

The Memoirs of Faculty of Engineering
Fukuyama University
The 19th issue, No.2, Mar., 1996

関数近似問題に対する階層型ニューラルネットワークの 内部表現と汎化能力の関係

渡辺 栄治*

Relations between Internal Representations and Generalization Ability of Multi Layered
Neural Networks for Function Approximation Problems

Eiji WATANABE*

ABSTRACT

This paper discusses the relation between internal representations and generalization ability of multi-layered neural networks for function approximation problems. Here, the *over learning* problem, which makes the generalization ability poor, is divided into the two problems; the *excessive degrees of freedom* and the *non-uniqueness of weights*. These problems are carefully discussed from the viewpoint of internal representation; weights and outputs of hidden units. First, the excessive degrees of freedom problem is discussed by introducing the entropy for weights. Next, the *non-uniqueness of weights* problem is also discussed by introducing the principal components analysis method. Finally, the relations between internal representations and generalization ability of multi-layered neural networks are concretely discussed by numerical results.

キーワード: 階層型ニューラルネットワーク, 関数近似問題, 内部表現, 汎化能力, 過剰学習

Keywords: Multi-layered neural network, function approximation problem, internal representation, generalization ability, over learning

1. まえがき

階層型ニューラルネットワーク (階層型 NN) が有する学習能力, 並列処理能力, および汎化能力等に多大な関心が寄せられ, 各種分野への応用が図られている. 特に, 汎化能力は工学的応用に際して重要な能力であり, 適切な汎化能力を獲得するネットワーク構造や学習法に関する研究が急務な課題となっている.

関数近似問題を対象とする場合, 離散的に与えられた学習データを精度良く近似することが重要であり, 未学習パターンに対する補間能力が汎化能力として定義される. しかしながら, 実データに含まれる観測雑音や学習データが十分に得られない場合, 学習回数の増加と共に学習誤差が減少する一方で, 未学習データに対する誤差が増加することが指摘されている. このような現象は“過剰学習”と呼ばれており, 過剰学習を回避するための手法が多数提案されている [1], [2], [3], [4]. 過剰学習を引き起こす要因として以下の2点が挙げられる. まず, 十分な中間層ユニット数を設定すれば, 3層

NN が任意の連続関数を任意の精度で近似できることが証明されている [5] が, 過剰な中間層ユニット数を設定した場合, 階層型 NN による近似関数が過度に複雑になり汎化能力が劣化することがある. 次に, 階層型 NN に対する代表的な学習法である誤差逆伝搬学習法 (BP 法) では, 荷重係数初期値に対する解の一意性が保証されておらず [6], 階層型 NN による近似関数形状のパラツキも重要な要因であると考えられる.

本論文では, 階層型 NN の汎化能力改善学習法を構築するための一つのアプローチとして, 関数近似問題に対する階層型 NN の内部表現と汎化能力の関係について検討する [7]. ここでは, 汎化能力の劣化原因である“過剰学習”問題を“過剰自由度”および“荷重係数の非一意性”の2種類の問題に分割し, 内部表現 (荷重係数値や中間層ユニット出力値) の観点からこれら問題について検討する. まず, 荷重係数に関するエントロピーに着目した情報量規範を導入することにより, “過剰自由度”問題に関して検討する. 次に, 主成分分析法に基づいた中間層上の線形化精度評価手法を導入することにより, “荷重係数の非一意性”問題に関して検討する.

*情報処理工学科

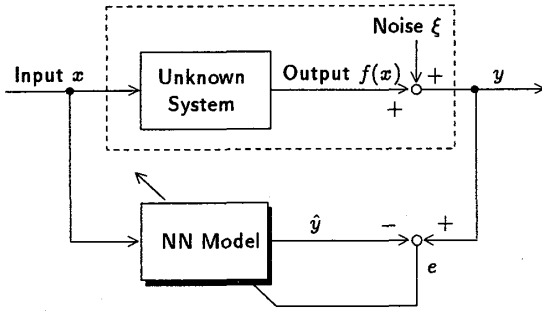


図 1: 階層型 NN による関数近似

最後に, “過剰学習” および “過剰自由度” の観点から階層型 NN の内部表現と汎化能力の関係について数値例を用いて具体的に検討する.

