社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

混合正規分布の特異点の非解析性について

渡辺　澄夫† 　山崎　啓介† 　青柳　美輝††

† 東京工業大学　精密工学研究所　〒 226-8503 横浜市緑区長津田 4259
†† 上智大学理工学部数学科　〒 102–8554 東京都千代田区紀尾井町 7–1
E-mail: †swatanab@pi.titech.ac.jp, ††k-yam@cs.pi.titech.ac.jp, †††miki-a@sophia.math.ac.jp

あらまし　本論文では、以下のことを述べる。(1) 混合正規分布のカルバック情報量が非解析的であることを示す。(2) カルバック情報量が非解析的である場合にも成立つ確率的複雑さの漸近展開の定理を証明し、混合正規分布がその条件を満たすことを示す。(3) 定理に基づいて混合正規分布のカルバック情報量と等価な多項式を導出する。

キーワード　混合正規分布、特異点、確率的複雑さ、非解析性、ゼータ関数

Kullback Information of Normal Mixture is not an Analytic Function

Sumio WATANABE†, Keisuke YAMAZAKI†, and Miki AOYAGI††

† PI Lab., Tokyo Institute of Technology　4259 Nagatsuda, Midori-ku, Yokohama 226-8503 Japan
†† Faculty of Science and Technology, Sophia University
7–1, Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan
E-mail: †swatanab@pi.titech.ac.jp, ††k-yam@cs.pi.titech.ac.jp, †††miki-a@sophia.math.ac.jp

Abstract　In this paper, we show three new points. Firstly, we show that the Kullback information of a normal mixture is not an analytic function of the parameter. Secondly, we prove the new theorem which gives the asymptotic expansion of the stochastic complexity of non-analytic learning machines. The normal mixture satisfies the condition of the theorem. And lastly, we show the polynomial which is equivalent to the Kullback information of normal mixture.

Key words　Normal mixture, singularities, stochastic complexity, non-analytic Kullback information, zeta function.

1. まえがき

本論では、パラメータ $w \in \mathbf{R}^d$ によって定まる $x \in \mathbf{R}^N$ の確率分布 $p(x|w)$ について、フィッシャー情報行列

$$
\begin{aligned}
I_{ij}(w) &= \int \frac{\partial L}{\partial w_i}\frac{\partial L}{\partial w_j} p(x|w)dx \\
L(w) &= \log p(x|w)
\end{aligned}
$$

の行列式 $(\det I(w))$ が 0 になるパラメータ w が存在するとき、$p(x|w)$ を特異モデルと呼ぶ。パラメータから確率分布への写像 $w \mapsto p(x|w)$ が、一対一でない学習モデルは特異モデルになることが多い。混合正規分布、多層パーセプトロン、ボルツマンマシン、ベイズネットワーク、縮小ランク回帰など、階層構造、対称構造、モジュール構造を持つ学習モデルは全て特異モデルである。特異モデルにおいては、統計的正則モデルの漸近理論は成立せず、最尤推定量や MAP 推定量の分布は正規分布に近づかない。またベイズ事後分布も正規分布とは異なる分布に法則収束する。

サンプルを発生している真の分布が $q(x) = p(x|w_0)$ であると仮定し、独立なサンプルを

$$X^n = (X_1, X_2, ..., X_n)$$

と書く。真の分布から学習モデル $p(x|w)$ までのカルバック情報量は

$$H(w) = \int q(x) \log \frac{q(x)}{p(x|w)} dx$$

である。我々はこれまで、カルバック情報量がパラメータ w の解析関数（テーラー展開が絶対収束する関数）であるときに、確率的複雑さ（周辺対数尤度、自由エネルギー）

$$
\begin{aligned}
F(X^n) &= -\log \int \exp(-nH_n(w))\varphi(w)dw \\
H_n(w) &= \frac{1}{n}\sum_{i=1}^{n} \log \frac{q(X_i)}{p(X_i|w)}
\end{aligned}
$$

