2006 年情報論的学習理論ワークショップ
2006 Workshop on Information-Based Induction Sciences (IBIS2006)
Osaka, Japan, October 31 - November 2, 2006.

ベイズ法における三層学習モデルの確率的複雑さ
Stochastic Complexities of Three Layered Learning Model in Bayesian Estimation

青柳　美輝*　　渡辺　澄夫†
Miki Aoyagi　　Sumio Watanabe

Abstract:
In this paper, we obtain the main terms of stochastic complexities for some three layered learning models in Bayesian estimation. In [1, 2], it was proved that the main term of the stochastic complexity for the hierarchical learning model is given by the maximum pole of the zeta function, which is defined by the integral of its Kullback function and an *a priori* probability density function. In this paper, our main purpose is to obtain the maximum pole and its order of the following function:

$$\int \Psi \quad = \quad \int (\sum_{n=1}^{H}\sum_{k=1}^{M}\sum_{j=1}^{N}(\sum_{i=1}^{H} a_{ki} b_{ij}^{p(n-1)+1})^2)^z \prod_{k=1}^{M}\prod_{i=1}^{H'} \mathrm{d}a_{ki} \prod_{i=1}^{H'}\prod_{j=1}^{N} \mathrm{d}b_{ij},$$

where M, N, H, H', p are natural numbers with $H' \le H \le 2H'$ and a_{ki}, b_{ij} $(i > H')$ are all constants. We call a learning model whose zeta function is $\int \Psi$, "a three layered learning model" in this paper. We use a recursive blowing up process for analyzing Ψ.

1　まえがき

階層型ニューラルネットワーク，縮小ランク回帰モデル，混合正規分布などは，パラメータが特定不能になり，フィッシャー計量が縮退するという特徴をもつ．このような階層構造，内部構造を持つものは，特異モデルとよばれており，パラメータを変えることできわめて多くの確率分布となりえることが証明されている．これらのモデルは，たとえば，モデル選択法 AIC[3], TIC[4], HQ[5], NIC[6], BIC[7], MDL[8] などは適用できず，古典的な学習理論の枠組みの中では捉えることができない．したがって，近年急速に多くの理論の研究が始まった．

論文 [1, 2] では，ベイズ法における確率的複雑さおよび汎化誤差とゼータ関数の関係を明らかにしている．確率密度関数 $q(x)$ を持つ確率変数 X の入力の例，$x^n = (x_1, x_2, \ldots, x_n)$ が与えられたとき，ある条件付密度関数 $q(y|x)$ によって，出力の例 $y^n = (y_1, y_2, \ldots, y_n)$ が得られたとする．すなわち，(x^n, y^n) をサンプル数 n の学習データとする．学習モデルを $p(y|x, w)$ とし，事前分布を $\psi(w)$ とする．このとき，パラメータの事後確率は，

$$p(w|(x^n, y^n)) = \frac{1}{Z_n}\psi(w)\prod_{i=1}^{n} p(y_i|x_i, w),$$

である．Z_n は正規化定数である．このとき，ベイズ推測による学習後の学習モデルは次で与えられる：

$$p(y|x, (x^n, y^n)) = \int p(y|x, w)p(w|(x^n, y^n))\mathrm{d}w.$$

カルバック情報量を

$$K(q||p) = \int q(y|x)\log\frac{q(y|x)}{p(y|x)}q(x)\mathrm{d}x\mathrm{d}y,$$

とする．汎化誤差 $G(n)$ はサンプル (x^n, y^n) に関する平均である $E_n\{\cdot\}$ を用いて，

$$G(n) = E_n\{\int q(y|x)\log\frac{q(y|x)}{p(y|x, (x^n, y^n))}q(x)\mathrm{d}x\mathrm{d}y\},$$

となる．一方，経験カルバック情報量を

$$K_n(w) = \frac{1}{n}\sum_{i=1}^{n}\log\frac{q(y_i|x_i)}{p(y_i|x_i, w)},$$

とし，この経験カルバック情報量を用いて

$$F(n) = -E_n\{\log\int \exp(-nK_n(w))\psi(w)\mathrm{d}w\},$$

と定義される確率的複雑さは，漸近展開可能であれば，

*東京工業大学精密工学研究所〒 226–8503 神奈川県横浜市緑区長津田 4259, e-mail aoyagi@cs.pi.titech.ac.jp,
Precision and Intelligence Laboratory, Tokyo Institute of Technology, 4259 Nagatsuda, Midori-ku, Yokohama, 226–8503
†東京工業大学精密工学研究所〒 226–8503 神奈川県横浜市緑区長津田 4259, e-mail swatanab@pi.titech.ac.jp,
Precision and Intelligence Laboratory, Tokyo Institute of Technology, 4259 Nagatsuda, Midori-ku, Yokohama, 226–8503