2. 階層型 NN による関数近似

本論文では, 静的な一入力一出力システムを対象とする. 入出力間の関係は未知なる連続な非線形関数 $f(x)$ により与えられ, 雑音 ξ が付加された出力 y が観測可能とする.

$$y_p = f(x_p) + \xi_p, \quad (p = 1, 2, \dots, P) \quad (1)$$

ここで, p はパターン番号を表し, 雑音 ξ_p は平均 0 および分散 σ^2 の統計量を有し, x_p と独立な確率変数とする. 雑音の分散 σ^2 および分布形状は任意とする. 本論文で扱う問題は, 学習データ (x_p, y_p) を用いて未知なる入出力関数 $f(x)$ を階層型 NN により推定することである. すなわち, 次式で定義される誤差関数 E を最小化することによりモデル $\hat{y} = \hat{f}(x)$ を構築することである (図 1).

$$E = \frac{1}{P} \sum_{p=1}^P (y_p - \hat{y}_p)^2 \quad (2)$$

階層型 NN において, p 番目の学習データに対する第 m 層第 i ユニットの入力 i_{pi}^m と出力 o_{pi}^m の関係は以下のように表される.

$$\begin{cases} i_{pi}^m = \sum_{j=1}^{n_{m-1}} w_{ij}^{m,m-1} o_{pj}^{m-1} + \theta_i^m \\ o_{pi}^m = f_u(i_{pi}^m) \end{cases} \quad (3)$$

ここで, $n_m (m = 1, 2, 3)$ は第 m 層内のユニット数を示す. また, $w_{ij}^{m,m-1}$ および θ_i^m は, 第 m 層第 i ユニットの第 $m-1$ 層第 j ユニットの間の荷重係数および第 m 層第 i ユニットのしきい値を表す. 更に, 入出力層ユニット特性関数 f_u には線形関数を, 中間層ユニット特性関

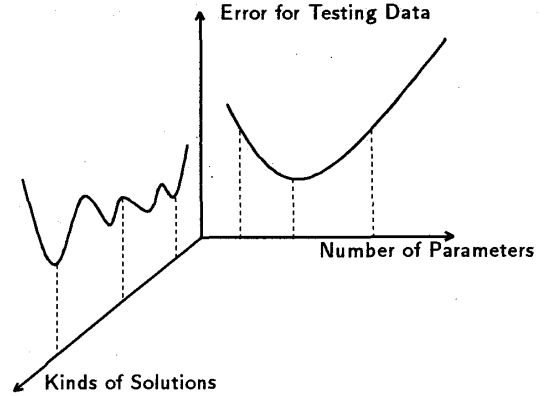


図 2: 関数近似問題における過剰自由度と解の非一意性

数 f_u には次式で与えられるシグモイド関数を用いる.

$$f_u(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

従って, 階層型 NN における入力 x_p および出力 $o_{p1}^3 (= \hat{y}_p)$ との関係は次式で与えられる.

$$o_{p1}^3 = \sum_{i=1}^{n_2} w_{1i}^{32} f_u(w_{i1}^{21} x_p + \theta_i^2) + \theta_1^3 \quad (5)$$

式 (5) より, 階層型 NN の入出力関係は $f_u(x)$ を基底関数とした非線形回帰とみなすことができる [8]. 従って, 誤差関数 E の最小化のもとで如何に適切な非線形変換 $f_u(x)$ を実現できるかが, 階層型 NN のモデリング精度に対して重要な要因であると考えられる.

3. 過剰学習

過剰学習とは, 学習回数の増加に伴い学習データに対する誤差が減少する反面, 未学習データに対する誤差が増加する現象を意味する. 次節以降では, 過剰学習の要因として階層型 NN の “過剰自由度” および “荷重係数の非一意性” に問題を絞り, それぞれ情報量規範および主成分分析法を導入することにより, 内部表現特性の観点から上述の問題点に対して定量的な指標を与える.

3.1 過剰自由度

十分な中間層ユニット数を設定すれば 3 層 NN が任意の連続関数を任意の精度で近似できることが証明されている [5]. しかし, 過剰な中間層ユニット数を設定した場合 (過剰自由度), 階層型 NN による近似関数が過度に複雑になり汎化能力が劣化することがある¹. 過

¹村田は ridge 関数をユニット特性関数として持つ 3 層 NN では, 多項式近似などで高次項の影響による近似関数の急激な複雑化から逃れることを示している [9].

