2002 年情報論的学習理論ワークショップ
2002 Workshop on Information-Based Induction Sciences (IBIS2002)
Fujiyoshida, Japan, Sep 8 - 10, 2002.

縮小ランク回帰モデルの学習曲線の解析

Learning Curves of Reduced Rank Regression in Bayesian Estimation

渡辺 一帆* 渡辺 澄夫†
Kazuho Watanabe Sumio Watanabe

Abstract: The reduced rank regression is a statistical model which estimates the conditional probability using a reduced rank linear operator. It is a kind of non-identifiable models such as layered neural networks and gaussian mixtures, therefore the learning theory of regular statistical models can not be applied. Recently, it has been known that the pole of the zeta function of the Kullback information has a mathematical relation with the learning efficiency in Bayesian estimation. In this paper, based on this theorem, we prove an upper bound of the learning curves of the reduced rank regression in Bayesian estimation. Also we show the usefulness of an algebraic geometrical method such as blowing-up for analyzing the learning machines with singularities.

1 はじめに

神経回路網や混合正規分布などの階層構造を持つ学習モデルが，パターン認識，時系列予測，システム制御などの様々な実問題に利用され，その有用性について実験的に確認されつつある．しかしながら，これらの階層構造を持つ学習モデルにおいては，パラメータと学習モデルとの対応が 1 対 1 ではなく，学習モデルのふるまいからパラメータが一意的には定まらない．このため，サンプル数が十分に大きい場合でも統計的正則モデルの統計的漸近理論を適用することができず，学習モデルの性質を解明することや最適な設計を行うための数学的基盤が未だに十分には確立されていない．この問題は，神経回路網を含む非常に広い学習モデルにおいて一般的に生じている問題であることが認識されるようになった [2]．

本論文では，特定不能な学習モデルである縮小ランク回帰モデルを考察し，その学習曲線を解析する．縮小ランク回帰モデルは，高次元入力から高次元出力への線形推論を小さなランクを持つ線形写像の中から見出すものであり，情報科学，生物学・医学，線形システムの制御，人文社会科学などにおいて多変量間の関係を見いだす方法のひとつとして広く応用されている．また，3 層パーセプトロンにおいて中間ユニットの応答が線形である場合に相当し，神経回路網の性質を理論的に調べる観点からの研究もなされている [4]．このモデルでは，入力から出力への写像は線形写像であるが，対数尤度はパラメータについて 4 次の多項式になり，原点に複雑な特異点を持つために，その学習は線形理論の枠組みでは取り扱うことができない．実際，最尤法を用いて予測を行うと，汎化誤差は（比例定数/学習例数）となるが，ここで比例定数は「パラメータ数/2」よりも大きな値になることが知られている [5]．

本論文では，縮小ランク回帰モデルの学習にベイズ推測を適用する場合を考え，ブローアップによって特異点を変換することにより，その推定精度の上限を求め，比例定数が「パラメータ数/2」よりもどの程度に小さいかを数理的に明らかにする．このことは縮小ランク回帰モデルを用いて予測や制御を行う場合に，最尤推定法を用いるよりもベイズ法を用いる方が適している可能性が高いことを示すものである．

2 縮小ランク回帰とベイズ推測

2.1 縮小ランク回帰モデル

縮小ランク回帰モデルとは，入力 $\mathbf{x} \in R^M$ から出力 $\mathbf{y} \in R^N$ への線形写像の中からもっとも相応しいランクの写像を学習によって見出すものである．行列 A および B を

$$A = \left(a_{ij}\right) \quad (1 \leq i \leq H, 1 \leq j \leq M)$$

*東京工業大学大学院総合理工学研究科電子機能システム専攻, 226-8503 神奈川県横浜市緑区長津田 4259, tel. 045-924-5017, e-mail kazuho23@pi.titech.ac.jp,
Tokyo Institute of Technology, 4259 Nagatuta, Midori-ku, Yokohama, 226-8503, Japan
†東京工業大学 精密工学研究所, e-mail swatanab@pi.titech.ac.jp
P&I Lab., Tokyo Institute of Technology, 4259 Nagatuta, Midori-ku, Yokohama, 226-8503, Japan

