

学習モデルにおける補題の証明

On the Proof of Lemmata in the Statistical Model

鷺野朋広

TOMOHIRO WASHINO

甲南大学大学院 自然科学研究科 知能情報学専攻

GRADUATE SCHOOL OF NATURAL SCIENCE, KONAN UNIVERSITY *

高橋正

TADASHI TAKAHASHI

甲南大学 知能情報学部 知能情報学科

DEPARTMENT OF INTELLIGENCE AND INFORMATICS, KONAN UNIVERSITY †

Abstract

学習モデルを中間ユニット数が H の 3 層パーセプトロンとする。真の分布を実現するパラメータの集合は、活性化関数を級数展開させると、有限個の多項式の零点集合として表される。初めに、 $H = 2$ のとき、真の分布が中間ユニット数が 0 と 1 で実現されるときの集合を求める。次に、一般の H に対して、真の分布より少ない中間ユニット数 $H_0 < H$ で実現されるときの集合を考える。

Abstract

The statistical model is a three-layer neural network with H hidden units. (By using Taylor expansion of activation function, the set of parameters that true density function is realizable by a statistical model is a common zero points defined by finite polynomials. First, in this case of hidden units equals two, we calculate equations defining equation of the algebraic set. The algebraic set is the set of parameters such that true density function is realizable by a statistical model with zero or one hidden unit. Next, for the general case, we determine equations define the algebraic set. The algebraic set is the set of parameters that true density function is realizable by a statistical model with H_0 is less than H hidden unit.

1 はじめに

階層構造を持ち、非線形なニューラルネットワークは正則モデルとは異なり特異モデルと呼ばれる。特異モデルでは、真の分布を実現するパラメータの集合が解析的集合をなす。真の分布より学習モデルが少ないユニット数で実現できる場合、真の分布を実現するパラメータが次元を持つ広がりとなり特異点を持つ。このとき、集合の定義方程式から定まるヤコビ行列の階数がパラメータによって異なり、特異点では行列式が 0 になる。ベイズ推定では、渡辺によって特異点解消定理を用いて汎化誤差の挙動が調べられている [1]。

損失関数を最小にする過程を学習と呼ぶ。特異モデルでは、最適なパラメータが 1 点には定まらず臨界点を持つ。臨界点 (鞍点, 極小点, 極大点) にとらわれて、学習が停滞し、過学習が起きる原因となっている [2]。

*d1523001@s.konan-u.ac.jp

†takahasi@konan-u.ac.jp

2 ヒルベルトの基底定理を応用した補題の証明

2.1 グレブナ基底と消去イデアル

定義 1

体 k に対し, $f \in k[x_1, x_2, \dots, x_d]$ をゼロでない多項式とする. f に現れる単項式のなかで単項式順序 $<$ に関して最大のものを先頭項とよび, $LT(f)$ と表す.

定義 2

単項式順序を固定する. イデアル I の有限部分集合 $G = \{g_1, \dots, g_t\}$ がグレブナ基底であるとは,

$$\langle LT(g_1), \dots, LT(g_t) \rangle = \langle LT(I) \rangle$$

を満たすことと定める.

定義 3

イデアル $I = \langle f_1, \dots, f_s \rangle \subset k[x_1, x_2, \dots, x_n]$ に対して, l 次の消去イデアル I_l であるとは,

$$I_l := I \cap k[x_{l+1}, \dots, x_d]$$

で定まる $k[x_{l+1}, \dots, x_d]$ のイデアルである.

定理 4 (消去定理)

$I \subset k[x_1, x_2, \dots, x_n]$ をイデアルとし, G を I の単項式順序 $<$ に関するグレブナ基底とする.

$0 \leq l \leq n$ に対して, 集合

$$G_l := G \cap k[x_{l+1}, \dots, x_d]$$

は l 次の消去イデアル I_l のグレブナ基底である.

2.2 4変数の場合

定理 5 (ヒルベルトの基底定理)

すべてのイデアル $I \subset k[x_1, x_2, \dots, x_n]$ は有限個の生成集合をもつ.

補題 6

4変数多項式 $g_k(a, b, c, d) := ab^{2k+1} + cd^{2k+1}$ に対して, $g_2, g_3 \in \langle g_0, g_1 \rangle$ が成り立つ.

