

神経回路網と幾何学の接点

福水 健次

情報・システム研究機構 統計数理研究所
総合研究大学院大学

RIMS共同研究: 離散力学系の分子細胞生物学への応用数理
「神経システムと代数幾何」パネルディスカッション
2009.1.5 – 1.9 京都大学

1

自己紹介

大学: 京都大学理学部数学教室 (ここ)

1989.- (株)リコー 中央研究所

1998.- 理化学研究所 脳科学総合研究センター

2000.- 統計数理研究所 現職
統計的学習、機械学習の研究

2

- ・ 自己紹介: 京大理数→1989 (株) リコー→1998 理研 BSI→2000 統数研で統計的学習・機械学習の研究
- ・ 脳も代数幾何もやっていない!

Outline

Singularities in neural network models

Geometry in neural networks, but not necessarily algebraic geometry.

Algebraic geometric approach to probabilistic networks

Algebraic geometry, but not necessarily neural networks.

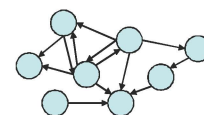
3

- ・ Singularity in neural network model (geometry in neural networks, but not necessarily algebraic geometry)
- ・ 神経回路モデルの特異点…数理モデルに抽象化ニューラルネットワーク
- ・ Algebraic geometric approach to probabilistic network (algebraic geometry, but not necessarily neural network)
- ・ 確率推論に現れるネットワークにおける代数幾何

Multilayer Perceptron

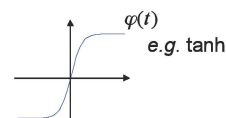
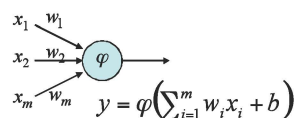
Mathematical model of neural computation

Brain as an information processing system



Network of processing units

Single neuron model

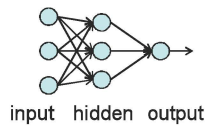


4

- ・ Mathematical model of neural computation
- ・ 脳の単純化とは、network of processing units
- ・ 脳のシナプスの複雑な連結→抽象化→ニューラルネットワーク (各 unit で、入力の線形和を引数とした非線形関数の値を出力)
- ・ single neuron model では、出力 $y = \varphi$ (e.g. tanh) ($\sum w_i$ (重み付け、興奮性ならば正、抑制性ならば負) x_i (入力) $+b$)

Multilayer Perceptron

Multilayer perceptron: feedforward network



$$f(x; \theta) = \sum_{j=1}^H b_j \sigma(a_j^T x + c_j) + d$$

a_j, b_j, c_j, d : parameters to learn (synaptic learning)

– It is *not* a precise model of biological neural networks, but it may capture some properties of neural computation:

- Network structure
- Local, parallel computation
- Simple computation by units of the same function
- Existence of hidden units

5

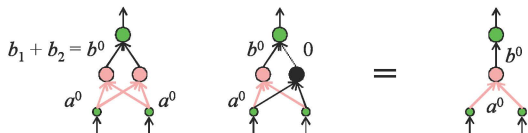
- 多層パーセプトロン (feedforward network: 前向きの信号の流のみを考える) は、入力 → hidden → 出力
- $f(x; \theta) = \sum b_j \sigma$ (e.g. tanh) $(a_j^T x + c_j$ (しきい値) $) + d$
- a_j, b_j, c_j, d が学ぶべきパラメータ

Symmetry and Singularity

Model: $f(x; \theta) = b_1 \varphi(a_1 x) + b_2 \varphi(a_2 x)$

Target: $f_0(x) = b^0 \varphi(a^0 x)$

Redundant representation



The parameter set $\{a_1 = a_2 = a^0, b_1 + b_2 = b^0\}$
 $\cup \{a_1 = a^0, b_1 = b^0, b_2 = 0, a_2 : \text{arbitrary}\}$ realizes $f_0(x)$.
 ∅ Unidentifiability

– c.f. Linearly parameterized model

The parameter is unique even if the model is surplus.

ex.) Polynomial $a_0 + a_1 x + \dots + a_H x^H$

6

- Symmetry and singularity
- 中間素子が2個あるネットワークを考えてみる
- 生物では実現すべき関数より冗長な素子があることが多い
- 冗長なモデルを使うと、パラメータの識別不能性
- parameter space における識別不能性 → function space における特異点