剰自由度は以下の条件下で生じることが指摘されている [1].

$$P < n \quad (6)$$

ここで、 P は学習データ数を、 n はモデルパラメータ数を表す。過剰自由度を抑制することを目的として、冗長な中間層ユニットや荷重係数の段階的な削除手法が提案されている [10], [11]。一方、情報量規範 (Akaike Information Criterion, AIC) に基づいて適切なモデルを選択するための手法が検討されている [12], [13], [14], [15]。ここで、AIC は統計的モデルの適切さをモデル誤差とモデルパラメータ数とのトレードオフから評価するための規範であり、次式で定義される [16].

$$\begin{aligned} AIC = & -2(\text{モデルの最大対数尤度}) \\ & +2(\text{モデルのパラメータ数}) \end{aligned} \quad (7)$$

ここで、非線形モデルである階層型 NN に対して“モデルのパラメータ数”をどのように評価するかが重要な論点であると考えられる。

3.1.1 AIC におけるモデルパラメータ数

戸田らは、中間層ユニット数決定規範として AIC, BIC, MDLP の 3 種類中で AIC によって決定される中間層ユニット数が、他の規範に比べて比較的多く選択されることを示している [12]。ここでは、パラメータ数 K として中間層ユニット数 n_2 が採用されている。

$$K = n_2 \quad (8)$$

また、栗田は AIC を導入した中間層ユニット数決定手法を提案しており、パラメータ数として全荷重係数の個数を採用している [13].

$$K = n_2(n_1 + 1) + n_3(n_2 + 1) \quad (9)$$

更に、Ishikawa は $|w_{ij}^{m,m-1}| \geq \theta$ を満足する荷重係数のみを K として採用している [14] が、しきい値 θ により K が大きく変動する問題点を有している。

加えて、渡辺は形式的情報量基準と呼ばれる新たな基準を提案している [15]。この基準では、[14] と同様に実質的なパラメータ数を次式により評価している。

$$K = \sum_{i,j} f_\alpha(w_{ij}) \quad (10)$$

ここで、 $f_\alpha(x)$ は以下の 2 条件を満足する関数である。

$$\begin{cases} f_\alpha(x) \rightarrow f_0(x) & \text{for } \alpha \rightarrow 0 \\ 0 \leq f_\alpha(x) \leq f_\alpha(y) \leq 1 & \text{for } |x| \leq |y| \end{cases} \quad (11)$$

具体的には、 $1 - \exp(-x^2/2\alpha^2)$ などが用いられる。しかしながら、関数 $f_\alpha(x)$ の選び方により最終的に獲得される NN 構造が異なる等の問題点を有している。

3.1.2 エントロピーの適用によるモデルパラメータ数の算出

第 m 層第 i ユニット入力値 i_{pi}^m と第 $m-1$ 層第 j ユニット出力値 o_{pj}^{m-1} との関係が次式で表現されることから、

$$i_{pi}^m = \sum_{j=1}^{n_{m-1}} w_{ij}^{m,m-1} o_{pj}^{m-1} + \theta_i^m \quad (12)$$

荷重係数 $w_{ij}^{m,m-1}$ の絶対的な大きさだけでなく、相対的な大きさが信号伝搬の上で重要な役割を果たす。この観点から、Ishikawa は次式で定義されるエントロピー H を提案している [17].

$$H = - \sum_{i,j} P_{ij} \log_2 P_{ij} \quad (13)$$

但し、

$$P_{ij} = \frac{|w_{ij}^{m,m-1}|}{\sum_{i,j} |w_{ij}^{m,m-1}|} \quad (14)$$

H は荷重係数値を相対的な尺度で評価しており数理的な意味が明確になっている。しかしながら、入力層と中間層間および中間層と出力層間における情報伝搬機構に着目すると、前者ではシグモイド関数による非線形変換が後者では線形重ね合わせ処理が行われているものと解釈できることから、本論文では各層間における荷重係数値の相対的な関係を考慮し、各層ごとに以下のようなエントロピーを与える。

$$H^m = - \sum_{i,j} P_{ij}^m \log_2 P_{ij}^m \quad (15)$$

但し、

$$P_{ij}^m = \frac{|w_{ij}^{m,m-1}|}{\sum_{i,j} |w_{ij}^{m,m-1}|} \quad (16)$$

ここで、各層内における荷重係数値 $|w_{ij}^{m,m-1}|$ がすべて同じ値であればエントロピー H^m は最大値 $\log_2 n_{m-1} n_m$ を取り得る。 H^m は各層における信号の表現能力を表しており、 2^{H^m} が近似的に第 $m-1$ 層と第 m 層間の支配的な荷重係数個数と解釈できることから、次式の K をパラメータ数として提案する。