$$B \quad = \quad \Big(b_{ij}\Big) \quad (1 \leq i \leq N, 1 \leq j \leq H)$$

とすると，入力 $\mathbf{x}$ が与えられたときの縮小ランク回帰モデルの出力 $\mathbf{y}$ は，

$$\mathbf{y} = BA\mathbf{x} + (\text{雑音})$$

と表される．以下，A，B は縮小ランク回帰モデルのパラメータであり，これを $w = (A, B)$ と書く．特に雑音として，平均 0，分散共分散行列が $\sigma^2 I (I$ は単位行列$)$ の正規分布を考えると，入力 $\mathbf{x}$ が与えられたときの，出力 $\mathbf{y}$ の分布 $p(\mathbf{y}|\mathbf{x}, w)$ は，

$$p(\mathbf{y}|\mathbf{x}, w) = \frac{1}{\sqrt{2\pi\sigma^2}^N} \exp(-\frac{1}{2\sigma^2}\|\mathbf{y} - BA\mathbf{x}\|^2) \qquad (1)$$

である．ここで $\|\cdot\|$ は R^N のユークリッドノルムを表す．入力 $\mathbf{x}$ が与えられたときの $\mathbf{y}$ の真の分布が中間ユニット数 H_0 のモデルで表されると仮定する．行列 A_0，B_0 を，

$$A_0 \quad = \quad \Big(a^*_{ij}\Big) \quad (1 \leq i \leq H_0, 1 \leq j \leq M)$$
$$B_0 \quad = \quad \Big(b^*_{ij}\Big) \quad (1 \leq i \leq N, 1 \leq j \leq H_0)$$

とすると，真の分布は，

$$q(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp(-\frac{\|\mathbf{y} - B_0A_0\mathbf{x}\|^2}{2\sigma^2}) \qquad (2)$$

と表される．縮小ランク回帰モデルは，高次元の情報から高次元への線形推論が，実質は低次元の情報により定まると考えられる問題に適用される．このとき，真の分布の中間ユニット数 H_0 は不明であることが多く，そのため，様々な中間ユニット数のモデルを学習させて比較することになる．真の分布に最も近いと思われるモデルを選んだり，予測誤差を最小にするモデルを選んだりする場合に，特に，学習モデルが真の分布よりも冗長な場合に生ずる現象を解明しておくことが必要になる．

2.2 ベイズ推測と汎化誤差

本論文では，ベイズ推測による学習を解析する．入力 $\mathbf{x}$ が密度関数 $q(\mathbf{x})$ を持つと仮定し，$(\mathbf{x}, \mathbf{y})$ の同時確率密度関数

$$q(\mathbf{x}, \mathbf{y}) = q(\mathbf{x})q(\mathbf{y}|\mathbf{x})$$

から，独立な n 個のサンプル

$$D = \{(\mathbf{x}_i, \mathbf{y}_i); i = 1, 2, ..., n\}$$

が得られたとする．パラメータ空間に事前分布 $\varphi(w)$ を設定する．このとき，ベイズ事後分布は

$$p(w|D) \propto \varphi(w) \prod_{i=1}^{n} p(\mathbf{y}_i|\mathbf{x}_i)$$

と定義される．またベイズ予測分布は

$$p(\mathbf{y}|\mathbf{x}, D) = \int p(\mathbf{y}|\mathbf{x}, w)p(w|D)dw$$

である．ベイズ汎化誤差 $G(n)$ は，真の分布からベイズ予測分布までのカルバック情報量をサンプル D の出方について平均することによって定義され，学習曲線と呼ばれる．

$$G(n) = E\Big[\int q(\mathbf{x}, \mathbf{y}) \log \frac{q(\mathbf{y}|\mathbf{x})}{p(\mathbf{y}|\mathbf{x}, D)} d\mathbf{x}d\mathbf{y}\Big]. \qquad (3)$$

ここで E はサンプル D の現れ方に関する平均を表す．

学習理論の課題のひとつは，学習曲線 $G(n)$ を解明し，その結果に基づいて，より小さな学習曲線を持つ学習法を構成することである．縮小ランク回帰モデルのように特異点を持つ学習モデルのベイズ汎化誤差については，次のような解決が知られている [9]．