証明 Mathematica を用いて

$$f1 = g0 - ab - cd; f2 = g1 - ab^3 - cd^3; f3 = g2 - ab^5 - cd^5;$$
$$\text{GroebnerBasis}[\{f1, f2, f3\}, \{a, b, c, d\}, \text{MonomialOrder} \rightarrow \text{Lexicographic}]$$

と入力すると, グレブナ基底が出力される. a, c を消去した消去イデアルを求めるため,

$$\text{GroebnerBasis}[\{f1, f2, f3\}, \{a, b, c, d\}, \{a, c\}, \text{MonomialOrder} \rightarrow \text{Lexicographic}]$$

と入力すると, $b^2d^2g0 - b^2g1 - d^2g1 + g2$ と出力されるため, $g_2 \in \langle g_0, g_1 \rangle$ である.

$$f4 = g3 - ab^7 - cd^7;$$
$$\text{GroebnerBasis}[\{f1, f2, f3, f4\}, \{a, b, c, d\}, \text{MonomialOrder} \rightarrow \text{Lexicographic}]$$

と入力すると、グレブナ基底が出力される。 a, c を消去した消去イデアルを求めるため、

`GroebnerBasis[{f1, f2, f3, f4}, {a, b, c, d}, {a, c, g2}, MonomialOrder -> Lexicographic]`

と入力すると、 $b^4d^2g_0 + b^2d^4g_0 - b^4g_1 - b^2d^2g_1 - d^4g_1 + g_3$ と出力されるため、 $g_3 \in \langle g_0, g_1 \rangle$ である。 ■

補題 7

4変数多項式 $g_k(a, b, c, d) := ab^{2k+1} + cd^{2k+1}$ に対して、イデアル $I_k = \langle g_0, g_1, g_2, \dots, g_k \rangle$ とする。このときイデアルの増大列 $I_0 \subset I_1 \subset I_2 \subset \dots$ は止まり、任意の $k \geq 1$ に対して、 $I_1 = I_k$ が成り立つ。

証明 Mathematica を用いて $g_2, g_3 \in \langle g_0, g_1 \rangle$ が成り立つ。

また、次の漸化式が成り立つ。

$$g_{k+1} = g_1(b^{2k} + d^{2k}) - g_0(b^2d^{2k} + b^{2k}d^2) + (b^2d^2)g_{k-1}.$$

よって任意の $k \geq 1$ に対して、 $g_k \in \langle g_0, g_1 \rangle$ が成り立つ。 ■

2.3 6変数の場合

補題 8

6変数多項式 $g_k(a, b, c, d, e, f) := ab^{2k+1} + cd^{2k+1} + ef^{2k+1}$ に対して、 $g_3, g_4, g_5 \in \langle g_0, g_1, g_2 \rangle$ が成り立つ。

証明 Mathematica を用いて a, b, c, d, e, h で基底を表すと計算が止まらなかったため、4変数の基底を用いて g_0, g_1, e, h で基底を表す。

$$f1 = g[0] - g_0 - eh; f2 = g[1] - g_1 - eh^3;$$

$$f3 = g[2] + b^2d^2g_0 - (b^2 + d^2)g_1 - eh^5; f4 = g[3] + (b^4d^2 + b^2d^4)g_0 - (b^4 + b^2d^2 + d^4)g_1 - eh^7;$$

`GroebnerBasis[{f1, f2, f3, f4}, {b, d, e, h}, MonomialOrder -> Lexicographic]`

と入力すると、グレブナ基底が出力される。 e, g_0, g_1 を消去した消去イデアルを求めるため、

`GroebnerBasis[{f1, f2, f3, f4}, {b, d, h}, {e, g0, g1}, MonomialOrder -> Lexicographic]`

と入力すると、次の式が出力されるため、 $g_3 \in \langle g_0, g_1, g_2 \rangle$ である。

$$b^2d^2h^2g[0] - b^2d^2g[1] - b^2h^2g[1] - d^2h^2g[1] + b^2g[2] + d^2g[2] + h^2g[2] - g[3].$$

$$f5 = g[4] + (b^6d^2 + b^4d^4 + b^2d^6)g_0 - (b^6 + b^4d^2 + b^2d^4 + d^6)g_1 - eh^9;$$

