

論文 4層フィードフォワードニューラルネットワーク の写像能力*

The Capabilities of A Four-layered Feedforward Neural Network

田村 震一 立石 雅彦
Shin'ichi Tamura Masahiko Tateishi

Neural network theorems state that, in the limit of infinitely many hidden units, a four-layered feedforward neural network is equivalent to a three-layered feedforward neural network. In the applications of neural networks, however, we can not use infinitely many hidden units. The capabilities of a neural network with a finite number of hidden units should be studied. In this paper, we first give another proof that a three-layered feedforward network with $N-1$ hidden units can realize any N input-target relations exactly. Based on the results of the proof, we then construct a four-layered feedforward network which can realize any N input-target relations with an arbitrarily small error using only $N/2+3$ hidden units. This shows that a four-layered feedforward network is superior to a three-layered feedforward network in terms of the number of parameters for a training data realization.

Key Words : Neural Network, 3-layered Feedforward Neural Network, 4-layered Feedforward Neural Network, Neural Network Mapping Capabilities.

1. 始めに

生物の脳神経系の数理モデルであるニューラルネットワークは人間が持つ高度な情報処理機能の実現の可能性がある。音声認識、画像認識などのパターン認識、雑音と信号の分離、高品質な音声合成、音の情景分析、情報圧縮などの信号処理、人間に代表される非線形システムモデル化、複雑なプラントの制御、故障診断、センサー信号の統号処理、さらには株価や天気などの高精度な予測等、様々な分野への応用が多くの研究機関で試みられている。

応用に用いられるニューラルネットワークはフィードフォワードネットワーク¹⁾⁻³⁾と呼ばれるタイプのものが主流であり、その中でも実現の容易さなどから3層型と4層型の数理モデルが主に使われる。3層フィードフォワードネットワークと4層フィードフォワードネットワークとはニューラルネットワークの構成要素である、神経細胞に相当するユニットをつなぐ結合方式が異なる。同じユニット数でも4層フィー

ドフォワードネットワークの方がその処理に要する計算量や記憶容量を少なくできるためのハードウェア化の点から有利である。しかし同じユニットを持つ3層と4層のフィードフォワードネットワークの能力が同じであるという理論的な保証は得られていない。応用する問題に応じて、研究者が自分の直感や試行錯誤によって3層フィードフォワードネットワークと4層フィードフォワードネットワークとを適当に使い分けているのが現状である。

本論文ではこの現状に理論の光をあてハードウェア化の際の計算量や記憶容量の最適化までも考慮に入れたニューラルネットワークモデル選択の設計指針を与える。すなわち、応用に際し4層フィードフォワードネットワークは3層フィードフォワードネットワークよりもはるかに少ないユニット数でその仕事を実現可能であるという証明を与え、同数のユニットの場合、4層フィードフォワードネットワークの方がより複雑な仕事ができることを明らかにする。

まず第2章で $N-1$ 個の隠れユニットを持つ3層フィードフォワードネットワークが任意の N 個の入出力関係を実現可能であるという事に新しい証明を与

* 1996年7月24日 原稿受付

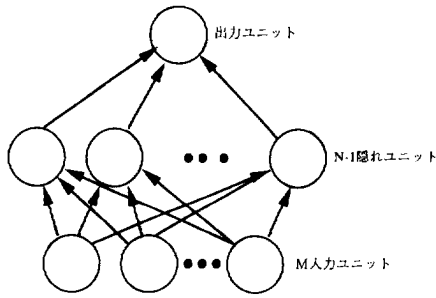


図1 3層フィードフォワードネットワーク

える。この証明の結果に基づいて第3章で任意の N 個の入出力関係を実現する4層フィードフォワードネットワークを構成する。構成に使われた隠れユニットの数が、3層ネットワークが必要とする $N-1$ 個よりも少ないことを示し4層ネットワークの優位性を示す。