真の分布からパラメータ w を持つ学習モデルまでのカルバック情報量 $K(w)$ を

$$K(w) = \int q(\mathbf{x}, \mathbf{y}) \log \frac{q(\mathbf{y}|\mathbf{x})}{p(\mathbf{y}|\mathbf{x}, w)} d\mathbf{x}d\mathbf{y} \qquad (4)$$

と定義する．カルバック情報量と事前分布 $\varphi(w)$ により定まるゼータ関数を

$$J(z) = \int K(w)^z \varphi(w)dw$$

と定めると，この関数は $Re(z) > 0$ では複素関数として正則であるが，複素平面全体まで有理型関数として解析接続することができ，その極はすべて負の実軸上にある有理数である [3]．その極の中で，最も原点に近い極を $-\lambda$ とし，その位数を m とすると，学習曲線 $G(n)$ が漸近展開できるならば

$$G(n) = \frac{\lambda}{n} - \frac{m-1}{n \log n} + o(\frac{1}{n \log n})$$

が成り立つ [9]．ここで，ゼータ関数の原点に一番近い極は，代数多様体 $\{w; K(w) = 0\}$ の特異点を解消することにより見出すことができる．任意の代数多様体はブローアップの有限回の繰り返しにより，正規交差特異点だけを持つものに変換できることが知られている（広中の定理）[6] が，特異点を解消する写像を具体的に構成することは，一般には，それほど容易ではない．しかしながら，ブローアップによってゼータ関数の極を見出すことは比較的容易であり，極が見出されると，その値は自動的に学習曲線の上限を与えるものになる．

本論文では，以上の方法に基づいて，ブローアップを適用することにより，縮小ランク回帰モデルのベイズ推測における学習曲線の上限値を求め，その値が正則な統計モデルよりも小さいことを証明する．

3　主定理

入力と事前分布について，次の仮定をおく．

条件 1 入力の分布 $q(\mathbf{x})$ について

$$\int x_k^2 q(\mathbf{x}) d\mathbf{x} < \infty \quad (1 \leq k \leq M)$$

が成り立つ．

条件 2 事前分布 $\varphi(w)$ は，A，B の各成分についてのルベーグ測度の直積に対する確率密度関数として定義され，そのサポートはコンパクトであり，

$$\begin{aligned} L &\equiv \max_{ij} |a_{ij}^*| + \max_{ij} |b_{ij}^*| \\ W &\equiv \{w \in R^d; |w_i| \leq 2L\} \end{aligned}$$

で定義される W の上で $\varphi(w) > 0$ であるとする．

以上の条件のもとで，本論文では次の定理を証明する．

定理 1 縮小ランク回帰モデルのベイズ汎化誤差の係数 λ は次の不等式を満たす．
(1) $M = N$ かつ $H - H_0$ が偶数のとき

$$\lambda \leq \frac{1}{4}\{(2M-1)H + (2M - 2H_0 + 1)H_0\}.$$

(2) $M = N$ かつ $H - H_0$ が奇数のとき

$$\lambda \leq \frac{1}{4}\{(2M-1)H + (2M - 2H_0 + 1)H_0 + 1\}.$$

(3) $M \neq N$ のとき

$$\lambda \leq \frac{1}{2}\{\min(M,N)H + \max(M,N)H_0 - H_0^2\}.$$

縮小ランク回帰モデルは，形式的には $H(M+N)$ 個のパラメータを持っているが，補題 2 でも示すように，任意の逆行列を持つ $H \times H$ 行列 C を用いて

$$\begin{aligned} A &\mapsto CA \\ B &\mapsto BC^{-1} \end{aligned}$$

とパラメータを変換しても推論は変化しない．すなわち，縮小ランク回帰モデルのパラメータ数は，この自由度を減算した $H(M+N) - H^2$ であると考えることができる．上記の定理は，ベイズ推測を用いると，λ がこの (パラメータ数/2) よりも，汎化誤差が小さくなることを述べている．