`GroebnerBasis[{f1, f2, f3, f4, f5}, {b, d, e, h}, MonomialOrder -> Lexicographic]`

と入力すると、グレブナ基底が出力される。 $e, g_0, g_1, g[3]$ を消去した消去イデアルを求めるため、

`GroebnerBasis[{f1, f2, f3, f4, f5}, {b, d, h}, {e, g0, g1, g[3]}, MonomialOrder -> Lexicographic]`

と入力すると、次の式が出力されるため、 $g_4 \in \langle g_0, g_1, g_2 \rangle$ である。

$$b^4d^2h^2g[0] + b^2d^4h^2g[0] + b^2d^2h^4g[0] - b^4d^2g[1] - b^2d^4g[1] - b^4h^2g[1] - 2b^2d^2h^2g[1] - d^4h^2g[1] - b^2h^4g[1] + b^4g[2] + b^2d^2g[2] + d^4g[2] + b^2h^2g[2] + d^2h^2g[2] + h^4g[2] - g[4].$$

$g[4] \in \langle g[0], g[1], g[2] \rangle$ より, $\langle g[0], g[1], g[2], g[3], g[4], g[5] \rangle = \langle g[0], g[1], g[2], g[3], g[5] \rangle$ が成立するので $\langle g[0], g[1], g[2], g[3], g[5] \rangle = \langle g[0], g[1], g[2] \rangle$ が成り立つことを調べる.

$$f6 = g[5] + (b^8 d^2 + b^6 d^4 + b^4 d^6 + b^2 d^8)g0 - (b^8 + b^6 d^2 + b^4 d^4 + b^2 d^6 + d^8)g1 - eh^{11};$$

$$\text{GroebnerBasis}[\{f1, f2, f3, f4, f6\}, \{b, d, e, h\}, \text{MonomialOrder} \rightarrow \text{Lexicographic}]$$

と入力すると, グレブナ基底が出力される. $e, g0, g1, g[3], g[4]$ を消去した消去イデアルを求めるため,

$$\text{GroebnerBasis}[\{f1, f2, f3, f4, f6\}, \{b, d, h\}, \{e, g0, g1, g[3], g[4]\}, \text{MonomialOrder} \rightarrow \text{Lexicographic}]$$

と入力すると, 次の式が出力されるため, $g_5 \in \langle g_0, g_1, g_2 \rangle$ である.

$$b^6 d^2 h^2 g[0] + b^4 d^4 h^2 g[0] + b^2 d^6 h^2 g[0] + b^4 d^2 h^4 g[0] + b^2 d^4 h^4 g[0] + b^2 d^2 h^6 g[0] -$$

$$b^6 d^2 g[1] - b^4 d^4 g[1] - b^2 d^6 g[1] - b^6 h^2 g[1] - 2b^4 d^2 h^2 g[1] - 2b^2 d^4 h^2 g[1] - d^6 h^2 g[1] -$$

$$b^4 h^4 g[1] - 2b^2 d^2 h^4 g[1] - d^4 h^4 g[1] - b^2 h^6 g[1] - d^2 h^6 g[1] + b^6 g[2] + b^4 d^2 g[2] + b^2 d^4 g[2] + d^6 g[2] +$$

$$b^4 h^2 g[2] + b^2 d^2 h^2 g[2] + d^4 h^2 g[2] + b^2 h^4 g[2] + d^2 h^4 g[2] + h^6 g[2] - g[5].$$

■

補題 9

6 変数多項式 $g_k(a, b, c, d, e, f) := ab^{2k+1} + cd^{2k+1} + ef^{2k+1}$ に対して, イデアル $I_k = \langle g_0, g_1, g_2, \dots, g_k \rangle$ とする. このときイデアルの増大列 $I_0 \subset I_1 \subset I_2 \subset \dots$ は止まり, 任意の $k \geq 2$ に対して, $I_2 = I_k$ が成り立つ.