$$K = (2^{H^2} + 2^{H^3}) \quad (17)$$

ここで、パラメータ数 K の最大値は $n_2(n_1 + n_3)$ である。

3.2 荷重係数の非一意性

中間層ユニット数が過剰でなくとも、観測雑音や過大な荷重係数値の影響により過剰学習が生じ汎化能力の劣化を招くことが指摘されており、(i) 出力層ユニット

変動の最小化 [18], (ii) 平均曲率や微係数の抑制 [19], (iii) 荷重係数値和の抑制 [11], 等の内部表現に関する評価項を付加した汎化能力改善学習法が提案されている。更に, BP 法では大域的最小値を獲得することが保証されていないため, 汎化能力が荷重係数値に依存する問題点が生じる。初期値設定に関しては, 学習の高速化に関する設定法 [20] が提案されているが, 荷重係数値と汎化能力の関係については十分に検討されていない。また, 荷重係数値が学習プロセスの中で変動することから, 本論文では学習終了後の階層型 NN の内部表現と汎化能力の関係について検討する。

3.2.1 主成分分析の導入による内部表現の解析

中間層上に中間的な表現が生成されることが指摘されている [21] が, 中間層が有する数理的な意味については明確では無く, 学習後の内部表現を直感的に把握することは困難である。更に, 階層型 NN における荷重係数値の非一意性 [6] も内部表現の解析を困難にしている。一方, Sekita ら [22] は主成分分析により, また Tamura ら [23] は特異値分解により, 冗長な中間層ユニットを削除するための手法を提案している。本節では, 主成分分析を中間層ユニット出力に適用することにより, 内部表現の適切な特性について検討する。

まず, 学習誤差 e_p を次式のように書き換える。

$$\begin{aligned} e_p &= y_p - \hat{y}_p \\ &= y_p - \left(\sum_{j=1}^{n_2} w_{1j}^2 o_{pj}^2 + \theta_1^3 \right) \end{aligned} \quad (18)$$

ここで, 中間層ユニット出力 o_{pi}^2 と学習誤差 e_p とは線形関係にある。回帰分析において, モデルの妥当性を評価するために残差分析や AIC による次数選定等が用いられる。前者では, 残差の規則性を見つけるものであり, 各項の“線形化”精度を評価しているものと考えられる。従って, 適切な回帰モデルが得られたとすれば, そのモデルを構成する各項は線形の枠組の中で精度良く圧縮できるものと考えられる。

以下では, “線形化精度”の一評価手法として主成分分析の適用について検討する。今, o_{pi}^2 の共分散行列 A の要素 a_{ij} を次式で与える。

$$a_{ij} = \frac{1}{P} \sum_{p=1}^P (o_{pi}^2 - \bar{o}_i^2)(o_{pj}^2 - \bar{o}_j^2) \quad (19)$$

但し, $\bar{o}_i^2 = 1/P \sum_{p=1}^P o_{pi}^2$. 直交行列 Q を用いて共分散行列 A を対角化することにより, A の固有値 $\lambda_i (i = 1, 2, \dots, n_2)$ が求められる。

$$Q^T A Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{n_2}) \quad (20)$$

$\tilde{n}_2 (< n_2)$ 個の変量を主成分と見なした場合, 変量 o_{pi}^2 を合成変量 $\tilde{o}_{pi}^2 (i = 1, 2, \dots, \tilde{n}_2)$ に次元圧縮できる。従って, 変量 o_{pi}^2 が有する情報を固有値の総和 $\sum_{i=1}^{\tilde{n}_2} \lambda_i$ とみなすことにより, 次元圧縮による情報損失を次式で評価することができる。

$$I_{Loss} = \sum_{i=1}^{n_2} \lambda_i - \sum_{i=1}^{\tilde{n}_2} \frac{1}{P} \sum_{p=1}^P \{\tilde{o}_{pi}^2\}^2 \quad (21)$$

ここで, 右辺第 2 項は合成変量 \tilde{o}_{pi}^2 の分散和を表す。また, 情報損失 I_{Loss} を次式のように書き換える。

$$\begin{aligned} I_{Loss} &= \frac{1}{P} \sum_{p=1}^P I_{Loss}^p \\ &= \frac{1}{P} \sum_{p=1}^P \left\{ \sum_{i=1}^{n_2} \lambda_i - \sum_{i=1}^{\tilde{n}_2} \{\tilde{o}_{pi}^2\}^2 \right\} \end{aligned} \quad (22)$$

上式において, I_{Loss}^p は学習パターンごとの情報損失を評価している。従って, 各学習パターン間における I_{Loss}^p のバラツキが少ない程, 中間層ユニットが各学習パターンに対して均等に“知識”を獲得したものと解釈できる。

