Kei Hirose, Yutaro Kabata, Keisuke Teramoto

In exploratory factor analysis, model parameters are usually estimated by maximum likelihood method. The
it is typically computed with continuous optimization methods, such as Newton-Raphson methods. With this
concave. Particularly, the estimates of unique variances can result in zero or negative, referred to as improp
estimate, we compute exact solutions to the multivariate algebraic equation by using algebraic computation
specific, Gröebner basis and cylindrical decomposition are employed, powerful tools for solving the multivar
initial value and estimation algorithm. We conduct Monte Carlo simulations to investigate the characteristics

図 17: arXiv:2402.08181

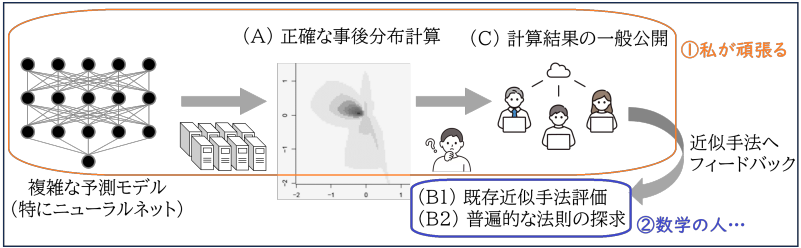
- 計算と数学の中間地点っぽさが面白い.
- 計算代数 = 制限が結構ある（が素人目にはよく分からない）.

代数統計には興味がある，が敷居が高い…

- 竹村先生，日比先生，栗木先生，原先生，青木先生，…
- 一介の統計学者が参入するのはハードルが高い．
- 栗木先生が最終講義で言及されていた論文： Drton, Kuriki, and Hoff (AoS2021), Existence and uniqueness of the Kronecker covariance MLE.

- よく聞いてみると面白いことが書いてあったり．
  - 計算代数コミュニティの人たちと仲良くやっていく…??
  - 鉄則として，数学力では戦わない…!

数学はやっぱり難しいので

- 数学ができる人と，数学で競ってはならない
  ⇒ 色々な経験的事実を数学者に提供すればよいのでは…

- 数学的に必要な「仮定」から外れた実験をやるプロジェクトにしたい．
- 数学に強い人はいつでもウェルカムです．