$G(n)=F(n+1)-F(n)$, という関係をもつ．

学習モデルのゼータ関数を

$J(z)=\int K(w)^z\psi(w)\mathrm{d}w,$

とおく．z は一複素変数である．この関数は $\mathrm{Re}(z)>0$ で正則で，$\mathbf{C}$ 全体に有理型関数として拡張できることが知られている．さらに極は負の有理数である．この関数 $J(z)$ の最も原点に近い極を $-\lambda$，その位数を θ とすると，

$$F(n)=\lambda\log n-(\theta-1)\log\log n+O(1),$$

$$G(n)\cong\frac{\lambda}{n}-\frac{\theta-1}{n\log n},$$

が成り立つ．ここで，$O(1)$ は n の有界な関数である．学習モデルのゼータ関数の極は，特異点解消を用いれば計算できる量である [9],[10].

しかし，実はこのような数理の確立にもかかわらず，計算を行うことは非常に困難であった．特異点解消は，広中の定理により，ブローアップの有限の手続きにより可能であることが保証されているが，具体的に求めるのは難しいとされている．また，計算機による代数計算により行う方式も提案されているが，情報学における学習モデルのカルバック情報量は中間ユニット数などのパラメータを含んでいるため，多項式の特異点解消よりも，さらに高度な面を含んでいる．また，他の問題点として，特異点が孤立していない，ニュートン図形が退化している，などがあげられる．これらの特徴を持つ特異点は非常に複雑で代数幾何の分野でもあまり研究されていない．実際，ゼータ関数の極の値に関しては，概均質ベクトル空間などの特別な場合にしか研究されていないのが実情である．

この論文では，次のゼータ関数を考察する：

$$\int\Psi = \int(\sum_{n=1}^{H}\sum_{k=1}^{M}\sum_{j=1}^{N}(\sum_{i=1}^{H}a_{ki}b_{ij}^{p(n-1)+1})^2)^z \prod_{k=1}^{M}\prod_{i=1}^{H'}\mathrm{d}a_{ki}\prod_{i=1}^{H'}\prod_{j=1}^{N}\mathrm{d}b_{ij}. \qquad (1)$$

ここで，M,N,H,H',p は自然数で $H'\le H\le 2H'$ を満たすものとし，$i>H'$ に対して，a_{ki},b_{ij} は定数とする．N 個の入力層，H' 個の隠れ層，M 個の出力層を持つ三層ニューラルネットワークは，$H-H'$ 個の隠れ層をもつものを真のモデルとすると，ゼータ関数は $p=2$ を代入した $\int\Psi$ となる．また，混合正規分布は，$p=1$, $M=1$, $\sum_{j=1}^{H'}a_{1j}=1$，$\sum_{j=H'+1}^{H}a_{1j}=-1$ であるゼータ関数 $\int\Psi$ が対応する [11].

特にこの論文では，M, N は任意であるが，$H'=2$ の場合を考察する．$M=N=1$ かつ任意の H, H' の場合は，すでに論文 [12]，[13] において明らかにされている．また，論文 [14] では，隠れ層の関数が線形の三層ニューラルネットワークである縮小ランクモデルの場合について考察している．このモデルは

$$\int(\sum_{k=1}^{M}\sum_{j=1}^{N}(\sum_{i=1}^{H}a_{ki}b_{ij})^2)^z\prod_{k=1}^{M}\prod_{i=1}^{H'}\mathrm{d}a_{ki}\prod_{i=1}^{H'}\prod_{j=1}^{N}\mathrm{d}b_{ij},$$

に対応する

2　主定理

行列 $C=(c_{ij})$ のノルムを $||C||=\sqrt{\sum_{i,j}|c_{ij}|^2}$ とする．簡単のため，微分形式 $\prod_{i=1}^{M}\prod_{j=1}^{H}\mathrm{d}a_{ij}$ の代わりに $\mathrm{d}a$ を使用する．

主定理 1

$$A=\begin{pmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\\ & \vdots \\ a_{M1} & a_{M2}\end{pmatrix} \text{ および } B_j=\begin{pmatrix} b_{1j} & b_{1j}^{p+1}\\ b_{2j} & b_{2j}^{p+1}\end{pmatrix},$$

$B=(B_1,\cdots,B_N)$ とする．

また，

$$\Psi=||AB||^{2z}\mathrm{d}a\mathrm{d}b. \qquad (2)$$

と定義する．このとき，$\int_{||AB||<1}\Psi$ の最大の極 $-\lambda$ およびその位数 θ は次のようになる．

(a)

(i) $N\ge M+1$ のとき $\lambda=M$ かつ $\theta=1$.