4. 数値計算

真のシステムは次式の非線形関数 $f(x)$ で表現されるものとする²。

$$f(x) = \frac{1}{1 + e^{-x}} + 1.8e^{-(x-1)^2} + 1.2e^{-(x+1)^2} \quad (23)$$

ここで, 入力パターン x_p は $[-2, 1]$ の範囲においてランダムに 100 個選び, 平均 0, 分散 0.25 のガウス乱数を雑音 ξ_p として用いた。学習データおよび非線形関数 $f(x)$ を図 3 に示す。また, 3 層 NN (線形-Sigmoid-線形) を用い, 各層のユニット数は $1 - n_2 - 1$ と設定した。更に, 学習係数 α を 0.01 と設定した。本論文では, 次式で表される補間誤差 E_{Inter} を関数近似問題における汎化能力として評価した。

$$E_{Inter} = \frac{1}{200} \sum_{p=1}^{200} (f(x_p) - \hat{y}_p)^2 \quad (24)$$

4.1 過剰自由度に関する検討

4.1.1 中間層ユニット数に関する検討

中間層ユニット数 n_2 を 1~20 まで変化させた場合の学習誤差および補間誤差を図 4 に示す。なお, 各中間層ユニット数ごとに 10 種類の荷重係数初期値を用い, 学習誤差および補間誤差は初期値に対する平均を表して

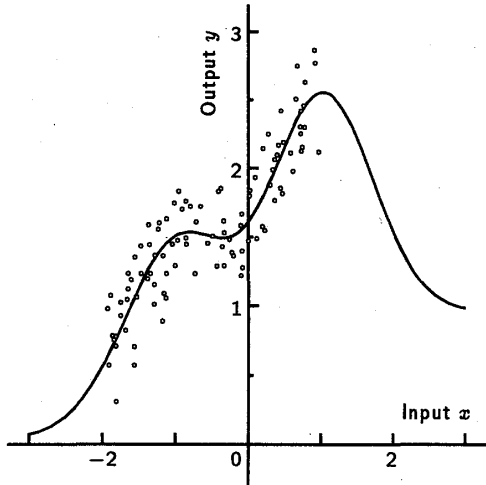


図 3: 学習データ

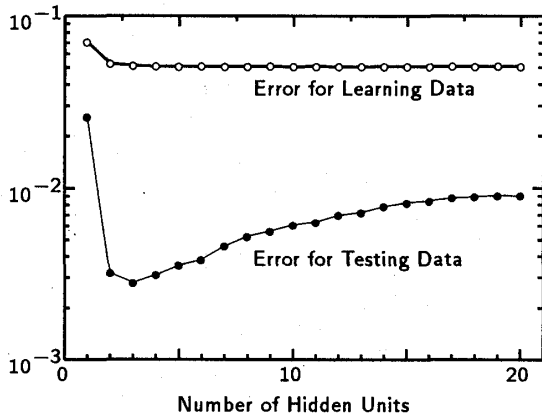


図 4: 中間層ユニット数に対する学習誤差と補間誤差の変動

いる。2個以上の中間層ユニット数に対して学習誤差がほぼ一定であるが、中間層ユニット数が4個以上の場合、補間誤差が増加しており過剰自由度の影響により汎化能力が劣化していることがわかる。

次に、式(17)で定義したモデルパラメータ数を用いた AIC の変動を図 5 に示す。但し、対数尤度の代わりに学習データに対する誤差を用いた。図 5 より、中間層ユニット数が 2 個のとき AIC が最小となっていることがわかる。図 4 および本図との比較において、補間誤差および AIC が最小となる中間層ユニット数が若干異なっているが、式(17)で定義した指標 K の有効性が示されている。