証明 Mathematica を用いて $g_3, g_4, g_5 \in \langle g_0, g_1, g_2 \rangle$ が成り立つ.
次の漸化式が成り立つ.

$$g_{k+2} = g_2(b^{2k} + d^{2k} + f^{2k}) - g_1(b^2 d^{2k} + b^2 f^{2k} + b^{2k} d^2$$

$$+ d^2 f^{2k} + b^{2k} f^2 + d^{2k} f^2) + g_0(b^2 d^2 f^{2k} + b^2 d^{2k} f^2 + b^{2k} d^2 f^2)$$

$$+ (d^2 f^2 + b^2 f^2 + b^2 d^2)g_k - 2(b^2 d^2 f^2)g_{k-1}.$$

よって任意の $k \geq 2$ に対して $g_k \in \langle g_0, g_1, g_2 \rangle$ が成り立つ.

■

3 ヒルベルトの基底定理を応用した補題の拡張

補題 10

$2d$ 変数 $(a_1, \dots, a_d, b_1, \dots, b_d)$ 多項式 $g_n(a_1, \dots, a_d, b_1, \dots, b_d) := \sum_{k=1}^d a_k b_k^{2n+1}$ について, イデアル $I_k = \langle g_0, g_1, g_2, \dots, g_k \rangle$ とする. このときイデアルの増大列 $I_0 \subset I_1 \subset I_2 \subset \dots$ は止まり, 任意の $k \geq d-1$ に対して, $I_{d-1} = I_k$ が成り立つ.

証明 次の漸化式が成り立つ.

$$g_{k+d-1} = g_{d-1}(b_1^{2k} + b_2^{2k} + \dots + b_d^{2k}) - g_{d-2}(b_1^2 b_2^{2k} + b_1^2 b_3^{2k} \dots) + g_{d-3}(b_1^2 b_2^2 b_3^{2k} + \dots) + \dots$$

$$+ g_0(b_1^{2k} b_2^2 b_3^2 \dots b_d^2 + \dots + b_1^2 b_2^2 b_3^2 \dots b_d^{2k}) + (b_1^2 b_2^2 + b_1^2 b_3^2 \dots + b_{d-1}^2 b_d^2)g_{k+d-3}$$

$$- 2(b_1^2 b_2^2 b_3^2 + \dots)g_{k+d-4} + \dots + (-1)^d (d-1)(b_1^2 b_2^2 b_3^2 \dots b_d^2)g_{k-1}.$$

よって任意の $k \geq d-1$ に対して, $g_k \in \langle g_0, g_1, g_2, \dots, g_{d-1} \rangle$ が成り立つ [3].

■

4 真の分布と学習モデル

定義 11

\mathbb{R}^1 に値をとる入力 X と出力 Y が同時確率密度関数 $q(x, y)$ に従う確率変数とする. 確率分布 $q(x, y)$ を真の分布という.

定義 12

入力 X が従う確率分布 $q(x)$ と, パラメータ w をもつ \mathbb{R}^1 から \mathbb{R}^1 への関数 $f(x, w)$ に対して確率密度関数

$$p(y|x, w) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{|y - f(x, w)|^2}{2}\right)$$

を用いて, 学習モデル $p(x, y|w)$ を次で定める.

$$p(x, y|w) := q(x)p(y|x, w).$$

真の分布を学習モデルが実現する場合, パラメータ w_0 が存在して, 次が成立する.

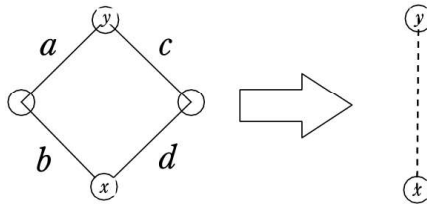
$$q(x, y) = p(x, y|w_0).$$

5 真の分布の中間ユニット数が $H_0 = 0$

学習モデルを入力ユニット数 1, 中間ユニット数 $H = 2$, 出力ユニット数 1, 活性化関数 $\tanh(x)$ である 3 層パーセプトロンとし, 真の分布を中間ユニット数 $H_0 = 0$ である 3 層パーセプトロンとする. $f(x, w)$ と真の分布を次で定める.

$$f(x, w) = a \tanh(bx) + c \tanh(dx), \quad q(x, y) = \frac{q(x)}{\sqrt{2\pi}} \exp\left(-\frac{|y|^2}{2}\right).$$

図 1 は真の分布が $H_0 = 0$ で実現される場合を表す.