(ii) $N=M$ のとき $\lambda=\frac{2N+p(M+N-1)}{2(p+1)}$ かつ $\theta=1$.

(iii) $N=M-1$ のとき $\lambda=N$ かつ $\theta=2$.

(iv) $N<M-1$ のとき $\lambda=N$ かつ $\theta=1$.

(b) $a_{k1}+a_{k2}=1$ の条件のもとで．

(i) $N>M+p$ のとき $\lambda=\frac{M+N}{2}$ かつ $\theta=1$.

(ii) $N=M+p$ のとき $\lambda=\frac{M+N}{2}$ かつ $\theta=2$.

(iii) $M-1<N<M+p$ のとき $\lambda=\frac{2N+p(M+N-1)}{2(p+1)}$ かつ $\theta=1$.

(iv) $N=M-1$ のとき $\lambda=N$ かつ $\theta=2$.

(v) $N<M-1$ のとき $\lambda=N$ かつ $\theta=1$.

主定理 2　$A=\begin{pmatrix} a_{11} & a_{12} & a_{13}^*\\ a_{21} & a_{22} & a_{23}^*\\ & \vdots & \\ a_{M1} & a_{M2} & a_{M3}^*\end{pmatrix}$ および $B_j=\begin{pmatrix} b_{1j} & b_{1j}^{p+1} & b_{1j}^{2p+1}\\ b_{2j} & b_{2j}^{p+1} & b_{2j}^{2p+1}\\ b_{3j}^* & {b_{3j}^*}^{p+1} & {b_{3j}^*}^{2p+1}\end{pmatrix}$, $B=(B_1,\cdots,B_N)$ とする．さらに，任意の k,j に対して，a_{k3}^*, b_{3j}^* は定数とし，2 個以上の $b_{3j}^*\neq 0$ が 0 ではなく，a_{k3}^* のどれかは 0 でないとしておく．

このとき，

$$\Psi = ||AB||^{2z} \mathrm{d}a\mathrm{d}b, \tag{3}$$

とおくと，$\int_{||AB||<1} \Psi$ の最大の極 $-\lambda$ およびその位数 θ は次のようになる．

(a)

(i) $N > M+1$ のとき $\lambda = \frac{2M+N}{2}$ かつ $\theta = 1$.

(ii) $N = M+1$ のとき $\lambda = \frac{2M+N}{2}$ かつ $\theta = 2$.

(iii) $N = M$ のとき $\lambda = \frac{3M+3N-1}{4}$ かつ $\theta = 1$.

(iv) $N = M-1$ のとき $\lambda = \frac{M+2N}{2}$ かつ $\theta = 2$.

(v) $N < M-1$ のとき $\lambda = \frac{M+2N}{2}$ かつ $\theta = 1$.

(b) $a_{k1} + a_{k2} = 1$ および $a^*_{k3} = -1$ の条件のもとで.

(i) $N > M+1$ のとき $\lambda = \frac{M+N}{2}$ かつ $\theta = 1$.

(ii) $N = M+1$ のとき $\lambda = \frac{M+N}{2}$ かつ $\theta = 2$.

(iii) $N = M$ のとき $\lambda = \frac{M+3N-1}{4}$ かつ $\theta = 1$.

(iv) $N = M-1$ のとき $\lambda = N$ かつ $\theta = 2$.

(v) $N < M-1$ のとき $\lambda = N$ かつ $\theta = 1$.

式 (1) のゼータ関数 $\int \Psi$ に対応する 2 つの例をここに示す．

例 1 N 個の入力層，H' 個の隠れ層，M 個の出力層を持つ三層ニューラルネットワークを考える．$H - H'$ 個の隠れ層をもつものを真のモデルとする．入力を $x = (x_j) \in \mathbb{R}^N$ とし，その確率密度関数を $q(x)$ とする．$q(x)$ はコンパクトサポート $\tilde{W}$ をもつとする．このとき，出力 $y = (y_k) \in \mathbb{R}^M$ は，$w = \{a_{ki}, b_{ij}; 1 \le k \le M, 1 \le i \le H', 1 \le j \le N\}$ および,