2. 3層フィードフォワードネットワークの写像能力の証明

$N-1$ 個の隠れユニットを持つ3層のフィードフォワードネットワークは任意の N 個の入出力関係を実現できることが知られている^{4),5)}。我々はその事実新しい証明を与える⁶⁾。我々の証明が対象とする3層フィードフォワードネットワークを図1に示す。

ネットワークは M 個の入力ユニット、 $N-1$ 個の隠れユニット、1個の出力ユニットから成る。入出力ユニットは線形ユニット、隠れユニットはシグモイド関数をもつ非線形ユニットである。入力層と隠れ層、隠れ層と出力ユニットは、各全結合になっている。入力ユニット、隠れユニットは、各、1から M 、1から $N-1$ と番号が付けられている。

以下の記号を定義する。

- $(\mathbf{x}^{(k)}, t^{(k)})$: k 番目の入出力関係
- $\mathbf{x}^{(k)}$: k 番目の M 次元入力ベクトル
- $t^{(k)}$: k 番目の教師出力
- \mathbf{t} : 教師出力ベクトル $t_k = t^{(k)}$ ($k=1, 2, \dots, N$)
- $o^{<i>}_k$: $\mathbf{x}^{(k)}$ を入力した時の i 番の隠れユニットの出力
- $s(x)$: シグモイド関数、 $1/(1+\exp(-x))$
- w_{ij} : j 番の入力ユニットから i 番目の隠れユニットへの結合荷重 ($i=1, 2, \dots, N-1, j=1, 2, \dots, M$)
- b_i : i 番の隠れユニットのバイアス値
- \mathbf{W} : 隠れ層から出力ユニットへの結合荷重と出力ユニットのバイアス値のベクトル

ル

W_i : $i-1$ 番の隠れユニットから出力ユニットへの結合荷重 ($i=2, 3, \dots, N$)

W_1 : 出力ユニットのバイアス値

空間はユークリッド空間を仮定する。 N 個の入出力関係を実現するには、ネットワークは以下の方程式を満たさねばならない。

$$\begin{aligned} \mathbf{O}\mathbf{W} &= \mathbf{t}, \quad O_{i,1} = 1 \quad (i=1, 2, \dots, N), \\ O_{k,i} &= o^{<i>}_k \quad (i=2, \dots, N, k=1, \dots, N) \end{aligned} \quad (1)$$

(1)の $N \times N$ マトリクス O を次のように書く。

$$O = (\mathbf{1}, \mathbf{o}^{<1>}, \mathbf{o}^{<2>}, \dots, \mathbf{o}^{<N-1>}) \quad (2)$$

$\mathbf{1}$ は要素がすべて1である定数列ベクトル、 $\mathbf{o}^{<i>}$ は $i+1$ 番目の列ベクトルである。証明すべきことは、方程式(1)が解けること、すなわち O をフルランクにできるということである。

$\mathbf{w}^{<i>}$ を入力層から i 番の隠れユニットへの結合荷重をすべてならべたベクトル、 (\cdot) を内積とすると、 $\mathbf{x}^{(k)}$ を入力した時の i 番の隠れユニットの出力は次のように書ける。

$$o^{<i>}_k = s((\mathbf{w}^{<i>}, \mathbf{x}^{(k)}) + b_i) \quad (k=1, 2, \dots, N) \quad (3)$$

$o^{<i>}$ は $\mathbf{o}^{<i>}$ の k 番目の要素、 $(\mathbf{w}^{<i>}, \mathbf{x}^{(k)})$ は i 番の隠れユニットへの入力である。

ユークリッド空間 R^N 内の曲線、 $\mathbf{c}(b_i)$ を考える。

$$\mathbf{c}(b_i) = \mathbf{o}^{<i>} = \begin{pmatrix} s(b_i + (\mathbf{w}^{<i>}, \mathbf{x}^{(1)})) \\ s(b_i + (\mathbf{w}^{<i>}, \mathbf{x}^{(2)})) \\ \vdots \\ s(b_i + (\mathbf{w}^{<i>}, \mathbf{x}^{(N)})) \end{pmatrix} \quad (4)$$

$$\begin{aligned} b_i \in I = [a, b] \quad (a < b, a, b \in R) \\ (i=1, 2, \dots, N-1) \end{aligned}$$