4　準備

4.1　ゼータ関数の性質

本論文では，ベイズ推測を考察するため，カルバック情報量と事前分布により定まるゼータ関数の性質が必要になる．一般に w の多項式 $f(w)$ とサポートがコンパクトな関数 $g(w) \geq 0$ とが与えられたとき，それらのゼータ関数は

$$\zeta(z) = \int f(w)^z g(w) dw \tag{5}$$

と定義される．ここで $g(w)$ は確率密度関数でなくてもよい．この関数は一変数 z の複素関数として有理型に複素平面全体に解析接続でき，その極は負の有理数になることが知られている．この関数の最も原点に近い極を $-\Lambda(f,g)$ と書く．このとき，次の基本的な関係が成り立つ [12]．
(1) $f_1(w) \leq f_2(w)$ かつ　$g_1(w) \geq g_2(w)$ ならば

$$\Lambda(f_1, g_1) \leq \Lambda(f_2, g_2).$$

(2) $f(w_1, w_2) = f_1(w_1) + f_2(w_2)$，$g(w_1, w_2) = g_1(w_1)g_2(w_2)$ ならば

$$\Lambda(f, g) = \Lambda(f_1, g_1) + \Lambda(f_2, g_2).$$

(3) w が d 次元であり，$f(w)$ が解析関数で，$g(w) > 0 (\forall w)$ が成り立つとき

$$\Lambda(f, g) \leq \frac{d}{2}.$$

特に，(1) から，式 (5) の積分を行う範囲を一部分に制限すると，対応する λ は制限する前の値以上の値になる．

4.2　補題とその証明

定義 1 R^M から R^N への線形写像を表す行列 $L = (L_{ij})$ （実数値）が与えられたとき，

$$\|L\| = \Big[\sum_{i=1}^{N}\sum_{j=1}^{M} L_{ij}^2\Big]^{1/2}$$

と定義する（異なる M，N の行列についても同様に定義する）．$\|\cdot\|$ は作用素としてのノルムとは一致しないが，ノルムの公理を満たす．特に，任意の L_1，L_2 について

$$\|L_1 + L_2\|^2 \leq 2(\|L_1\|^2 + \|L_2\|^2) \tag{6}$$

が成り立つ．

補題 1 式 *(4)* で定義されるカルバック情報量 $K(w)$ は次の不等式を満たす．ある $c_0 > 0$ が存在して，

$$K(w) \leq c_0 \|BA - B_0 A_0\|^2.$$

(補題 1 の証明概略) カルバック情報量 $K(w)$ の $\mathbf{y}$ に関する積分を行い，条件 1 を考えればよい．

まず，一般的に次の不等式が成り立つ．

補題 2 $f(w) \equiv \|BA - B_0A_0\|^2$ とおくと，次の不等式が成り立つ．

$$\Lambda(f, \varphi) \leq \frac{1}{2}(N + M - H)H. \tag{7}$$

この不等式は真の推論のランクに依存しない．

(補題 2 の証明概略) 変数変換により $f(w)$ は $MH + NH - H^2$ 個のパラメータにのみ依存するようにできることと，ゼータ関数の性質 (3) から導かれる．

行列 A の最初の H_0 行までが作る行列（A_0 と同じ型の行列） を A_1 とし，残りの行が作る行列を A_2 とする．行列 B の最初の H_0 列までが作る行列（B_0 と同じ型の行列）を B_1 とし，残りの列が作る行列を B_2 とする．$w_1 = (A_1, B_1)$, $w_2 = (A_2, B_2)$ とおいて

$$\begin{aligned} f_1(w_1) &\equiv \|B_1A_1 - B_0A_0\|^2 \\ f_2(w_2) &\equiv \|B_2A_2\|^2 \end{aligned}$$

と定義すると， 式 (6) から，

$$f(w) \leq 2f_1(w_1) + 2f_2(w_2) \tag{8}$$

が成り立つ．また

$$\begin{aligned} \varphi_1(w_1) &= \min_{\|w_2\| \leq 1} \varphi(w_1, w_2) \\ \varphi_2(w_2) &= \begin{cases} 1 & (\|w_2\| \leq 1) \\ 0 & (\text{otherwise}) \end{cases} \end{aligned}$$