$$f_k(x,w) = \sum_{i=1}^{H'} a_{ki} \tanh(\sum_{j=1}^{N} b_{ij} x_j)$$

としたとき，$y_k = f_k(x,w) + (\text{noise})$ で与えられる．ここで学習モデルを,

$$p(y|x,w) = \frac{1}{(2\pi)^{M/2}} \exp(-\frac{1}{2}||y - f(x,w)||^2),$$

とする．真のモデルを

$$p(y|x,w^*) = \frac{1}{(2\pi)^{M/2}} \exp(-\frac{1}{2}||y - f(x,w^*)||^2),$$

とし，学習モデルに含まれているものとする．ここで，$w^* = \{a^*_{ki}, b^*_{ij}; 1 \le k \le M, H'+1 \le i \le H, 1 \le j \le N\}$, $f_k(x,w^*) = \sum_{i=H'+1}^{H} (-a^*_{ki}) \tanh(\sum_{j=1}^{N} b^*_{ij} x_j)$, である．事前分布 $\psi(w)$ を $C^\infty-$ 関数で，コンパクトサポート W を持ち，$\psi(w^*) > 0$ であるとする．

このとき，$-\lambda$ および θ は，式 (1) に $p = 2$ を代入した $\int \Psi$ に対する最大の極および位数になる．

これは，論文 [1] の Lemma 5 とテーラー展開を用いて証明される．

例 2 混合正規分布

$$p(x|w) = \frac{1}{(2\pi)^{N/2}} \sum_{i=1}^{H'} a_{1i} \exp(-\frac{\sum_{j=1}^{N}(x_j - b_{ij})^2}{2}),$$

$\sum_{i=1}^{H'} a_{1i} = 1$ を考える．ここで，$w = \{a_{1i}, b_{ij}; 1 \le i \le H', 1 \le j \le N\}$ である学習モデルに含まれる真の分布を

$$p(x|w^*) = \frac{1}{(2\pi)^{N/2}} \sum_{i=H'+1}^{H} (-a^*_{1i}) \exp(-\frac{\sum_{j=1}^{N}(x_j - b^*_{ij})^2}{2}),$$

$\sum_{i=H'+1}^{H} a^*_{1i} = -1$ とする．w^* は $w^* = \{a^*_{1i}, b^*_{ij}; H'+1 \le i \le H, 1 \le j \le N\}$ で定義される．事前分布 $\psi(w)$ は，$C^\infty-$ 関数でコンパクトサポート W を持つものとし，$\psi(w^*) > 0$ であるとしておく．このとき，ゼータ関数は，$p = 1$，$M = 1$ を式 (1) に代入した $\int \Psi$ である [11].

3 主定理１の証明

次の式 $(*)$ における自然数 k にたいする帰納法とブローアップで証明する．

$$(*) \qquad \Psi^* = \{v_{11}^{2k}(d_{11}^2 + \cdots + d_{M1}^2) + v_{11}^{2k} \left\| \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{M2} \end{pmatrix} (B_1, \cdots, B_N) \right\|^2 \}^z v_{11}^{2N-1+(k-1)(M+N-1)} \mathrm{d}d\mathrm{d}v\mathrm{d}a\mathrm{d}b.$$

ここで，$B_1 = v_{11}^{p+1-k} b'_{21}$ および $B_j = b'_{2j}$, $(j \ge 2)$ である．

多様体 $\{b_{ij} = 0, i = 1, 2, 1 \le j \le N\}$ にそって，式 (2) の Ψ をブローアップする．

$b_{11} = v_{11}$, $b_{ij} = v_{11} b_{ij}$ $(i,j) \ne (1,1)$ とおく．このとき,

$$\Psi = \left\| \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ & \vdots \\ a_{M1} & a_{M2} \end{pmatrix} (B_1, \cdots, B_N) \right\|^{2z} v_{11}^{2N-1} \mathrm{d}v\mathrm{d}a\mathrm{d}b$$

となる．ここで，$B_1 = \begin{pmatrix} v_{11} & v_{11}^{p+1} \\ v_{11} b_{21} & v_{11}^{p+1} b_{21}^{p+1} \end{pmatrix}$ および $B_j = \begin{pmatrix} v_{11} b_{1j} & v_{11}^{p+1} b_{1j}^{p+1} \\ v_{11} b_{2j} & v_{11}^{p+1} b_{2j}^{p+1} \end{pmatrix}$, $j \ge 2$ である．

論文 [14] における Lemma 2，3 をもちいると $\int_W \Psi$ の最大の極とその位数は，次で定義される $\int_W \Psi'$ の最大の極とその位数に等しい．

$$\Psi' = \left\| \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ & \vdots \\ a_{M1} & a_{M2} \end{pmatrix} (B_1,\cdots,B_N) \right\|^{2z} v_{11}^{2N-1}\mathrm{d}v\mathrm{d}a\mathrm{d}b,$$