(4)において結合荷重ベクトルは、 $k \neq k'$ ならば、 $(\mathbf{w}^{<i>}, \mathbf{x}^{(k)}) \neq (\mathbf{w}^{<i>}, \mathbf{x}^{(k')})$ という条件を満足しているとする。条件が成立しない結合荷重ベクトルがあれば、微小な外乱ベクトルをその結合荷重ベクトルに加えればよい。

この曲線が、 $N-1$ 次元以下の次元の R^N の部分空間に含まれないことを示す。 $\mathbf{c}(b_i)$ が $N-1$ 次元の部分空間に入っていると仮定する。この時、この部分空間に直交するベクトル、 \mathbf{n} が存在して以下の式が成り立つ。

$$\begin{aligned} \forall b_i \in I = [a, b], \\ (\mathbf{n}, \mathbf{c}(b_i) - \mathbf{c}(a)) = n_1 \cdot s(b_i + d_1) \\ + n_2 \cdot s(b_i + d_2) + \dots \\ + n_N \cdot s(b_i + d_1) - z \end{aligned}$$

$$=0$$

$$d_k = (\mathbf{w}^{<i>, \mathbf{x}^{(k)} \quad (k=1, 2, \dots, N), \quad (5)$$

$$z = (\mathbf{n}, \mathbf{c}(a))$$

n_N がゼロでないとして (5) 式を変形して以下の式を得る。

$$s(b_i) = \sum_{p=1}^{N-1} \alpha_p \cdot s(b_i + e_p) + z' \quad (6)$$

$$\alpha_p = -n_p/n_N, \quad z' = z/n_N, \quad e_p = d_p - d_N,$$

$$e_p \neq e_{p'}, \quad b_i \in I' = [a + d_N, b + d_N]$$

パラメーター、 α_p ($p=1, 2, \dots, N-1$)、 z' は任意の $\Theta \in I'$ について以下の線型方程式を満足しなければならない。

$$\sum_{p=1}^{N-1} \alpha_p \cdot s(\Theta + e_p) + z' = s(\Theta) \quad (7.1)$$

$$\sum_{p=1}^{N-1} \alpha_p \cdot s^{(1)}(\Theta + e_p) + z' = s^{(1)}(\Theta) \quad (7.2)$$

$$\sum_{p=1}^{N-1} \alpha_p \cdot s^{(2)}(\Theta + e_p) + z' = s^{(2)}(\Theta) \quad (7.3)$$

⋮

$s^{(n)}(\cdot)$ はシグモイド関数の n 階微分である。シグモイド関数は有限次数の多項式ではないから、この線型方程式は無数個ある。しかし、調節可能な自由パラメーターは $\alpha_p, \alpha_{p'}, \dots, \alpha_{N-1}, z'$ の N 個しかなく、(7) 式をすべて満たすことは不可能である。したがって、曲線 $\mathbf{c}(b_i) = \mathbf{o}^{<i>, b_i \in I = [a, b]$ ($a < b, a, b \in R$) は $N-1$ 次元以下の次元の部分空間に入ることはない。

以上のことから、 i 番 ($i=1, 2, \dots, N-1$) の隠れユニットについて任意の区間 $[a, b] \subset R$ から N 個のバイアス値、 $b_i^{(1)}, b_i^{(2)}, \dots, b_i^{(N)}$ を、対応するベクトル、 $\mathbf{c}(b_i^{(1)}), \mathbf{c}(b_i^{(2)}), \dots, \mathbf{c}(b_i^{(N)})$ が一次独立になるように選ぶことがわかる。これは、各の隠れユニットのバイアス値をうまく調節して、列ベクトル、 $\mathbf{1}, \mathbf{o}^{<i>, \mathbf{o}^{<2>, \dots, \mathbf{o}^{<N-1>$ を一次独立にすることが可能であることを意味する。つまり、隠れユニットのバイアス値を調節することによって、マトリクス、 \mathbf{O} をフルランクにすることが可能であり、 $N-1$ 個の隠れユニットを持つ 3 層フィードフォワードネットワークが任意の N 個の入出力関係を実現できることが示されたことになる。