と定義すると

$$\varphi(w) \geq \varphi_1(w_1)\varphi_2(w_2)$$

が成り立つので，

$$\begin{aligned} \lambda &= \Lambda(f, \varphi) \\ \mu &= \Lambda(f_1, \varphi_1) \\ \nu &= \Lambda(f_2, \varphi_2) \end{aligned}$$

と定義すると，

$$\lambda \leq \mu + \nu \tag{9}$$

が成り立つ．そこで，以下では，μ，ν のそれぞれの値の上限を求める．μ は，真の推論と学習モデルの大きさが一致する場合に相当し，補題 2 より

$$\mu \leq \frac{H_0(M + N - H_0)}{2}$$

が成り立つ．ν は，真の推論が $y = 0$ の場合に相当する．

定義 2 パラメータの集合 C を

$$C = \{w \in R^{(M+N)H}; \|w_i\| \leq 1 \ \ ({}^{\forall}i)\}$$

と定義する．

$$J(z) = \int_C \|BA\|^{2z} dw$$

の最も原点に近い極を $\lambda(H)$ と定義する．

特に，式 (9) における ν は，$\nu = \lambda(H - H_0)$ である．

補題 3 任意の $1 \leq k \leq H - 1$ について

$$\lambda(H) \leq \lambda(H - k) + \lambda(k) \tag{10}$$

が成り立つ．

(補題 3 の証明) また式 (6) より，式 (8) を導くのとまったく同じ方法で示すことができる．(補題 3 の証明終).

補題 4 次の不等式が成り立つ．

$$\lambda(H) \leq \frac{\min(M, N)H}{2}.$$

(補題 4 の証明) 定義から

$$\|BA\|^2 = \sum_{i=1}^{N} \sum_{k=1}^{M} \{\sum_{j=1}^{H} b_{ij}a_{jk}\}^2$$

次の変数変換 $(a, b, w') \mapsto w$ を用いる．

$$\begin{aligned} a_{11} &= a, \\ a_{jk} &= aa'_{jk} \quad ((j, k) \neq (1, 1)) \\ b_{11} &= b, \\ b_{ij} &= bb'_{ij} \quad ((i, j) \neq (1, 1)). \end{aligned}$$

定数倍を省略し，積分範囲を (a, b, w') の全ての変数について $[-1, 1]$ に限ったゼータ関数は

$$\begin{aligned} J(z) &= \int a^{2z}b^{2z}K_1(w')^z a^{MH-1}b^{NH-1} da db dw' \\ &= \frac{1}{2z + MH}\frac{1}{2z + NH}\int K_1(w')^z dw' \end{aligned}$$

となる．ここで

$$K_1(w') = \sum_{i=1}^{N}\sum_{k=1}^{M}\{\sum_{j=1}^{H} b'_{ij}a'_{jk}\}^2$$

とおいた (ただし $a'_{11} = b'_{11} = 1$)．この $J(z)$ は $z = -\frac{MH}{2}, -\frac{NH}{2}$ を極として持つ．このことから，補題が得られた．(補題 4 の証明終)．

補題 5 次の不等式が成り立つ．

$$\lambda(2) \leq \frac{M+N-1}{2}$$

(補題 5 の証明) 考察するゼータ関数は

$$J(z) = \int \{\sum_{i=1}^{N}\sum_{k=1}^{M}(b_{i1}a_{1k} + b_{i2}a_{2k})^2\}^z dw$$

である．

$$\begin{aligned} b_{i2} &= b'_{i2} \\ b_{i1} &= b'_{i2}b'_{i1}, \\ a_{1k} &= a'_{1k} \\ a_{2k} &= a'_{1k}a'_{2k} \end{aligned}$$

の変数変換を考え，

$$f(w') = \sum_{i=1}^{N}\sum_{k=1}^{M} b'^2_{i2}a'^2_{1k}(b'_{i1} + a'_{2k})^2 \qquad (11)$$

とすると，

$$J(z) = \int f(w')^z \prod_{i=1}^{N}|b'_{i2}| \prod_{j=1}^{M}|a'_{1j}|\, dw' \qquad (12)$$

そこで

$$b'_{i1} + a'_{21} = p_i \quad (1 \leq i \leq N)$$
$$b'_{11} + a'_{2k} = q_k \quad (2 \leq k \leq M)$$