真の分布が $H_0 = 0$ で実現される場合

図 1: $H = 2 \Rightarrow H_0 = 0$ の場合

5.1 真の分布を実現するパラメータの集合

学習モデルと真の分布が等しくなるパラメータの解析的集合を考える.

$$W_0 := \{w \in \mathbb{R}^4 | p(x, y|w) = q(x, y)\} = \{w \in \mathbb{R}^4 | a \tanh(bx) + c \tanh(dx) = 0\}.$$

$\tanh(x)$ を級数展開すると、定義方程式は次で表される。

$$\sum_{k=0}^{\infty} \frac{2^{2k+2}(2^{2k+2}-1)B_{2k+2}x^{2k+1}}{(2k+2)!} (ab^{2k+1} + cd^{2k+1}).$$

ここで、 B_n はベルヌーイ数である。 $\{x^{2k+1}\}$ は一次独立であるから、 W_0 は $g_k(a, b, c, d) := ab^{2k+1} + cd^{2k+1}$ と定めると、無限個の多項式の共通零点である。

$$W_0 = \{w \in \mathbb{R}^4 | g_0 = g_1 = \dots = 0\}.$$

イデアル $I_k = \langle g_0, g_1, g_2, \dots, g_k \rangle$ とすると、イデアルの増大列 $I_0 \subset I_1 \subset I_2 \subset I_3 \subset \dots$ は止まる。補題 7 より次が成り立つ。

$$V(I_k) = \{w \in \mathbb{R}^4 | g_0 = g_1 = \dots = g_k = 0\} = \{w \in \mathbb{R}^4 | ab + cd = ab^3 + cd^3 = 0\}.$$

5.2 代数的集合の計算

Mathematica を用いて $f1 = ab + cd; f2 = ab^3 + cd^3$; と入力する。
 a を消去した消去イデアルを求めるため、次のように入力する。

$$\text{GroebnerBasis}[\{f1, f2\}, \{a, b, c, d\}, \{a\}, \text{MonomialOrder} \rightarrow \text{Lexicographic}].$$

$b^2cd - cd^3$ と出力され、 $b^2cd - cd^3 = cd(b^2 - d^2) = 0$ が成り立つ。

よって $c = 0$ 又は $d = 0$ 又は $b = \pm d$ である。

(i) $c = 0$ のとき、 $ab = 0$ より、 $(0, b, 0, d)$, $(a, 0, 0, d)$ 。

(ii) $d = 0$ のとき、 $ab = 0$ より、 $(0, b, c, 0)$, $(a, 0, c, 0)$ 。

(iii) $b = \pm d$ のとき、 $(\pm a + c)d = 0$ 。よって $\pm a + c = 0$ 又は $d = 0$ である。

$\pm a + c = 0$ のとき、 $(a, \pm d, \mp a, d)$ (複合同順)。 $d = 0$ のとき、 $(a, 0, c, 0)$ 。

真の分布を実現するパラメータの集合は

$(a, 0, 0, d)$, $(a, 0, c, 0)$, $(0, b, 0, d)$, $(0, b, c, 0)$, $(a, \pm d, \mp a, d)$ (複合同順) である。

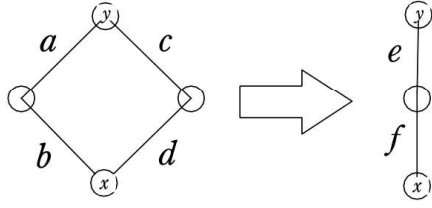
6 真の分布を表す中間ユニット数が $H_0 = 1$

学習モデルを入力ユニット数 1, 中間ユニット数 $H = 2$, 出力ユニット数 1, 活性化関数 $\tanh(x)$ である 3 層パーセプトロンとし、真の分布を中間ユニット数 $H_0 = 1$ である 3 層パーセプトロンとする。

$f(x, w)$ と真の分布を次で定める。

$$f(x, w) = a \tanh(bx) + c \tanh(dx), \quad q(x, y) = \frac{q(x)}{\sqrt{2\pi}} \exp\left(-\frac{|y - e \tanh(fx)|^2}{2}\right).$$