が次の漸近展開を持つことを証明してきた [22] [23]。

$$F(X^n) = \lambda \log n - (m-1)\log\log n + R + o(1)$$

ここで、有理数 $(-\lambda)$ と自然数 m は、それぞれゼータ関数 [5] [10]

$$\zeta(z) = \int H(w)^z \varphi(w) dw$$

の最も原点に近い極とその位数であり、また R はある確率変数である。確率的複雑さは、ベイズ学習においては学習モデル $p(x|w)$ と事前分布 $\varphi(w)$ の対数尤度を表しているので、モデルの設計や選択に利用される。また、その増分 $F(X^{n+1}) - F(X^n)$ の平均値は、汎化誤差に等しいことが知られているので、ベイズ学習における最も基本的な量である。

ところで、これまでに導かれた結果では、特異点解消定理 [5] [9] やｂ関数 [10] などの数学的方法を利用するために、カルバック情報量が解析関数であることが仮定されてきたが、本論で示すように、混合正規分布においては、カルバック情報量は一般には解析関数ではない。この論文では、そのような場合にも成立つ定理を与え、混合正規分布のカルバック情報量がある多項式と等価になることを明らかにし、その多項式に関する特異点解消が学習係数を与えることを明らかにする。

2. 混合正規分布の非解析性

本論文では、次のような学習モデルを考える。

$$p(x|w) = \sum_{h=1}^{H} a_h \frac{1}{(2\pi)^{N/2}} \exp(-\frac{\|x-b_h\|^2}{2}) \qquad (1)$$

ここで $a_1, ..., a_H$ は、非負のパラメータで、

$$a_1 + a_2 + \cdots + a_H = 1$$

を満たすとする。また $b_h \in \mathbf{R}^N$ である。学習モデルのパラメータは

$$w = (a_1, .., a_H, b_1, .., b_H)$$

である。真の分布は、次の式であるとする。

$$q(x) = \sum_{h=1}^{H_0} a_h^* \frac{1}{(2\pi)^{N/2}} \exp(-\frac{\|x-b_h^*\|^2}{2}) \qquad (2)$$

ここで $a_1^*, ..., a_{H_0}^*$ は正の定数で、

$$a_1^* + a_2^* + \cdots + a_{H_0}^* = 1$$

を満たすとする。カルバック情報量は

$$H(w) = \int q(x) \log \frac{q(x)}{p(x|w)} dx$$

である。

［定理 1］ 混合正規分布のカルバック情報量 $H(w)$ は w の解析関数でないことがある。

（証明）解析的でない例があることを示す。以下、$N = 1$ とする。

$$f(x|b) = \frac{1}{(2\pi)^{1/2}} \exp(-\frac{(x-b)^2}{2})$$

とおく。学習モデルが

$$p(x|w) = (1-a)f(x|b) + af(x|c)$$

であり、真の分布が

$$q(x) = f(x|0)$$

である場合を考える。多変数の解析関数は各変数毎にも解析関数であるから、ある変数について解析的にならないことを示せば十分である。$b = 0, c > 0$ のとき、$H(w)$ が a について解析関数でないことを示す。

$$\log \frac{q(x)}{p(x|w)} = -\log \left[1 + a(e^{cx - \frac{c^2}{2}} - 1)\right]$$

であるから、

$$\frac{\partial^k H}{\partial a^k} = \frac{(-1)^k}{k} \int q(x) \left[e^{cx-\frac{c^2}{2}} - 1\right]^k dx$$

である。固定した $c > 0$ に対して

$$e^{cx-\frac{c^2}{2}} - 1 = e^{cx/2}$$

をみたす x を x_0 とおくと $x > x_0$ では

$$e^{cx-\frac{c^2}{2}} - 1 > e^{cx/2}$$

が成立つ。従って

$$\begin{aligned} \left|\frac{\partial^k H}{\partial a^k}\right| &\geqq \frac{1}{k} \int_{x_0}^{\infty} q(x) e^{kcx/2} dx \\ &= \frac{1}{k} \exp(k^2c^2/8) A_n \end{aligned}$$

ここで

$$A_k = \int_{x_0}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp(-\frac{(x-kc/2)^2}{2}) dx$$