以上の証明から、 N 個の入出力関係を実現する際に、次の二つのことがいえる。

- (1) 入力層から隠れ層への結合荷重は、ほとんど任意に選ぶことができる。
- (2) 各の隠れユニットのバイアス値は、任意の実数の区間から選ぶことができる。

3. 4 層フィードフォワードネットワークの構成

4 層フィードフォワードネットワークは任意の N 個の入出力関係、 $(\mathbf{x}^{(k)}, t^{(k)})$ ($k=1, 2, \dots, N$)、 $\mathbf{x}^{(k)} \in R^M$ 、 $t^{(k)} \in R$ を 3 層フィードフォワードネットワークよりも少ない隠れユニットで実現できる事を示す。簡単のため、 N は偶数であるとする。

まず始めに、図 2 に示す 3 層サブネットワークを考える。

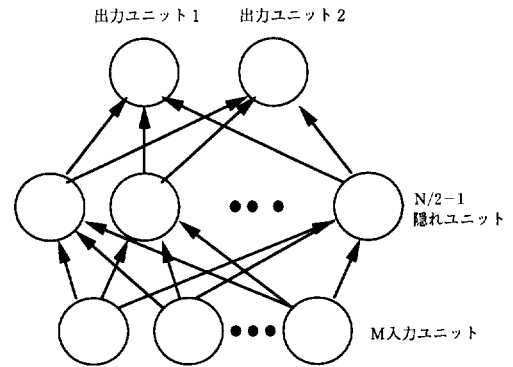


図 2 3 層サブネットワーク

サブネットワークは M 個の線形入力ユニット、 $N/2-1$ 個の非線形隠れユニット、二つの非線形出力ユニットからなる。非線形ユニットはシグモイド関数を持つ。入力層と隠れ層、隠れ層と出力層は、各々全結合を形成しており、入力層と隠れ層の結合は、乱数を使って決定されているとする。入力空間 R^M において N 個の入力ベクトル、 $\mathbf{x}^{(k)}$ ($k=1, 2, \dots, N$) を半分に分ける超平面を一つ決定する。データが有限個なのでこのような超平面は常に存在する。この超平面によって分けられた入力ベクトルの組を V_1, V_2 と名付け、入力ベクトルの番号を以下のように付け直す。

$$\mathbf{x}^{(k)} \in V_1 \quad (k=1, 2, \dots, N/2),$$

$$\mathbf{x}^{(k)} \in V_2 \quad (k=N/2+1, \dots, N) \quad (8)$$

また、 N 個の教師出力から、次の新たな教師出力を作る。

$$t^{(k)} = t^{(k)}/C + 0.5 \in (0, 1) \quad (k=1, 2, \dots, N) \quad (9)$$

C は正の定数で、 $t^{(k)}$ が区間 $(0, 1)$ に入るように適当に決める。

第 2 章で述べたように、このサブネットワークの隠れユニットのバイアス値を調節することで、 V_1 のすべてのベクトルを入力して形成される $N/2 \times N/2$ のマトリクス、 \mathbf{O} をフルランクにすることができる。したがって、出力ユニット 1 は、 V_1 の入力ベクトルに対し、