図 2 は真の分布が $H_0 = 1$ で実現される場合を表す。



真の分布が $H_0=1$ で実現される場合

図 2: $H = 2 \Rightarrow H_0 = 1$ の場合

6.1 真の分布を実現するパラメータの集合

学習モデルと真の分布が等しくなるパラメータの解析的集合を考える.

$$W_0 := \{w \in \mathbb{R}^4 | p(x, y|w) = q(x, y)\} = \{w \in \mathbb{R}^4 | a \tanh(bx) + c \tanh(dx) = e \tanh(fx)\}.$$

$\tanh(x)$ を級数展開して, $g_k(a, b, c, d, e, f) := ab^{2k+1} + cd^{2k+1} - ef^{2k+1}$ と定めると, 定義方程式は次で表される.

$$\sum_{k=0}^{\infty} \frac{2^{2k+2}(2^{2k+2} - 1)B_{2k+2}x^{2k+1}}{(2k+2)!} g_k(a, b, c, d, e, f).$$

ここで B_n はベルヌーイ数である. $\{x^{2k+1}\}$ は一次独立であるから, W_0 は無限個の多項式の共通零点である.

$$W_0 := \{w \in \mathbb{R}^4 | g_0 = g_1 = \dots = 0\}.$$

イデアル $I_k = \langle g_0, g_1, g_2, \dots, g_k \rangle$ とすると, イデアルの増大列 $I_0 \subset I_1 \subset I_2 \subset I_3 \subset \dots$ は止まる. 補題 9 より次が成り立つ.

$$V(I_k) = \{w \in \mathbb{R}^4 | g_0 = g_1 = \dots = g_k = 0\} = \{w \in \mathbb{R}^4 | ab + cd - ef = ab^3 + cd^3 - ef^3 = ab^5 + cd^5 - ef^5 = 0\}.$$

6.2 代数的集合の計算

Mathematica を用いて $f_1 = ab + cd - ef$; $f_2 = ab^3 + cd^3 - ef^3$; $f_3 = ab^5 + cd^5 - ef^5$; と入力し, a, c を消去した消去イデアルを求めるため, 次のように入力する.

$$\text{GroebnerBasis}[\{f_1, f_2, f_3\}, \{a, c, e, b, d, f\}, \{a, c\}, \text{MonomialOrder} \rightarrow \text{Lexicographic}].$$

$b^2d^2ef - b^2ef^3 - d^2ef^3 + ef^5$ と出力され, $b^2d^2ef - b^2ef^3 - d^2ef^3 + ef^5 = ef(f^2 - b^2)(f^2 - d^2) = 0$ が成り立つ. このとき $b = \pm f$ 又は $d = \pm f$ である.

(i) $b = f$ のとき, $f_1 = (a - e)f + cd$; $f_2 = (a - e)f^3 + cd^3$; $f_3 = (a - e)f^5 + cd^5$; と入力し, a を消去した消去イデアルを求めるため, 次のように入力する.

$$\text{GroebnerBasis}[\{f_1, f_2, f_3\}, \{a, c, e, d, f\}, \{a\}, \text{MonomialOrder} \rightarrow \text{Lexicographic}].$$

$cd^3 - cdf^2$ と出力され, $cd(f^2 - d^2) = 0$ が成り立つ.

(ii) $b = -f$ のとき, $f_1 = (-a - e)f + cd$; $f_2 = (-a - e)f^3 + cd^3$; $f_3 = (-a - e)f^5 + cd^5$; と入力し, a を消去した消去イデアルを求めるため, 次のように入力する.

$$\text{GroebnerBasis}[\{f_1, f_2, f_3\}, \{a, c, e, d, f\}, \{a\}, \text{MonomialOrder} \rightarrow \text{Lexicographic}].$$

$cd^3 - cdf^2$ と出力され, $cd(f^2 - d^2) = 0$ が成り立つ.

よって (i),(ii) より, $b = \pm f$ のとき, $cd(f^2 - d^2) = 0$ が成り立つ.

このとき $c = 0$ 又は $d = 0$ 又は $d = \pm f$ である.

$c = 0$ 又は $d = 0$ のとき, $(\pm a - e)f = 0$ であり, $(\pm e, \pm f, 0, d), (\pm e, \pm f, c, 0)$ (複合同順).