ここで $B_1 = \begin{pmatrix} v_{11} & 0 \\ v_{11}b_{21} & v_{11}^{p+1}\begin{vmatrix} 1 & 1 \\ b_{21} & b_{21}^{p+1} \end{vmatrix} \end{pmatrix}$ および $B_j = \begin{pmatrix} 0 & 0 \\ v_{11}\begin{vmatrix} 1 & b_{1j} \\ b_{21} & b_{2j} \end{vmatrix} & v_{11}^{p+1}\begin{vmatrix} 1 & b_{1j}^{p+1} \\ b_{21}^{p+1} & b_{2j}^{p+1} \end{vmatrix} \end{pmatrix}$, $(j \geq 2)$ である．

そこで，変数 $a_{11},\ldots,a_{M1}$ から $d_{11},\ldots,d_{M1}$ へ，式

$$\begin{pmatrix} d_{11} \\ d_{21} \\ \vdots \\ d_{M1} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ & \vdots \\ a_{M1} & a_{M2} \end{pmatrix} \begin{pmatrix} 1 \\ b_{21} \end{pmatrix},$$

で変数変換する．

$\mathbf{a_{i1}+a_{i2}=1}$ の場合． $a_{i1}+a_{i2}=1$ であれば，$d_{i1}=a_{i1}(1-b_{21})+b_{21}$ となる．もし，$b_{21}=1$ であれば，$d_{i1}=1$ となり，ブローアップはもう必要ない．

したがって，

$$\Psi' = \{v_{11}^2(d_{11}^2+\cdots+d_{M1}^2) + v_{11}^2 \left\| \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{M2} \end{pmatrix} (B_1,\cdots,B_N) \right\|^2 \}^z v_{11}^{2N-1}\mathrm{d}d\mathrm{d}v\mathrm{d}a\mathrm{d}b,$$

$B_1 = v_{11}^p \begin{vmatrix} 1 & 1 \\ b_{21} & b_{21}^{p+1} \end{vmatrix}$,
$B_j = \begin{pmatrix} \begin{vmatrix} 1 & b_{1j} \\ b_{21} & b_{2j} \end{vmatrix} & v_{11}^p \begin{vmatrix} 1 & b_{1j}^{p+1} \\ b_{21}^{p+1} & b_{2j}^{p+1} \end{vmatrix} \end{pmatrix}$, $(j \geq 2)$ となる．

$b'_{21} = \begin{vmatrix} 1 & 1 \\ b_{21} & b_{21}^{p+1} \end{vmatrix}$ かつ $b'_{2j} = \begin{vmatrix} 1 & b_{1j} \\ b_{21} & b_{2j} \end{vmatrix}$ とおけば，$B_j = \begin{pmatrix} b'_{2j} & v_{11}^p((b'_{2j}+b_{21}b_{1j})^{p+1}-(b_{21}b_{1j})^{p+1}) \end{pmatrix}$, $(j \geq 2)$ であるから，さらに，論文 [14] の Lemma 2，3 より，

$$\Psi'' = \{v_{11}^2(d_{11}^2+\cdots+d_{M1}^2) + v_{11}^2 \left\| \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{M2} \end{pmatrix} (B_1,\cdots,B_N) \right\| \}^{2z} v_{11}^{2N-1}\mathrm{d}d\mathrm{d}v\mathrm{d}a\mathrm{d}b,$$

および $B_1 = v_{11}^p b'_{21}$，$B_j = b'_{2j}$, $(j \geq 2)$ としてよい．

これで，帰納法 $(*)$ の $k=1$ が示された．

多様体 $\{v_{11}=b'_{22}=\cdots=b'_{2N}=d_{11}=\cdots=d_{M1}=0\}$ に沿って Ψ^* のブローアップを行う．次の (I), (II), (III) の場合が得られる．

(I) Ψ^* に $d_{11}=u_{11}, v_{11}=u_{11}v_{11}, b'_{2j}=u_{11}b'_{2j}, (j \geq 2)$, $d_{i1}=u_{11}d_{i1}, (i \geq 2)$ を代入すると，

$$\Psi'^* = \{v_{11}^{2k}u_{11}^{2k+2}(1+d_{21}^2+\cdots+d_{M1}^2) + v_{11}^{2k}u_{11}^{2k+2} \left\| \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{M2} \end{pmatrix} (B_1,\cdots,B_N) \right\|^2 \}^z \times v_{11}^{2N-1+(k-1)(M+N-1)} u_{11}^{2N-1+k(M+N-1)}\mathrm{d}d\mathrm{d}u\mathrm{d}v\mathrm{d}a\mathrm{d}b,$$