• target: $f_0(x) = b_0 d(a_0 x)$

• model: $f(x; \theta) = b_1 d(a_1 x) + b_2 d(a_2 x)$

• パラメータの識別不能性が現れる

こんな感じのネットワーク

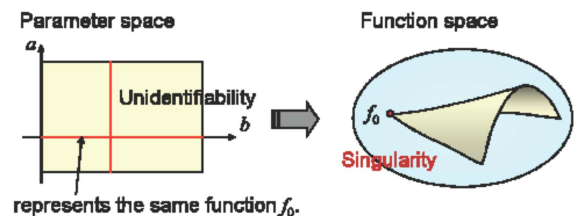
○ → ● → ⊙ →

○ → ● ↑

- 多項式では識別不能性は起こらない
- ノイズがあるときは冗長性により機能が向上するのではないかな？

Symmetry and Singularity

Singularity caused by unidentifiability

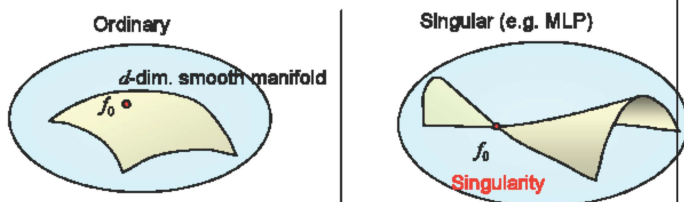


7

- Singularity caused by unidentifiability
- パラメータ空間 → 関数空間
- 特異点に対応

Smooth Manifold v.s. Singular Model

Parametric model in function space



Interesting behaviors around singularity

- Statistical behavior of the estimator
- Trajectory of the parameter in learning process
- The singularity may be locally infinite dimensional
- More geometric approach? Algebraic geometry helps?

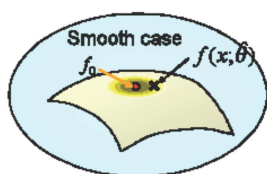
8

- Smooth manifold vs. singular model
- 普通の Smooth manifold: d 次元、スムーズな manifold
- Singular (e.g. MLP): 特異点が生じると幾何的構造が複雑になる
- 学習の誤差、学習の trajectory が singularity を通らないと学習できない場合がある
- Singularity が局所的に無限次元かもしれない

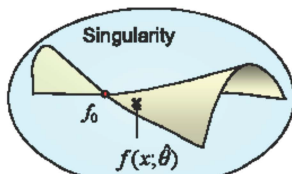
Statistical Estimation

$\hat{\theta}$: optimal parameter by learning. $\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N (Y_i - f(X_i; \theta))^2$

$\hat{\theta}$ is a random variable.



Asymptotically normal
information geometry



??? algebraic geometry helps?

Reference: 「特異モデルの統計学」
榎水, 栗木, 竹内, 赤平 (岩波書店)

Bayes estimation & integral. Approach by algebraic geometry (Watanabe)

9

- 統計的推定
- 多層パーセプトロンは Hebb 則的学習
- 普通は Optimal parameter by learning の 2 乗誤差を少なくする
- ベイズ推定では integral. approach by algebraic

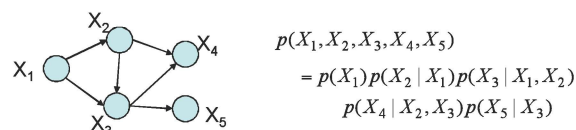
geometry

- プラトール現象とは、学習の誤差の減り方に停滞期が存在する→singularity がある程度かかわっているかもしれない

Bayesian Network

Simpler and more abstract network

Probabilistic network with finite states



Bayesian network / graphical model
Used for

- modeling of data
- expressing causal relations

10

- Bayesian network = Simpler and more abstract network。より簡単で、それぞれの素子が有限状態だけ取る。条件付き確率を表す ($p(X_1, X_2, X_3) = p(X_1)p(X_2 | X_1)p(X_3 | X_1, X_2)$)

Algebraic Statistics

Network with two nodes of binary states

General



$$p_{ij} = p(X=i, Y=j) \quad i, j \in \{0,1\}$$

$$\Delta = \{p = (p_{00}, p_{01}, p_{10}, p_{11}) \in \mathbb{R}^4 \mid p_{ij} > 0 (\forall i, j), \sum_{i,j \in \{0,1\}} p_{ij} = 1\}$$

Independent



$$\Gamma = \{p \in \Delta \mid p_{ij} = P(X=i)P(Y=j)\}$$

$$\Gamma = \{(\alpha_0 \beta_0, \alpha_0 \beta_1, \alpha_1 \beta_0, \alpha_1 \beta_1) \in \mathbb{R}^4 \mid \alpha_i \in \mathbb{R}, \beta_j \in \mathbb{R}\} \cap \Delta$$

$$\frac{\alpha_i}{\alpha_0 + \alpha_1} = P(X=i), \quad \frac{\beta_j}{\beta_0 + \beta_1} = P(Y=j)$$