とおいた。$k \to \infty$ のとき、$A_k \to 1$ なので十分大きな k について $A_k > 0.9$ としてよい。a についての収束半径を r とすれば

$$\frac{1}{r} \geqq \limsup_{k\to\infty} \left[\frac{1}{k \cdot k!} \exp(k^2c^2/8) 0.9\right]^{1/k} = \infty$$

従って、収束半径は０である。（証明終り）

注意 学習モデルのパラメータを $a_k = \alpha_k^2$ と置き換えて α_k をパラメータと考えても、k の代わりに $2k$ を考えれば上記と同様の理由により、カルバック情報量は非解析的である。

3. 主 定 理

上記のように混合正規分布のカルバック情報量が解析関数とは限らないので、カルバック情報量が解析関数でない場合の確率的複雑さの解析は、情報学上も十分な意義を有すると考えら

れる。以下では、混合正規分布を含む一般のモデルに対する定理を与える。定理が成立つための条件を述べる。

[条件 1] 事前分布 $\varphi(w)$ はコンパクトサポートであり、C^∞ 級関数 $\varphi_0(w)$ と解析関数 $\varphi_1(w)$ とにより、

$$\varphi(w) = \begin{cases} \varphi_0(w) & (\varphi(w) > 0) \\ 0 & (\text{otherwise}) \end{cases}$$

と表されるとする。$\varphi(w)$ のサポートを W と書く。

この条件 1 は、カルバック情報量が解析関数である場合にも同じ条件が課されていた。次の条件 2 が、本論文で新たに導入するものである。実数 $t > 0$ の関数

$$S(t) = \frac{-\log t + t - 1}{(t-1)^2}$$

を考える。$t = 1$ は除去可能な特異点で、$S(t)$ は解析関数であり、$S(t) > 0$ である。

[条件 2] $\mathbf{R}^N$ 上のある可測関数 $M(x)$ が存在して

$$\sup_{w\in W} S\left(\frac{p(x|w)}{q(x)}\right) \leqq M(x)$$

が成立つ。さらに内積

$$(u, v) = \int u(x)v(x)M(x)q(x)dx$$

によって定まるヒルベルト空間を $\mathcal{H}$ とするとき、写像

$$W \ni w \mapsto \frac{p(x|w)}{q(x)} - 1 \in \mathcal{H}$$

は、$\mathcal{H}$ の位相で解析的である。(すなわち、$p(x|w)/q(x) - 1$ の w に関するテーラー展開が、ヒルベルト空間の位相で絶対収束する)。

以上の条件のもとで次の定理が得られる。

[定理 2]　条件 1，2 を仮定する。このとき、解析関数

$$K(w) = \int \left[\frac{p(x|w)}{q(x)} - 1\right]^2 M(x)q(x)dx$$

のゼータ関数

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

の原点に一番近い極を $(-\lambda)$、その位数を m とすると、

$$F(X^n) = \lambda \log n - (m-1)\log\log n + R + o(1)$$

の漸近展開ができる。ここで R はある確率変数である。

(証明) W はコンパクトだから、W について局所的な議論を行えば十分である。$H(w)$ は解析関数とは限らないが無限回微分できる関数である。一方、$K(w)$ は解析関数なので、特異点解消定理を適用することができる。ある多様体 $\mathcal{M}$ から W の近傍への解析関数 $g : \mathcal{M} \ni u \mapsto w \in W$ が存在して、

$$K(g(u)) = u_1^{2k_1} \cdots u_d^{2k_d}$$

$$|g'(u)| = b(u)u_1^{h_1} \cdots u_d^{h_d}$$

とできる [9] [5]。ここで $|g'(u)|$ は $w = g(u)$ のヤコビアンで、$k_1, .., k_d, h_1, .., h_d$ は非負の整数、$b(u) > 0$ である。従って

$$a(x, u) = \frac{p(x|g(u))/q(x) - 1}{u_1^{k_1} \cdots u_d^{k_d}}$$

とおくと

$$\int a(x, u)^2 M(x)q(x)dx = 1$$

である。また定義より

$$\begin{aligned} H(w) &= \int q(x) \log \frac{q(x)}{p(x|w)} dx \\ &= \int q(x) \left[\frac{p(x|w)}{q(x)} - 1\right]^2 S\left(\frac{p(x|w)}{q(x)}\right) dx \\ &\leqq \int q(x) \left[\frac{p(x|w)}{q(x)} - 1\right]^2 M(x)dx \\ &= K(w) \end{aligned}$$