$d = \pm f$ のとき, $(\pm a \pm c - e)f = 0$ (複合同意) であり, $\pm a \pm c = e$ のとき,

$(\lambda e, \pm f, \pm(1 \mp \lambda)e, \pm f)$

(第2成分と第3成分の λ , 第4成分と第3成分が複合同順, 第2成分と第4成分が複合同意).

(iii) $d = \pm f$ のとき, (a, b) と (c, d) は対称性があるので (i),(ii) より,

$(\lambda e, \pm f, \pm(1 \mp \lambda)e, \pm f)$

(第2成分と第3成分の λ , 第3成分と第4成分が複合同順, 第2成分と第4成分が複合同意),

$(0, b, \pm e, \pm f), (a, 0, \pm e, \pm f)$ (複合同順).

真の分布を実現するパラメータの集合は

$(a, 0, \pm e, \pm f), (0, b, \pm e, \pm f), (\pm e, \pm f, c, 0), (\pm e, \pm f, 0, d)$ (複合同順),

$(\lambda e, \pm f, \pm(1 \mp \lambda)e, \pm f)$

(第2成分と第3成分の λ , 第4成分と第3成分が複合同順, 第2成分と第4成分が複合同意).

7 真の分布を表す中間ユニット数が $H_0 = m$

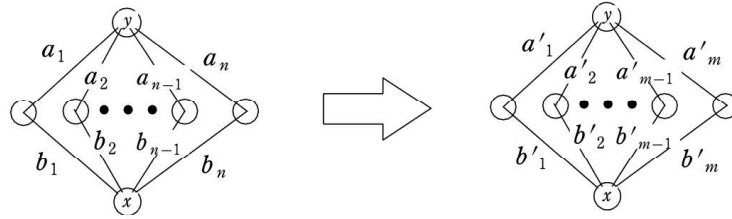
学習モデルを入力ユニット数 1, 中間ユニット数 $H = n$, 出力ユニット数 1, 活性化関数 $\tanh(x)$ である 3 層パーセプトロンとし, 真の分布を中間ユニット数 $H_0 = m$ である 3 層パーセプトロンとする.

$f(x, w)$ と真の分布を次で定める.

$$f(x, w) = a_1 \tanh(b_1 x) + \cdots + a_n \tanh(b_n x)$$

$$q(x, y) = \frac{q(x)}{\sqrt{2\pi}} \exp\left(-\frac{|y - a'_1 \tanh(b'_1 x) + \cdots + a'_m \tanh(b'_m x)|^2}{2}\right).$$

図 3 は真の分布が $H_0 = m$ で実現される場合を表す.



真の分布が $H_0 = m$ で実現される場合

図 3: $H = n \Rightarrow H_0 = m$ の場合

7.1 真の分布を実現するパラメータの集合

学習モデルと真の分布が等しくなるパラメータの解析的集合を考える.

$$W_0 := \{w \in \mathbb{R}^{2n} | p(x, y|w) = q(x, y)\}$$

$$= \{w \in \mathbb{R}^{2n} | a_1 \tanh(b_1 x) + \cdots + a_n \tanh(b_n x) = a'_1 \tanh(b'_1 x) + \cdots + a'_m \tanh(b'_m x)\}.$$

$\tanh(x)$ を級数展開して,

$$g_k(a_1, \dots, a_n, b_1, \dots, b_n, a'_1, \dots, a'_m, b'_1, \dots, b'_m) := \sum_{i=1}^n a_i b_i^{2k+1} - \sum_{i=1}^m a'_i b_i'^{2k+1}$$

と定めると, 定義方程式は次で表される.

$$\sum_{k=0}^{\infty} \frac{2^{2k+2}(2^{2k+2} - 1)B_{2k+2}x^{2k+1}}{(2k+2)!} g_k(a_1, \dots, a_n, b_1, \dots, b_n, a'_1, \dots, a'_m, b'_1, \dots, b'_m)$$

ここで B_n はベルヌーイ数である. $\{x^{2k+1}\}$ は一次独立であるから, W_0 は次の無限個の多項式の共通零点である.

$$W_0 := \{w \in \mathbb{R}^{2n} | g_0 = g_1 = \dots = 0\}.$$

ここでイデアル $I_k = \langle g_0, g_1, g_2, \dots, g_k \rangle$ とすると, イデアルの増大列 $I_0 \subset I_1 \subset I_2 \subset I_3 \subset \dots$ は止まる.