Toric variety with toric ideal $\langle p_{00}p_{11} - p_{01}p_{10} \rangle$

11

- Algebraic statistics、2 値の値を取る ($p_{00}, p_{01}, p_{10}, p_{11}$)
- Toric variety with toric ideal
- 一般に有限状態のグラフィカルモデルが toric variety に対応

Algebraic Statistics

- In general, a graphical model with finite states corresponds to a toric variety.
- Algebraic geometric methods are applied to statistical problems.
✂ Algebraic statistics.
- Any link with neuro science?

References:

- Pistone, Riccomango and Wynn. *Algebraic Statistics*. 2001.
- Pachter and Sturmfels. *Algebraic Statistics and Computational Biology*. 2005.
- Geiger, Meek and Sturmfels. On the toric algebra of graphical models. *Annals of Statistics* 34 (2006) etc.

12

ディスカッション

- 多層パーセプトロンをつなげていけば愛が生まれるか？→精度は良くなる
- 同じ層内で関係を持たせる、または時間遅れの関係性を入れればどうなるか？本質的に性能が変わるか？
- 出力を入力にすることの数学的な意味は？
- ニューラルネットを複雑にして文字認識に使う
- 簡単にフィードバックがある場合、任意のネットワークでは難しい
- Back propagation
- 問題の複雑さとニューラルネットワークの複雑さの関係→ある関数のクラスを考えて、関数を近似する場合→ユニットの数と、近似誤差のワースト値の関係は分かっている
- 中間素子の数、worst case の近似性能の関係、素子を増やす近似精度が上がる
- 特異点があった方が良いのか？ベイズ推論では特異点の方が得=冗長な方が良い。ニューラルネットでは特異点が必ず出る。事前確率を置くと特異点で

得をする

186

- 加藤：入来先生の「わかるとは何か」 学習=状態の集合をカテゴライズ、部分集合をどれだけ認識するか 分け方をどれだけよくするか
- 全ての現象を理解することはできない
- 多義化 = 集合が無限に広がっている
- 概念の構成 = 集合との対応化
- 入来：わかるとは 階層 論文では「上の階層の原因になっている あるいは下の階層が上に影響している」として解釈する
- 階層が無限に続く
- ニューラルネット = パラメータ最適化学習
- わかり方の問題、わかる対象の問題
- 可塑性のないシナプスの方が多い
- 愛とか正義の連合野はほとんど活動しない
- 多層パーセプトロンは問いに対して常に正しい答えを出すのか？
- 中間素子のパターンに対する発火状況のエントロピーを用いて学習状態を判断する試みもあった。ただし、それで良いのか？せいぜい数十個のニューロン。検証すべき
- 中間素子の状態が情報の representation ではないか？
- 特異点をはさまれたものの状態？サイエンスとして検証しうるか
- 中間レイヤーをモニタリングすることは可能
- 脳内にはあるか
- パーセプトロンは小脳に似ている、だから有名になった
- 全てのニューロンは可塑性がない。時期によって異なる
- 中間層が何かと結合していない方が難しい
- 近似の他のメジャー 確率的な予測がどれだけ当たるか
- 連続関数の近似
- 多層になったときはもっと違うメジャーが必要
- 制限にして学習をしても結果は良くない 決まった定式化はない
- ランダムにつながったネットワークで自発的に階層ができるか？
- 作れるか？作り込むことは可能。自然に発生する

か？

- ・振動子として素子を考える。階層性やクラスターができる
- ・大腦のように乖離した構造があると機能ができる
- ・本能はどこにあるのか？本能＝原始的な脳。鳥ではわかっている
- ・自己抑制ができる人は大腦が発達している
- ・どれだけ複雑なことができるのか？やっている研究は？

・文字列認識（手書きの郵便番号の認識）…ニューラルネットワークのパラメータをヒューリスティックに調節したものの方が性能がよかった

・音声認識（時間構造が入る→構造が増えると難しくなる→あまりニューラルネットワークを使われることがなくなってきた）

・ループがある時のパラメータ調整は？

簡単なFBがあるような系では、研究がされている
任意のネットワークについてはよくわからない

・自発的に階層を作るようなネットワークはあるのだろうか？

⇒作りこむことは可能だともわれる

同期に応じて結合が強化されるようなネットワークは存在している

ある種のクラスターである

constraint が無い状態で、自発的に階層を作るとしたら、大腦の構造も納得できる