である。一方、コンパクト集合 C について

$$\begin{aligned} H(g(u)) &= \int q(x) \left[\frac{p(x|g(u))}{q(x)} - 1\right]^2 S\left(\frac{p(x|g(u))}{q(x)}\right) dx \\ &= \int q(x) \left[a(x, u)u_1^{k_1} \cdots u_d^{k_d}\right]^2 S\left(\frac{p(x|g(u))}{q(x)}\right) dx \\ &\geqq u_1^{2k_1} \cdots u_d^{2k_d} \int_C q(x)a(x, u)^2 S\left(\frac{p(x|g(u))}{q(x)}\right) dx \end{aligned}$$

ここでコンパクト集合 C を十分大きくとれば、定数 $\alpha > 0$ が存在して

$$H(g(u)) \geqq \alpha u_1^{2k_1} \cdots u_d^{2k_d}$$

とできる。以上より、無限回微分可能な関数 $\alpha(u) > 0$ が存在して

$$H(g(u)) = \alpha(u)u_1^{2k_1} \cdots u_d^{2k_d}$$

となる。次に

$$\begin{aligned} nH_n(g(u)) &\equiv nH(g(u)) \\ &\quad + \sum_{i=1}^n \left[\log \frac{q(X_i)}{p(X_i|g(u))} - H(g(u))\right] \\ &= n\alpha(u)u_1^{2k_1} \cdots u_d^{2k_d} \\ &\quad + \sqrt{n}u_1^{k_1} \cdots u_d^{k_d} \xi_n(u) \end{aligned}$$

とかける。ここで経験過程

$$\begin{aligned} \xi_n(u) &\equiv \frac{1}{\sqrt{n}} \frac{1}{u_1^{k_1} \cdots u_d^{k_d}} \sum_{i=1}^n \left[\log \frac{q(X_i)}{p(X_i|g(u))} - H(g(u))\right] \\ &= \frac{1}{\sqrt{n}} \frac{1}{u_1^{k_1} \cdots u_d^{k_d}} \sum_{i=1}^n \Big[-\log(1 + a(X_i, u)u_1^{k_1} \cdots u_d^{k_d}) \\ &\qquad -\alpha(u)u_1^{2k_1} \cdots u_d^{2k_d}\Big] \end{aligned}$$

は、u について無限回微分可能であり、コンパクト集合上に

あるのでドンスカークラス[17] に属し、$\mathcal{M}$ 上の正規確率過程 $\xi(u)$ に法則収束する。以上から

$$\begin{aligned} F(X^n) &= -\log \int \exp(-nH_n(w))\varphi(w)dw \\ &= -\log \sum \int \exp(-nH_n(g(u)))\varphi(g(u))|g'(u)|du \\ &= -\log \sum \int \int \frac{1}{n}\delta(\frac{t}{n} - u_1^{2k_1}\cdots u_d^{2k_d}) \\ &\quad \times \exp(-\alpha(u)t + \xi_n(u)t^{1/2})\varphi(g(u))|g'(u)|du\, dt \end{aligned}$$

となる。(和 $\sum$ は多様体 $\mathcal{M}$ を座標ごとに分割して積分するときの座標の和である。$\mathcal{M}$ は一般に向き付け可能な多様体でない)。

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

の原点に一番極とその位数は、

$$\zeta_0(z) = \sum \int \left[u_1^{2k_1}\cdots u_d^{2k_d}\right]^z \varphi(g(u))|g'(u)|du$$

の原点に一番近い極とその位数と一致する。$\zeta_0(z)$ の極を原点に近い方から順に $(-\lambda_k)$ $(k = 1, 2, ...,)$ とし、その位数を m_k とすれば、

$$\begin{aligned} \mathcal{D}(u) &\equiv \frac{1}{n}\delta(\frac{t}{n} - u_1^{2k_1}\cdots u_d^{2k_d})|g'(u)| \\ &= \sum_{k=1}^{\infty}\sum_{m=1}^{m_k} c_{km} t^{\lambda_k - 1}(-\log t)^{m-1} D_{km}(u) \end{aligned}$$