および $B_1 = v_{11}^{p+1-k}u_{11}^{p-k}b'_{21}$，$B_j = b'_{2j}$, $(j \geq 2)$ となる．したがって，極 $-\dfrac{2N+(k-1)(M+N-1)}{2k}$ および $-\dfrac{2N+k(M+N-1)}{2k+2}$ が得られる．

(II) Ψ^* に $b'_{22}=v_{22}, v_{11}=v_{22}v_{11}, b'_{2j}=v_{22}b'_{2j}, (j \geq 3)$, $d_{i1}=v_{22}d_{i1}, (i \geq 1)$ を代入すれば，

$$\Psi'^* = \{v_{11}^{2k}v_{22}^{2k+2}(d_{11}^2+\cdots+d_{M1}^2) + v_{11}^{2k}v_{22}^{2k+2} \left\| \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{M2} \end{pmatrix} (B_1,\cdots,B_N) \right\|^2 \}^z v_{11}^{2N-1+(k-1)(M+N-1)} v_{22}^{2N-1+k(M+N-1)}\mathrm{d}d\mathrm{d}v\mathrm{d}a\mathrm{d}b,$$

および，$B_1 = v_{11}^{p+1-k}v_{22}^{p-k}b'_{21}$, $B_2=1$, $B_j=b'_{2j}$, $(j \geq 3)$ となる．

再度，論文 [14] の Lemma 2，3 を用いると，次の $\int_W \Psi''^*$ を考えればよい．

$$\Psi''^* = \{v_{11}^{2k}v_{22}^{2k+2}(d_{11}^2+\cdots+d_{M1}^2) + v_{11}^{2k}v_{22}^{2k+2} \left\| \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{M2} \end{pmatrix} \right\|^2 \}^z v_{11}^{2N-1+(k-1)(M+N-1)} v_{22}^{2N-1+k(M+N-1)}\mathrm{d}d\mathrm{d}v\mathrm{d}a\mathrm{d}b.$$

$\mathbf{a_{i1}+a_{i2}=1}$ の場合． $a_{i1}+a_{i2}=1$ であれば，$a_{i2}=(1-v_{11}^{2k-2}v_{22}^{2k}d_{i1})/(1-b_{21})$ となるため，ブローアップは終了する．

多様体 $\{a_{12}=\cdots=a_{M2}=d_{11}=\cdots=d_{M1}=0\}$ にそって Ψ''^* をブローアップする．

Ψ''^{*} に $d_{11}=u_{11}$, $d_{i1}=u_{11}d_{i1}$, $(i\geq 2)$, $a_{i2}=u_{11}a_{i2}$, $(i\geq 1)$ を代入すると，

$$\Psi'''^{*}=\{v_{11}^{2k}v_{22}^{2k+2}u_{11}^{2}(1+d_{21}^{2}+\cdots+d_{M1}^{2}) +v_{11}^{2k}v_{22}^{2k+2}u_{11}^{2}\left\|\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{M2}\end{pmatrix}\right\|^{2}\}^{z}v_{11}^{2N-1+(k-1)(M+N-1)} v_{22}^{2N-1+k(M+N-1)}u_{11}^{2M-1}\mathrm{d}d\mathrm{d}u\mathrm{d}v\mathrm{d}a\mathrm{d}b,$$

を得る．したがって，極 $-\dfrac{2N+(k-1)(M+N-1)}{2k}$ および $-\dfrac{2N+k(M+N-1)}{2k+2}$, $-M$ が得られる．

(III) Ψ^{*} に $b'_{2j}=v_{11}b'_{2j}$, $(j\geq 2)$, $d_{i1}=v_{11}d_{i1}$, $(i\geq 1)$ を代入すると，

$$\Psi'^{*}=\{v_{11}^{2k+2}(d_{11}^{2}+\cdots+d_{M1}^{2}) +v_{11}^{2k+2}\left\|\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{M2}\end{pmatrix}(B_1,\cdots,B_N)\right\|^{2}\}^{z} v_{11}^{2N-1+k(M+N-1)}\mathrm{d}d\mathrm{d}v\mathrm{d}a\mathrm{d}b,$$

$B_1=v_{11}^{p-k}b'_{21}$, $B_j=b'_{2j}$, $(j\geq 2)$ が得られる．したがって，$k+1$ の場合の帰納法の仮定 $(*)$ を得る．