対応する新たな教師出力、 $t^{(k)}$ を出力する事が可能である。出力ユニット1のバイアス値と出力ユニット1と隠れ層間の結合荷重から成るベクトル、 w_1 は以下の方程式から決定される。

$$Ow_1 = u, \quad u_k = s^{-1}(t^{(k)}) \quad (k=1, 2, \dots, N/2) \quad (10)$$

$s^{-1}(\cdot)$ はシグモイド関数の逆関数である。

V_2 のすべてのベクトルから形成されるもう一つの $N/2 \times N/2$ マトリクス、 O' を考える。隠れ層のバイアス値は V_1 のベクトルについて調整されているのであるから、このマトリクスがフルランクである保証はない。 V_1 の入力に調節された隠れ層のバイアス値を以下のように書く。 O' がフルランクでないと仮定する。

$$B = (b_1, b_2, \dots, b_{N/2-1})^t \quad (t:\text{transpose}) \quad (11)$$

第2章で述べたように、 O' をフルランクにするバイアス値は、各の隠れユニットについて任意の実数の区間から選ぶことができる。したがって、任意の小さな値、 $e > 0$ に対し、 O' をフルランクにする隠れ層のバイアス値ベクトル、 B' を以下のように選ぶことが可能である。

$$B' = (b'_1, b'_2, \dots, b'_{N/2-1})^t, \quad b'_i \in [b_i - e, b_i + e] \quad (i=1, 2, \dots, N/2-1) \quad (12)$$

マトリクス、 O の行列式はバイアス値の連続関数になっており B では、仮定からゼロでない。従って、(12) の e を十分小さくとれば、マトリクス、 O の B' での行列式はゼロにならない。従って、我々は常にマトリクス、 O と O' の両方をフルランクにするようにバイアス値を選択することができる。このように選択したバイアス値を用いて、出力ユニット2が、 V_2 の入力ベクトルに対し、対応する新たな教師出力、 $t^{(k)}$ を出力するように、隠れ層と出力ユニット2の間の結合荷重と出力ユニット2のバイアス値を出力ユニット1の場合と同様に決定する。この時点では、出力ユニット1は V_2 入力に対し、出力ユニット2は V_1 入力に対して、正しい値を出力しない。

ここで、新たに二つのシグモイド関数を持つ隠れ非線形ユニット、A と B をこのサブネットワークに付け加える。それを図3に示す。

他の隠れユニットと同様にユニットA、Bは入力層と全結合をしている。しかし、ユニットAは出力ユニット1のみと、ユニットBは出力ユニット2のみと結合している。

ユニットAについて述べる。入力層とユニットA

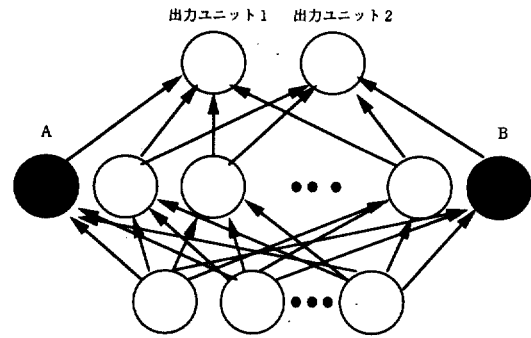


図3 隠れユニット, A, B

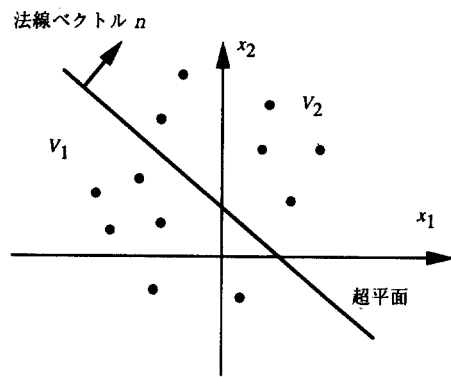


図4 超平面とその法線ベクトル

との結合荷重ベクトルを入力空間 R^M で N 個の入力ベクトルを二つに分けている超平面の法線ベクトル、 n に正の数、 T を掛けたベクトル、 Tn に等しくとる。そして、ユニットAのバイアス値を超平面上のベクトルが入力された時、ユニットAの出力が0.5になるように調整し、 T を十分大きくする。こうすることで、ユニットAは V_1 と V_2 に対する判別器になる(図4)。

ユニットAは V_1 入力に対して、ゼロに近い値を、 V_2 入力に対しては1に近い値を出力する。ユニットAと出力ユニット1との結合荷重を大きな正の数、 U にマイナス符号を付けた値、 $-U$ とすることで、出力ユニット1はユニットAから以下の値を受け取る。