となる（$D_{km}(u)$ はある超関数。具体的に書くこともできるが煩雑になる）ので、これを $F(X^n)$ の定義式に代入すると定理を得る。(証明終り)。

4. 混合正規分布

次に混合正規分布が定理２の二つの条件を満たすことを示す。条件１は、事前分布についての条件なので、事前分布をこれを満たすように選べばよい。条件２を示す。パラメータのとりうる集合（コンパクト）を W とする。

［補題 1］ 式 (1) で表される学習モデル $p(x|w)$ と式 (2) で表される $q(x)$ について、ある定数 $A_0, A_1, B_0, B_1 > 0$ が存在して、任意の $w \in W$、任意の $x \in \mathbf{R}^N$ について

$$e^{-A_0\|x\| - A_1} \leqq \frac{p(x|w)}{q(x)} \leqq e^{B_0\|x\| + B_1}$$

が成立つ。

(証明)

$$\begin{aligned} \frac{p(x|w)}{q(x)} &= \frac{\sum a_h \exp(-\|x - b_h\|^2/2)}{\sum a_h^* \exp(-\|x - b_h^*\|^2/2)} \\ &\leqq \frac{\sum a_h \exp(-\|x - b_h\|^2/2)}{a_1^* \exp(-\|x - b_1^*\|^2/2)} \\ &= \frac{1}{a_1^*}\sum \exp(-\{\|x - b_h\|^2 - \|x - b_1^*\|\}/2) \\ &= \sum \frac{\exp((\|b_h^*\|^2 - \|b_h\|^2)/2)}{a_1^*} \exp((b_h - b_1^*)\cdot x) \\ &= \left[\sum_h \frac{\exp((\|b_1^*\|^2 - \|b_h\|^2)/2)}{a_1^*}\right] \\ &\quad \times \exp(\left[\max_h \|b_h - b_1\|\right]\|x\|) \end{aligned}$$

より右側が得られる。左側は、逆数について同じ議論を行えばよい。(証明終り)

［補題 2］ 式 (1) で表される学習モデル $p(x|w)$ と式 (2) で表される $q(x)$ について、ある定数 $C_0, C_1 > 0$ が存在して、任意の $w \in W$、任意の $x \in \mathbf{R}^N$ について

$$S\left(\frac{p(x|w)}{q(x)}\right) \leqq C_0\|x\| + C_1$$

が成立つ。

(証明) $S(t)$ の定義式より、$S(t) > 0$ $(0 < t < \infty)$ は連続関数であり、$S(t)$ が発散するのは $t \to 0$ のときだけである。これより、ある定数 $D_0, D_1 > 0$ が存在して、

$$S(t) \leqq \max[-D_0 \log t, D_1]$$

である。これに補題１を適用すればよい。(証明終り)。

［定理 3］ 式 (1) で表される学習モデル $p(x|w)$ と式 (2) で表される $q(x)$ を考える。補題 1,2 の定数 $B_0, B_1, D_0, D_1 > 0$ を用いて

$$M(x) = C_0\|x\| + C_1$$

とおく。このとき、$p(x|w)/q(x) - 1$ は定理２の前提条件２を満たす。

(証明)

$$\begin{aligned} u(x) &= \frac{p(x|w)}{q(x)} - 1 \\ &= \frac{\sum_h a_h \exp(-\|b_h\|^2/2)\exp(b_h\cdot x)}{\sum_h a_h^* \exp(-\|b_h^*\|^2/2)\exp(b_h^*\cdot x)} - 1 \end{aligned}$$

とおく。また、これのテーラー展開を J 項まで取って

$$u_J(x, w) = \sum_{j=0}^{J} \frac{1}{j!}\frac{\sum_h a_h \exp(-\|b_h\|^2/2)(b_h\cdot x)^j}{\sum_h a_h^* \exp(-\|b_h^*\|^2/2)\exp(b_h^*\cdot x)} - 1$$