とおいて，b'_{i1} のかわりに p_i を，a'_{2k} のかわりに q_k を用いると

$$\begin{aligned} f(w') &= \sum_{i=1}^{N} a'^2_{11}b'^2_{i2}p_i^2 + \sum_{k=2}^{M} a'^2_{1k}b'^2_{12}q_k^2 \\ &\quad + \sum_{i=2}^{N}\sum_{k=2}^{M} a'^2_{1k}b'^2_{i2}(p_i + q_k - p_1)^2 \end{aligned}$$

となる．さらに

$$\begin{aligned} p_1 &= p, \\ p_i &= pp'_i \ (2 \leq i \leq N) \\ q_k &= pq'_k \ (2 \leq k \leq M) \end{aligned}$$

の変数変換を考えると，

ゼータ関数 (12) は，

$$\begin{aligned} J_0(z) &= \int p^{2z+M+N-2}dp \\ &= \frac{1}{2z+M+N-1} \end{aligned}$$

を因数として含む．以上より補題 5 が得られた．(補題 5 の証明終)．

5 主定理の証明

(定理の証明) 式 (9) より，λ の上限は，μ と ν の上限の和で与えられる．μ は，この補題 (2) において $H = H_0$ とおいたときの不等式を満たすから，

$$\mu \leq \frac{1}{2}(N + M - H_0)H_0. \qquad (13)$$

が成り立つ．

次に ν を考える．式 (10) をくりかえして用いることにより

$$\lambda(H) \leq \begin{cases} \frac{H}{2}\lambda(2) & (H \text{ が偶数のとき}) \\ \frac{H-1}{2}\lambda(2) + \lambda(1) & (H \text{ が奇数のとき}) \end{cases} \qquad (14)$$

が得られる．$\lambda(1)$ は補題 4 において $H = 1$ の場合を考えて，

$$\lambda(1) \leq \frac{\min(M,N)}{2} \qquad (15)$$

従って，補題 (5) と (14)，(15) より，

$$\lambda(H) \leq \begin{cases} \frac{M+N-1}{2}\frac{H}{2} \\ \quad (H \text{ が偶数のとき}) \\ \frac{M+N-1}{2}\frac{H-1}{2} + \frac{\min(M,N)}{2} \\ \quad (H \text{ が奇数のとき}) \end{cases} \qquad (16)$$

が成り立つ．

$$\begin{cases} \frac{M+N-1}{2} < \min(M,N) & (M = N \text{ のとき}) \\ \frac{M+N-1}{2} \geq \min(M,N) & (M \neq N \text{ のとき}) \end{cases}$$

であるから，定数 ν は H を $H - H_0$ にすることで，

$$\nu \leq \begin{cases} \frac{2M-1}{2}\frac{H-H_0}{2} \\ \quad (M = N \text{ かつ } H - H_0 \text{ が偶数}) \\ \frac{2M-1}{2}\frac{H-H_0}{2} + \frac{1}{4} \\ \quad (M = N \text{ かつ } H - H_0 \text{ が奇数}) \\ \min(M,N)\frac{H-H_0}{2} \\ \quad (\text{それ以外}) \end{cases} \qquad (17)$$

を満たす．μ，ν の上限が得られたので，式 (9)，(13)，(17) より，定理の不等式が得られる．(定理の証明終)．

6　考察

6.1　特異点の様子とゼータ関数

縮小ランク回帰モデルの最も簡単なケースとして，学習モデルの中間ユニット数が 1 で真の分布が $\mathrm{y} = 0$ である場合には，カルバック情報量 $K(w)$ は，ある定数 $c_1, c_2 > 0$ が存在して

$$\begin{aligned} c_1 f(w) &\leq K(w) \leq c_2 f(w) \\ f(w) &= \{\sum_{i=1}^{N} b_{i1}^2\}\{\sum_{j=1}^{M} a_{1j}^2\} \end{aligned}$$