²ここで、ここで取り上げた非線形関数 $f(x)$ は連続かつ変曲点を有することから、3層階層型 NN により十分に近似可能な関数である。

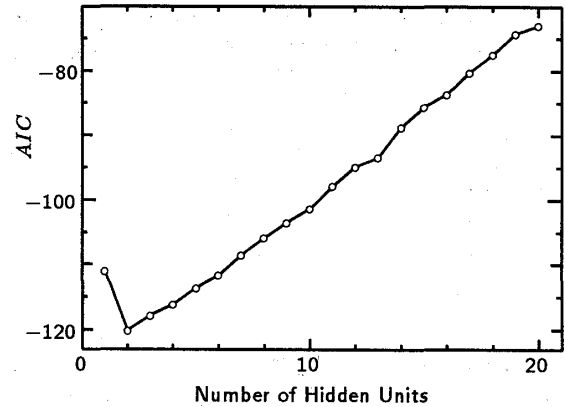


図 5: 中間層ユニット数に対する AIC の変動

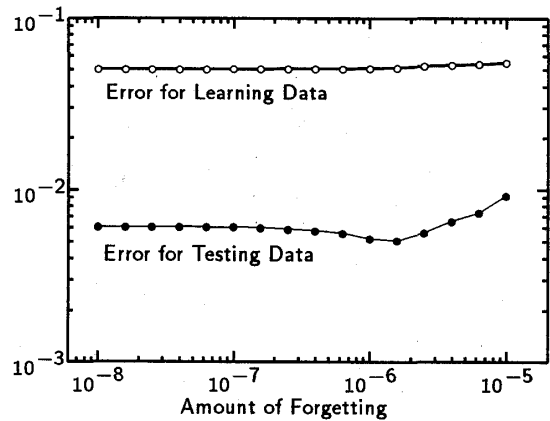


図 6: 忘却量に対する学習誤差と補間誤差の変動

4.1.2 構造化による影響

更に、過剰自由度を低減化するための一学習法として忘却付き学習法 [11] を用いて、階層型 NN の構造化による影響について検討する。忘却量 ϵ を $10^{-8} \sim 10^{-5}$ まで変化させた場合の学習誤差および補間誤差を図 6 に示す。ここで、初期構造として 10 個の中間層ユニット数を設定した。なお、各忘却量ごとに 10 種類の荷重係数初期値を用い、学習誤差および補間誤差は初期値に対する平均を表している。

また、式(17)で定義したモデルパラメータ数を用いた AIC の変動を図 7 に示す。中間層ユニット数あるいは全荷重係数個数を AIC におけるモデルパラメータ数として用いる場合、忘却付き学習法などの構造化学習法による内部表現の相違を的確に捉えることが困難であると考えられる。一方、図 6 および図 7 との比較において、補間誤差および AIC が最小となる忘却量が若干異なっているが、式(17)で定義した指標 K が他のパラメータ個数の算出手法とは異なり、内部表現に着目しており提案した指標の有効性が示されていると考えら