とおく。$u_J(x, w)$ は $w = (a_h, b_h)$ の多項式であり、x については条件２のヒルベルト空間に属するから、ヒルベルト空間に値を取る解析関数である。解析関数が一様収束すれば解析関数であるから、以下、$J \to \infty$ のとき $u_J(x, w) \to u(x, w)$ が $w \in W$ についてヒルベルト空間の位相で一様収束することを示す。

$$\begin{aligned} T_J &\equiv \sup_{w\in W} \|u(w) - u_J(w)\|^2 \\ &= \sup_{w\in W} \int q(x)M(x) \\ &\quad \left[\sum_{j>J} \frac{1}{j!}\frac{\sum_h a_h e^{-\|b_h\|^2/2}(b_h\cdot x)^j}{\sum_h a_h^* e^{-\|b_h^*\|^2/2} e^{b_h^*\cdot x}}\right]^2 dx \\ &\leqq \int q(x)M(x)\sum_{j>J}\frac{1}{j!} e^{B_0\|x\| + B_1}\|x\|^j dx \end{aligned}$$

$\to 0$

となる（ルベーグの収束定理を用いた）ので、定理を得る。（証明終り）

5. 学習係数

以上に述べたことから、次の幾つかのゼータ関数

$$\zeta_i(z)=\int [H_i(w)]^z\varphi(w)dw$$

の最も原点に近い極は全て等しく位数も等しい。

$$H_1(w) = \int q(x)M(x)[p(x|w)/q(x)-1]^2dx$$
$$H_2(w) = \int_C q(x)[p(x|w)/q(x)-1]^2dx$$
$$H_3(w) = \int_C q(x)[p(x|w)-q(x)]^2dx$$
$$H_4(w) = \int_C [p(x|w)-q(x)]^2dx$$

混合正規分布の学習係数 λ は、上のいずれかのゼータ関数の原点に近い極 (すべて共通である）を求めることによって得られる。特に

$$H_5(w)=\int [p(x|w)-q(x)]^2dx$$

とおくと、ある定数 $e_1,e_2>0$ が存在して

$$e_1H_4(w)\leqq H_5(w)\leqq e_2H_1(w)$$

であるから、$H_5(w)$ についてのゼータ関数について調べればよい。フーリエ変換を用いて

$$H_5(w)=\int\left|\sum a_h e^{ib_h\cdot t}-\sum a_h^* e^{ib_h^*\cdot t}\right|^2 e^{-\frac{\|t\|^2}{2}}dt$$

が得られる。$(b_h\cdot t)$ と $(b_h^*\cdot t)$ について展開すれば、

$$H_5(w) = \int e^{-\frac{\|t\|^2}{2}}dt \left|\sum_{j=1}^{\infty}\frac{i^j}{j!}\{\sum a_h(b_h\cdot t)^j-\sum a_h^*(b_h^*\cdot t)^j\}\right|^2$$

となる。$x=(x_1,..,x_d)$ と $y=(y_1,..,y_d)$ の多項式

$$p_k(x,y)=\sum_{h=1}^{d}x_hy_h^k$$

について、$k\geqq H+1$ の $p_k(x,y)$ はそれ未満の k の $p_k(x,y)$ によって生成される多項式イデアルに含まれる [20] [27] ことと、$t=(t_1,t_2,..,t_d)$ とするとき関数

$$\{t_1^{\alpha_1}\cdots t_d^{\alpha_d}\}$$

が独立であることとを用いると、$H_5(w)$ は次の $H_6(w)$ と等価である。

$$b_h=(b_{h,1},b_{h,2},...,b_{h,N})$$

と書く。

$$H_6=\sum_{k=1}^{H+H_0}\sum_{\alpha_1+\cdots+\alpha_N=k}\left(\sum_{h=1}^{H}a_h\prod_{j=1}^{N}b_{hj}^{\alpha_j}-\sum_{h=1}^{H_0}a_h^*\prod_{j=1}^{N}b^{*\alpha_j}_{hj}\right)^2$$