を満たす．すなわち，代数多様体 $\{w; f(w) = 0\}$ は $a_{1j} = 0$ または $b_{i1} = 0$ によって定義される．このとき，ゼータ関数

$$J(z) = \int f(w)^z dw$$

の極は，$a_{1j} = 0$ の近くから得られる $z = -M/2$ と $b_{i1} = 0$ の近くから得られる $z = -N/2$ の二つになり，M，N の小さい値が，$J(z)$ のより原点近い極を与えることになる．事後分布は，

$$p(w) \propto \exp(-nK(w))$$

であるが，上記のことは，この事後分布において，M，N の小さい方のパラメータは原点に集中し，そうでない方のパラメータは, 大局的な広がりを持っていることと対応する．従って，M，N の大きさの大小に依存して, 事後確率に従うパラメータの広がり方が変化する．一般のケースでは，特異点の様子は複雑になるが，特異点を解消した空間では同様の状況が生じているために結果に場合わけが必要になる．

6.2　事前分布の選択

本論文では，縮小ランク回帰モデルについて，行列 A, B の両方の各成分についてのルベーグ測度を規準として正値を取る密度関数を用いた．神経回路網などのモデルでは，ジェフリーズの事前分布を用いると，正則な統計モデルの時と同等のモデル選択規準が得られることが知られている [10][12]．縮小ランク回帰モデルでは，パラメータを A，B の成分に選ぶと，フィッシャー情報行列の行列式は常に 0 なるので，ジェフリーズの事前分布は存在しない．ジェフリーズの事前分布を定義できるようにパラメータの制限を行い，その場合の汎化誤差を解析するのは，今後に残された問題である．また，縮小ランク回帰モデルにおいて, モデル選択によって，適切なランクのモデルを選ぶことは，中心的な課題のひとつである．神経回路網のモデル選択において，確率的複雑さの最小化による方法は既に研究されている [10][7]．今後, 本論文の結果を基盤として，縮小ランク回帰モデルの選択について検討を行いたい．

7　結論

代数幾何的な手法を用いて，縮小ランク回帰モデルのベイズ汎化誤差の上限を与え，証明を行った．また具体的な例に適用することによって，カルバック情報量のゼータ関数の極を求めるためにブローアップを用いる方法の有効性と汎用性を明らかにした．

本研究は部分的に科学研究費補助金 12680370 の援助を受けた．

参考文献

[1] Amari, S., Murata, N. “Statistical theory of learning curves under entropic loss criterion,” *Neural Computation*, Vol.5, pp.140-153, 1993.

[2] 甘利俊一, 尾関智子, 朴慧暎, “階層モデルにおける学習と推論ー特異構造を持つ統計モデル,” 電子情報通信学会誌 VOL.J85-DII No.5,pp.701-708,2002.

[3] M.F.Atiyah, “Resolution of Singularities and Division of Distributions,” *Communications of Pure and Applied Mathematics* , Vol.13, pp.145-150,1970.

[4] P.F.Baldi, K. Hornik,“Learning in linear neural networks: a survey,” *IEEE Trans. on Neural Networks*,Vol.6,pp.837-858, 1995.

[5] K. Fukumizu, “Generalization error of linear neural networks in unidentifiable cases,” *Lecture Notes on Computer Science*, Vol.1720, 51-62, 1999.

[6] H.Hironaka, H. “Resolution of singularities of an algebraic variety over a field of characteristic zero,” *Annals of Mathematics*, Vol.79, pp.109-326,1964.

[7] 西上功一郎, 渡辺澄夫,“特異モデルのモデル選択における最良予測と知識獲得,” 信学技報, NC2001-151, pp.143-150, 2002.

[8] 渡辺一帆, 渡辺澄夫,“ランク縮小写像のベイズ汎化誤差の解析,” 信学技報, NC2001-149, pp.127-134, 2002.

[9] S.Watanabe,“Algebraic analysis for nonidentifiable learning machines,” *Neural Computation*, Vol.13, No.4, pp.899-933,2001.

[10] S.Watanabe,“Algebraic information geometry for learning machines with singularitiess,” *Advances in Neural Information Processing Systems*, Vol.13, pp.329-336. 2001.

[11] S. Watanabe, ”Algebraic geometrical methods for hierarchical learning machines,” Neural Networks, Vol.14, No.8,pp.1049-1060, 2001.

[12] 渡辺澄夫, “特異点を持つ学習モデルと事前分布の代数幾何,” 人工知能学会誌, Vol.1 6，No.2, 2001.