$n + m := d$ とすると, 補題 10 より次が成り立つ.

$$\begin{aligned} V(I_k) &= \{w \in \mathbb{R}^k | g_0 = g_1 = \dots = g_k = 0\} \\ &= \{w \in \mathbb{R}^{2n} | \sum_{i=1}^n a_i b_i - \sum_{i=1}^m a'_i b'_i = \sum_{i=1}^n a_i b_i^3 - \sum_{i=1}^m a'_i b_i'^3 = \dots = \sum_{i=1}^n a_i b_i^{2d-1} - \sum_{i=1}^m a'_i b_i'^{2d-1} = 0\}. \end{aligned}$$

7.2 代数的集合のパラメータ表示

パラメータ w を $w := \{a_1, b_1, a_2, b_2, \dots, a_n, b_n\}$ と表示すると次で表される.

$$w = \{a'_1, b'_1, \dots, a'_i, b'_i, \dots, a'_m, b'_m, a_0, b_0, a_1, b_1, \dots, a_i, b_i, \dots, a_h, b_h\}.$$

初めに b'_i ($1 \leq i \leq m$) に対して, $b_i^2 = b_j^2$ を満たす b_j ($1 \leq j \leq n$) の個数を l_j とする. $l_1 \geq l_2 \geq \dots \geq l_m \geq 1$ となるように $(a_i, b_i), (a'_i, b'_i)$ の順番を入れ替える. このとき $1 \leq i \leq m$ に対して, 次で表される.

$$\{a'_i, b'_i\} := \left\{ \lambda_i^1 a'_i, \pm b'_i, \lambda_i^2 a'_i, \pm b'_i, \dots, \lambda_i^{l_i-1} a'_i, \pm b'_i, \pm(1 \mp \lambda_i^1 \mp \lambda_i^2 \dots \mp \lambda_i^{l_i-1}) a'_i, \pm b'_i \right\}$$

(第 $2k$ 成分と第 $2l_i - 1$ 成分の λ_i^k , 第 $2l_i$ 成分と第 $2l_i - 1$ 成分が複合同順, 第 $2k$ 成分と第 $2l_i$ 成分が複合同意).

次に $\sum_{i=1}^m l_i := n'$ とする. n'' ($n' \leq n'' \leq n$) に対し, $a_i b_i = 0$ ($n' + 1 \leq i \leq n''$), $a_i b_i \neq 0$ ($n'' + 1 \leq i \leq n$) となるように (a_i, b_i) の順番を入れ替える. このとき次で表される.

$$\{a_0, b_0\} := \{a_{n'+1}, b_{n'+1}, \dots, a_i, b_i, \dots, a_{n''}, b_{n''}\}.$$

最後に b_i ($n'' + 1 \leq i \leq n$) に対して, $b_i^2 = b_j^2$ を満たす b_j ($n'' + 1 \leq j \leq n$) の個数を r_j とする. $r_1 \geq r_2 \geq \dots \geq r_h \geq 2$ となるように (a_i, b_i) の順番を入れ替える. $n'' + \sum_{k=1}^{i-1} r_k := j$, $n'' + \sum_{k=1}^h r_k := n$ とする. このとき $1 \leq i \leq h$ に対して, 次で表される.

$$\{a_i, b_i\} := \{a_{j+1}, \pm b_{j+r_i}, a_{j+2}, \pm b_{j+r_i}, \dots, a_{j+r_i-1}, \pm b_{j+r_i}, \mp a_{j+1} \mp a_{j+2} \dots \mp a_{j+r_i-1}, b_{j+r_i}\}$$

(第 $2k$ 成分と第 $2r_i - 1$ 成分の a_{j+k} が複合同順, 第 $2k$ 成分が複合同意).

参 考 文 献

- [1] 渡辺澄夫, 代数幾何学と学習理論. 森北出版株式会社, 2006.
- [2] K. Fukumizu and S. Amari, "Local minima and plateaus in hierarchical structures of multilayer perceptrons," *Neural Networks*, vol. 13, no. 3, pp. 317-327, 2000.
- [3] 鷺野朋広, ヒルベルトの基底定理を応用した補題の拡張, preprint.