ここで、条件 $a_1+\cdots+a_H=1,\ a_1^*+\cdots+a_{H_0}^*=1$ および $a_1,..,a_H\geqq 0$ がある。このような多項式の特異点解消は、まだ得られていないが、最近、幾つかの考察がなされるようになってきた [2] [3] [4] [30] [31] [32] [33] [34]。ここで第 H 番目の

$$a_H=a_1^*+\cdots+a_{H_0}^*-a_1-\cdots-a_{H-1}$$

を代入して、$a_1-a_1^*,\cdots,a_{H_0}-a_{H_0}^*,a_{H_0+1},\cdots,a_{H-1}$ についてのブローアップを行うことで、学習係数の上界として

$$\lambda\leqq\frac{1}{2}(NH_0+H-H_0-1)$$

が得られる。

6. 考　　察

本論文ではカルバック情報量が解析関数でない場合を考えたが、条件１、条件２を満たさない学習モデルでは、定理が成立たない場合がある。例えば、学習モデルが

$$p(y|x,a)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(y-\exp(-\frac{x}{a}))^2)$$

であり、真の分布が

$$q(y|x)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}y^2)$$

であり、$q(x)$ が正規分布のとき、カルバック情報量は

$$H(a) = \frac{1}{2}\int\exp(-\frac{2x}{a}))q(x)dx = \frac{\sqrt{2\pi}}{2}\exp(-\frac{1}{a^2})$$

となる。

$$H(a)=0\Longleftrightarrow a=0$$

であり、$H(a)$ は $a=0$ で無限回微分可能であるが、

$$c_1H_1(a)\leqq H(a)\leqq c_2H_1(a)$$

をみたすような定数 $c_1,c_2>0$ と多項式 $H_1(a)$ は存在しない。この例は条件２を満たさない。

混合正規分布は、学習モデル $p(x|w)$ は w について解析的であるが、対数関数 $\log(\ )$ の収束半径が有限であるために、カルバック情報量の解析性が失われてしまう。一方、上記であげた例では、学習モデル $p(x|w)$ そのものが解析的でないので、本質的に非解析的である。この例は必ずしも自然なものではないが、例えば、

$$Y=\frac{1}{1+e^{-x/a}}+\text{雑音}$$

によって、真の分布

$$Y=\begin{cases}1 & (x>0)\\ 0 & (\text{otherwise})\end{cases}+\text{雑音}$$

を学習する例は、やはり $a=0$ が解となるが、これは神経回路網の学習においては頻出する状況であり、今後は、このようなケースも重要な研究対象となるかもしれない。

7. 結　　論

混合正規分布のカルバック情報量が解析関数でない場合があることを示し、確率的複雑さの漸近展開を与える定理を証明した。混合正規分布の場合は、カルバック情報量が解析関数でなくても、解析関数であるときと同等の結果が成立つことがわかった。

謝　　辞

この研究は科学研究費補助金 1550030 を受けた。

文　　献

[1] Amari,S., Park,H., and Ozeki,T. ”Geometrical Singularities in the Neuromanifold of Multilayer Perceptrons,” *Advances in Neural Information Processing Systems*, 14, 1,2002.

[2] 青柳美輝，渡辺澄夫，“特異点解消と学習理論への応用,” 信技報,NC2003-26.pp.25-30,2003.

[3] M. Aoyagi, S. Watanabe, “The generalization error of reduced rank regression in Bayesian estimation,” to appear in International Symposium on Information Theory and its Applications, (Parma, Italy), 2004.

[4] 青柳美輝、渡辺澄夫、“縮小ランクモデルの汎化誤差と特異点解消,” In this Issue.

[5] Atiyah, M. F. ”Resolution of singularities and division of distributions.,” Communications of Pure and Applied Mathematics, 13, 145-150,1970.

[6] Dacunha-Castelle, D., Gassiat, E. ”Testing in locally conic models, and application to mixture models,” Probability and Statistics, 1, 285-317,1997.

[7] Hagiwara,K.”On the problem in model selection of neural network regression in overrealizable scenario,” *Neural Computation*,14,1979-2002.

[8] Hartigan, J. A. ”A failure of likelihood asymptotics for normal mixtures.,” Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, 2, 807-810,1985.

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

[10] Kashiwara, M. ”B-functions and holonomic systems. Inventiones Mathematicae,” 38, 33-53,1976.