明らかに，帰納法の仮定 $(*)$ は $k=p+1$ で終了し，このとき，

$$\Psi^{*}_{p+1}=\{v_{11}^{2(p+1)}(d_{11}^{2}+\cdots+d_{M1}^{2}) +v_{11}^{2(p+1)}\left\|\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{M2}\end{pmatrix}(B_1,\cdots,B_N)\right\|^{2}\}^{z} v_{11}^{2N-1+p(M+N-1)}\mathrm{d}d\mathrm{d}v\mathrm{d}a\mathrm{d}b,$$

$B_1=b'_{21}$, $B_j=b'_{2j}$, $(j\geq 2)$ となる．

多様体 $\{b_{21}=\cdots=b_{2N}=d_{11}=\cdots=d_{M1}=0\}$ にそって Ψ^{*}_{p+1} をブローアップする．(I'), (II') の場合が現れる．

(I') Ψ^{*}_{p+1} に $b'_{21}=v_{21}$, $b'_{2j}=v_{21}b'_{2j}$, $(j\geq 2)$, $d_{i1}=v_{21}d_{i1}$, $(i\geq 1)$ を代入する．このとき，

$$\Psi'^{*}_{p+1} = \{v_{11}^{2(p+1)}v_{21}^{2}(d_{11}^{2}+\cdots+d_{M1}^{2}) +v_{11}^{2(p+1)}v_{21}^{2}\left\|\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{M2}\end{pmatrix}\right\|^{2}\}^{z} v_{11}^{2N-1+p(M+N-1)}v_{21}^{M+N-1}\mathrm{d}d\mathrm{d}v\mathrm{d}a\mathrm{d}b,$$

となる．

$\mathbf{a_{i1}+a_{i2}=1}$ の場合. $a_{i2}=(1-v_{11}^{2p}v_{21}^{2}d_{i1})/(1-b_{21})$ よりブローアップは終了する．

多様体 $\{a_{12}=\cdots=a_{M2}=d_{11}=\cdots=d_{M1}=0\}$ にそって，Ψ'^{*}_{p+1} をブローアップする．

Ψ'^{*}_{p+1} に $d_{11}=u_{11}$, $d_{i1}=u_{11}d_{i1}$, $(i\geq 2)$, $a_{i2}=u_{11}a_{i2}$, $(i\geq 1)$ を代入すると，

$$\Psi''^{*}_{p+1} = \{v_{11}^{2(p+1)}v_{21}^{2}u_{11}^{2}(1+d_{21}^{2}+\cdots+d_{M1}^{2}) +v_{11}^{2(p+1)}v_{21}^{2}u_{11}^{2}\left\|\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{M2}\end{pmatrix}\right\|^{2}\}^{z} v_{11}^{2N-1+p(M+N-1)}v_{21}^{M+N-1}u_{11}^{2M-1}\mathrm{d}d\mathrm{d}u\mathrm{d}v\mathrm{d}a\mathrm{d}b.$$

したがって，極 $-\dfrac{2N+p(M+N-1)}{2(p+1)}$, $-\dfrac{M+N}{2}$ および $-M$ を得る．

(II') Ψ^{*}_{p+1} に $d_{11}=u_{11}$, $b'_{2j}=u_{11}b'_{2j}$, $(j\geq 1)$, $d_{i1}=u_{11}d_{i1}$, $(i\geq 2)$ を代入すると，

$$\Psi'^{*}_{p+1} = \{v_{11}^{2(p+1)}u_{11}^{2}(1+d_{21}^{2}+\cdots+d_{M1}^{2}) +v_{11}^{2(p+1)}u_{11}^{2}\left\|\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{M2}\end{pmatrix}(B_1,\cdots,B_N)\right\|^{2}\}^{z} v_{11}^{2N-1+p(M+N-1)}u_{11}^{M+N-1}\mathrm{d}d\mathrm{d}u\mathrm{d}v\mathrm{d}a\mathrm{d}b,$$

および，$B_1=b'_{21}$，$B_j=b'_{2j}$, $(j\geq 2)$ を得る．

したがって極 $-\dfrac{2N+p(M+N-1)}{2(p+1)}$ および $-\dfrac{M+N}{2}$ を得る．

これらの結果から，極 $-\dfrac{2N+(k-1)(M+N-1)}{2k}$, $(k=1,\cdots,p+1)$, $-\dfrac{M+N}{2}$, M が得られる．