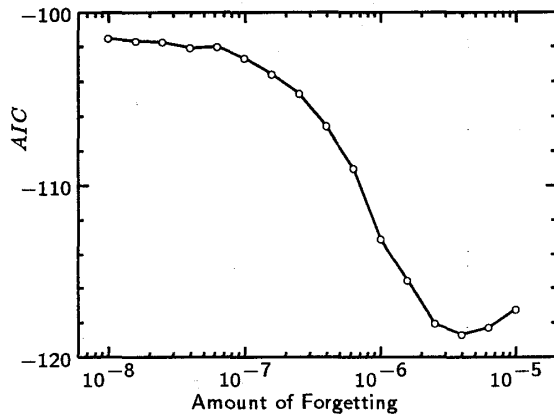


図 7: 忘却量に対する AIC の変動

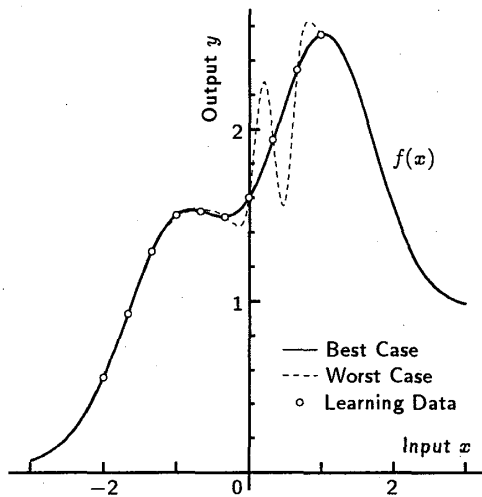


図 8: 近似関数

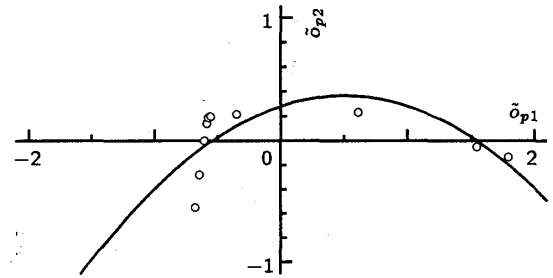
れる。

4.2 荷重係数の非一意性に関する検討

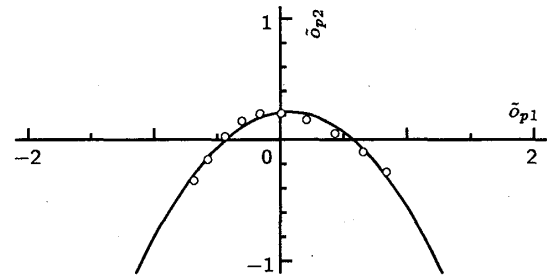
3.2.1節で検討した内部表現の解析手法を用いて、階層型 NN における“荷重係数の非一意性”問題について検討する。式(23)で与えられる非線形関数 $f(x)$ を対象とし、観測雑音が存在しないケースについて検討する。ここで、入力パターン x_p は $[-2, 1]$ において均等に 10 個選んだ。また、中間層ユニット数 n_2 を 5 と設定し、100 種類の荷重係数初期値を用いた。更に、各荷重係数初期値に対して学習誤差が 0.0001 未満になるまで学習を行った。

図 8 に補間誤差が最小および最大となる近似関数を示す。学習誤差が同一でも荷重係数の非一意性により、補間誤差の間に差異が生じることが示されている。

3.2.1節で検討した内部表現の解析手法において、固有値の寄与率から第 2 主成分までを採用した。第 1 お



(a) 補間誤差が最大



(b) 補間誤差が最小

図 9: 第 1 および第 2 主成分の分布 (曲線は主成分に対する 2 次の回帰モデルを表す)

表 1: 補間誤差 E_{Inter} および 情報損失 I_{Loss}

seed	E_{Inter}	I_{Loss}
42	3.1382×10^{-2}	0.135
5	1.0853×10^{-3}	0.102
86	1.3183×10^{-5}	0.044
69	1.1544×10^{-5}	0.038

よび第 2 主成分の分布を図 9 に示す。補間誤差が最小となる場合の主成分が均等に分布しているのに対して、最大の場合には片寄った分布を呈していることがわかる。

更に、荷重係数初期値に対して補間誤差が良好なケースと劣悪なケースにおける情報損失 I_{Loss} (式(21)) を表 1 に示す。ここで、“seed” は荷重係数初期値の種の番号を表している。本表において、補間誤差が小さくなる場合は情報損失が小さくなっており、中間層上において適切な線形化が行われているものと解釈できる。

加えて、補間誤差が最小および最大の階層型 NN に対する学習パターンごとの情報損失 I_{Loss}^p (式(22)) の分布を図 10 に示す。本図においても、補間誤差が小さくなる場合は情報損失 I_{Loss}^p が均等に小さくなっており、中間層上において適切な線形化が行われているものと解釈できる。

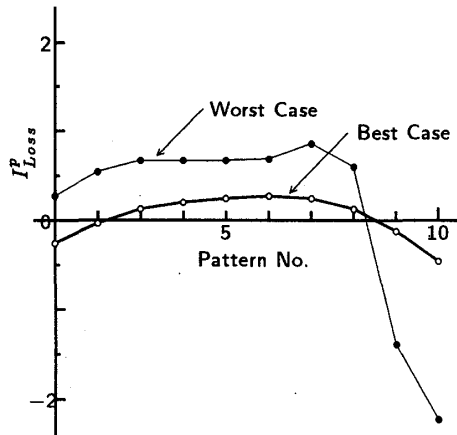


図 10: I_{Loss}^p の分布

5. むすび

本論文では、階層型 NN の汎化能力改善学習法を構築するための一つのアプローチとして、階層型 NN の内部表現と汎化能力の関係について検討した。ここでは、汎化能力の劣化を意味する“過剰学習”問題を“過剰自由度”および“荷重係数の非一意性”の2種類の問題に分割し、内部表現(荷重係数値や中間層ユニット出力値)の観点からこれらの2種類の問題について検討した。まず、荷重係数に関するエントロピーに着目した情報量規範を導入することにより、“過剰自由度”問題に関して検討した。次に、主成分分析法の導入による中間層上の線形化精度評価手法を導入することにより、“荷重係数の非一意性”問題に関して検討した。最後に、数値計算結果から“荷重係数に関するエントロピーに着目した情報量規範”および“主成分分析法の導入による中間層上の線形化精度評価手法”と汎化能力との関係を具体的に示した。

今後、さまざまな非線形関数に対して本評価手法の適用を図る予定である。

謝辞

日頃ご指導頂く清水光教授に感謝致します。

参考文献

- [1] S. Amari, N. Murata, K. R. Müller, and H. Yang. Asymptotic statistical theory of overtraining and cross-validation. Tech. Report METR 03-95, 1995.
- [2] 小川英光, 山崎一孝. ニューラルネットの汎化能力と過学習. 信学技報 NC91-75, 1993.
- [3] 渡辺栄治. 関数近似問題に対する階層型ニューラルネットワークの汎化能力改善学習法. 信学論, Vol. J79-D-II,

No. 5, 1996.