[11] 高橋克之、渡辺澄夫、”特異的な学習モデルにおける MCMC 法の評価,” 2003 年情報論的学習理論ワークショップ,21003.

[12] 中島伸一、渡辺澄夫,”ハイパーパラメータ最適化法における汎化誤差について,” In this issue.

[13] 永田賢二、渡辺澄夫、”カルバック情報量の分割による特異モデルの学習係数の計算アルゴリズム” 情報論的学習理論 (IBIS) 、pp.41-46 、2003.

[14] 中野 修弘、高橋 克之、渡辺 澄夫,” 特異的な学習モデルにおけるメトロポリス法の挙動について ,” ニューロコンピューティング研究会,NC2003-191,pp.157-162,2004.3.18

[15] 西上功一郎, 渡辺澄夫, “特異な学習モデルの選択における事前分布の影響について,” 電子情報通信学会誌, Vol.J86-D-II,No.1,pp.119-129, 2003.

[16] Rusakov, D. & Geiger, D. ”Asymptotic model selection for naive Bayesian networks,” *Proceedings of UAI'02*, 2002.

[17] van der Vaart, A.W,&Weller,J ”Weak Convergence and Empirical Processes,” Springer,1996.

[18] 渡辺一帆、渡辺澄夫、“縮小ランク回帰モデルのベイズ汎化誤差について,” 電子情報通信学会誌, Vol.J86-A,No. 3, pp278-287,2003.

[19] Watanabe,S. ”A generalized Bayesian framework for neural networks with singular Fisher information matrices,”*International Symposium on Nonlinear Theory and Its Applications*,2,207-210,1995.l

[20] 渡辺澄夫, “ ベイズ法による階層型統計モデルの推定誤差について, ” 電子情報通信学会誌 A, Vol.J81-A, No.10, pp.1442-1452, 1998.

[21] Watanabe, S. ”Algebraic analysis for singular statistical estimation,” Lecture Notes in Computer Science, 1720, 39-50,1999.

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

[23] S. Watanabe, “Algebraic geometrical methods for hierarchical learning machines,” Neural Networks, Vol.14, No.8, pp. 1409–1060, 2001.

[24] 渡辺澄夫, “代数的な特異点をもつ学習モデルの学習誤差と汎化誤差, ” 信学論, Vol.J84-A, No.1, pp. 99–108, 2001.

[25] S. Watanabe, “Algebraic information geometry for learning machines with singularities,” Advances in Neural Information Processing Systems, Vol.13, pp. 329–336, 2001.

[26] 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.16, No.2, pp.308-315, March, 2001. (招待論文）.

[27] S. Watanabe, ”Learning efficiency of redundant neural networks in Bayesian estimation,” IEEE Transactions on Neural Networks, 12(6),pp.1475-1486,2001.

[28] 渡辺澄夫,“特異モデルとベイズ学習,” 日本神経回路学会誌,Vol.10,No.4,pp.211-219,2003.

[29] S. Watanabe and S. Amari, “Learning coefficients of layered models when the true distribution mismatches the singularities” Neural Computation, Vol.15, No.5, pp. 1013–1033, 2003.

[30] K.Yamazaki, S.Watanabe,”A probabilistic algorithm to calculate the learning curves of hierarchical learning machines with singularities,” Trans. on IEICE, Vol.J85-D-II,No.3,pp.363-372,Mar. 2002.

[31] K.Yamazaki, S.Watanabe,“Singularities in mixture models and upper bounds of stochastic complexity.” International Journal of Neural Networks, Vol.16, No.7, pp.1029-1038,2003.

[32] K. Yamazaki, S.Watanabe, “Stochastic complexity of Bayesian networks,” Proc. of International Conference on Uncertainty in Artificial Intelligence, CD-ROM, 2003.

[33] K.Yamazaki, S.Watanabe,“Stochastic complexities of hidden Markov models,” Proc. of International Conference on Neural Networks for Signal Processing, 2003.

[34] 山崎啓介、青柳美輝、渡辺澄夫、“ニュートン図形を用いた確率的複雑さの解析法,” In this issue.