$\mathbf{a_{i1}+a_{i2}=1}$ の場合. M のみが現れない．

主定理 2 の証明は省略する．ブローアップを用いて証明できるが，定数項が存在するため場合わけが多くなりさらに複雑な証明が必要である．

4 おわりに

この論文では，三層学習モデルの確率的複雑さ，汎化誤差を考察した．

混合正規分布
$p(x|w) = \frac{1}{(2\pi)^{N/2}} \sum_{i=1}^{2} a_{1i} \exp(-\frac{\sum_{j=1}^{N}(x_j - b_{ij})^2}{2})$, $a_{11} + a_{12} = 1$, は，真の分布が
$p(x|w^*) = \frac{1}{(2\pi)^{N/2}} \exp(-\frac{\sum_{j=1}^{N}(x_j - b_j^*)^2}{2})$ のとき，$p = 1$, $M = 1$ を主定理 2 (b) に代入した結果によって，確率的複雑さ，汎化誤差が得られる．

主定理 1 (b) は，真のモデルが
$p(x|w^*) = \frac{1}{(2\pi)^{N/2}} \exp(-\frac{\sum_{j=1}^{N} x_j^2}{2})$, である場合に対応している．明らかに，主定理 1 (b) と主定理 2 (b) が $p = 1$ のとき同じ結果になる．これは，正規分布の中心が 0 かまたは $(b_1^*, \ldots, b_N^*) \neq 0$ であるかの違いだけであるからである．

これらの結果の応用として，MCMC 法の精度の評価や改良の数学的基盤を作ることができる [15], [16]．また，局所ベイズ法の結果との比較などが今後の課題である [17].

今後は，任意の H, H' の場合について考察できるよう，この方法の改良を行いたい．

謝辞

この研究は科学研究費補助金 16700218 を受けた．

参考文献

[1] S. Watanabe, Algebraic analysis for nonidentifiable learning machines, *Neural Computation*, 13 (4), 2001, 899-933.

[2] S. Watanabe, Algebraic geometrical methods for hierarchical learning machines, *Neural Networks*, 14 (8) 2001, 1049-1060.

[3] H. Akaike, A new look at the statistical model identification, *IEEE Trans. on Automatic Control*, 19 1974, 716-723.

[4] K. Takeuchi, Distribution of an information statistic and the criterion for the optimal model, *Mathematical Science*, 153, 1976, 12-18.

[5] E. J. Hannan and B. G. Quinn, The determination of the order of an autoregression. *Journal of Royal Statistical Society, Series B*, 41, 1979, 190-195.

[6] N. J. Murata, S. G. Yoshizawa, and S. Amari, Network information criterion - determining the number of hidden units for an artificial neural network model, *IEEE Trans. on Neural Networks*, 5 (6), 1994, 865-872.

[7] G. Schwarz, Estimating the dimension of a model, *Annals of Statistics*, 6 (2) 1978, 461-464.

[8] J. Rissanen, Universal coding, information, prediction, and estimation, *IEEE Trans. on Information Theory*, 30 (4) 1984, 629-636.

[9] H. Hironaka, Resolution of Singularities of an algebraic variety over a field of characteristic zero, *Annals of Math*, 79 1964, 109-326.

[10] 渡辺澄夫, 萩原克幸, 赤穂昭太郎, 本村陽一, 福水健二, 岡田真人, 青柳美輝, 学習システムの理論と実現, 森北出版，2005, p. 195.

[11] S. Watanabe, K. Yamazaki and M. Aoyagi, Kullback Information of Normal Mixture is not an Analytic Function, *Technical report of IEICE*, NC2004, 2004, 41-46.

[12] M. Aoyagi and S. Watanabe, Resolution of Singularities and the Generalization Error with Bayesian Estimation for Layered Neural Network, *IEICE Trans.*, J88-D-II, 10, 2005, 2112-2124, (English version : *Systems and Computers in Japan,* John Wiley & Sons, Inc. (in press).

[13] M. Aoyagi, The zeta function of learning theory and generalization error of three layered neural perceptron, *RIMS Kokyuroku, Recent Topics on Real and Complex Singularities*, 2006, (in press).

[14] M. Aoyagi and S. Watanabe, Stochastic Complexities of Reduced Rank Regression in Bayesian Estimation, *Neural Networks*, 18, 2005, 924-933.

[15] N. Nakano, K. Takahashi and S. Watanabe, On the evaluation criterion of the MCMC Method in Singular Learning machines, *IEICE Trans.*, J88-D-II, 10, 2005, 2011-2020.

[16] K. Nagata and S. Watanabe, A proposal and effectiveness of the optimal approximation for Bayesian posterior distribution, *Workshop on Information-Based Induction Sciences,* 2005, 99-104.

[17] S. Takamatsu, S. Nakajima and S. Watanabe, Generalization Error of Localized Bayes Estimation in Reduced Rank Regression, *Workshop on Information-Based Induction Sciences,* 2005, 81-86.