- [4] David S. Chen and Ramesh C. Jain. A robust back propagation learning algorithm for function approximation. *IEEE Trans. on Neural Networks*, Vol. 5, No. 3, pp. 467-479, 1994.
- [5] 船橋賢一. ニューラルネットワークの capability について. 信学技報 MBE88-52, 1988.
- [6] 萩原克幸, 戸田尚宏, 臼井支朗. 階層型ニューラルネットワークにおける結合重みの非一意性と AIC. 信学論, Vol. J76-D-II, No. 9, pp. 2058-2065, 1993.
- [7] 渡辺栄治. 関数近似問題に対する階層型ニューラルネットワークの内部表現と汎化能力の関係に関する一検討. 信学技報 NC95-164, 1996.
- [8] 萩原克幸, 戸田尚宏, 臼井支朗. 離散的に基底を可変とする関数表現の二乗誤差規範の下での統計的性質. 信学技報 NC94-98, 1995.
- [9] 村田昇. 3層ニューラルネットワークの関数近似誤差—ridge 型関数による積分表現定理. 信学論, Vol. J78-A, No. 9, pp. 1204-1212, 1995.
- [10] R. Reed. Pruning algorithms—a survey. *IEEE Trans. on Neural Networks*, Vol. 4, No. 5, pp. 740-747, 1993.
- [11] 石川眞澄. コネクションリストモデルの忘却を用いた構造化学習. 信学技報 MBE88-144, 1988.
- [12] 戸田尚宏, 萩原克幸, 臼井支朗. 多層ニューラルネットワークによるデータフィッティング—情報統計量による構造決定—. 信学技報 NC89-100, 1989.
- [13] 栗田多喜夫. 情報量基準による3層ニューラルネットワークの隠れ層のユニット数の決定. 信学論, Vol. J73-D-II, No. 11, pp. 1872-1878, 1990.
- [14] M. Ishikawa. Structural learning in neural networks. In *Proc. of the 3rd International Conference on Fuzzy Logic, Neural Nets and Soft Computing*, pp. 37-44, Iizuka, JAPAN, 1994.
- [15] 渡辺澄夫. 情報量基準の変形によるニューラルネット最適化の一手法. 信学技報 NC93-52, 1993.
- [16] 坂元慶行, 石黒真木夫, 北川源四郎. 情報量統計学. 共立出版, 1978.
- [17] M. Ishikawa and H. Uchida. A structural learning of neural networks based on an entropy criterion. In *Proc. of IEEE/INNS Joint Conference on Neural Networks*, Vol. II, pp. 375-380, Beijing, CHINA, 1992.
- [18] 木村義政. 誤差及び出力変動を最小化する学習アルゴリズムとその低品質手書き漢字認識への応用. 信学論, Vol. J73-D-II, No. 6, pp. 840-847, 1990.
- [19] 鈴木紳, 河原英紀. 平均曲率に基づいた神経回路網の評価基準. 信学論, Vol. J75-D-II, No. 3, pp. 637-645, 1992.
- [20] 賈, 戸田, 臼井. ニューラルネットワークにおける逆伝搬学習アルゴリズムの初期値設定に関する一考察. 信学論, Vol. J73-D-II, No. 8, pp. 1179-1158, 1990.

- [21] 入江文平, 川人光男. 多層パーセプトロンによる内部表現の獲得. 信学論, Vol. J73-D-II, No. 8, pp. 1173-1178, 1990.
- [22] I. Sekita, T. Kurita, D. K. Y. Chiu, and H. Asoh. A method to reduce redundant hidden nodes. *IEICE Trans. on Inf. and Syst.*, Vol. E77-D, No. 4, pp. 443-449, 1994.
- [23] S. Tamura, M. Tateishi, M. Matsumoto, and S. Akita. Determination of the number of redundant hidden units in a three-layered feed-forward neural networks. In *Proc. of IEEE/INNS Joint Conference on Neural Networks*, Vol. I, pp. 335-338, Nagoya, JAPAN, 1993.