V_1 入力の場合：ゼロに近い負値、 $-e_i$ ($i=1, 2, \dots, N/2$)

V_2 入力の場合：大きな負値、 $-E_i$ ($i=N/2+1, \dots, N$)

T と U を大きくすることで、 e_i, E_i をゼロ、無限大にいくらかでも近付ける事ができる。出力ユニット1は、他の隠れユニットからは、 V_1 入力の時、対応する教師出力となる信号を、 V_2 入力の時、ある定まった信号を受け取る。したがって、出力ユニット1は V_1 入力の時

時、対応する教師出力にいくらでも近い値を、また、 V_2 入力の場合は、ゼロにいくらでも近い値を出力可能となる。同様にしてユニット B を出力ユニット 2 に対し調整することで、図 3 のネットワークの出力を以下のようにすることができる。

V_1 入力の場合：出力ユニット 1 は、対応する教師出力にいくらでも近い値を、出力ユニット 2 はゼロにいくらでも近い値を出力する。

V_2 入力の場合：出力ユニット 1 は、ゼロにいくらでも近い値を、出力ユニット 2 は対応する教師出力にいくらでも近い値を出力する。

ネットワークの構成を完成するには、このネットワークに最終の線形出力ユニットを一つ加えればよい。図 5 に完成した 4 層フィードフォワードネットワークを示す。

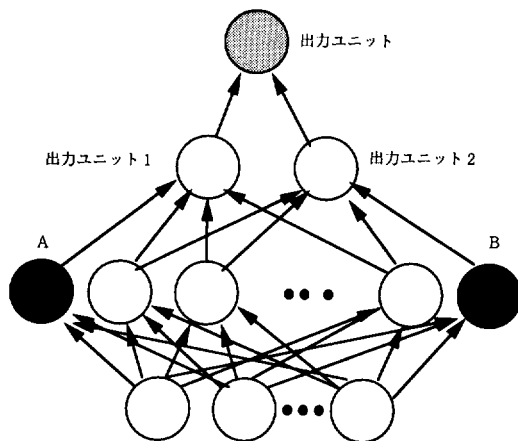


図 5 4 層フィードフォワードネットワーク

もとの教師出力、 $t^{(k)}$ を作り出すためには、サブネットワークの出力ユニット 1, 2 から最終の出力ユニットへの結合荷重は (9) 式の C に、また、出力ユニットのバイアス値は $0.5C$ に設定すれば良い。以上より、四層フィードフォワードネットワークは、 $N/2+3$ 個の隠れユニットで、 N 個の入出力関係を任意の精度で実現可能であることが分かる。

4. 結 論

$N-1$ 個の隠れユニットをもつ 3 層フィードフォワードネットワークが任意の N 個の入出力関係を実現できるのに対し、4 層フィードフォワードネットワークでは、 $N/2+3$ 個の隠れユニットにより任意の N 個の入出力関係を実現できることを証明し、さらにその構成法を示した。これより 4 層フィードフォー

ドネットワークが 3 層フィードフォワードネットワークより優れた写像能力を持つことを示した。応用に際しては、適用事例毎に最適なネットワークモデル、隠れユニット数を選択することが重要である。本論文の結果は 4 層フィードフォワードネットワークが 3 層フィードネットワークのおよそ 1/2 の隠れユニット数で同等の機能を実現できることを示す。また適用事例に依らない最適な隠れユニット数の選択の指針を与える。さらにハードウェア化における計算量、記憶容量の決定を容易にする。

＜参 考 文 献＞

- 1) K. Hunahashi : "On the approximate realization of continuous mapping by neural networks," Neural Networks, vol. 2 (1989), p. 183
- 2) G. Cybenko : "Continuous valued neural networks with two hidden layers are sufficient," Mathematics of Controls, Signals and Systems, vol. 2 (1989), p. 303
- 3) K. Hornik, M. Stinchcombe and H. White : "Multilayer feedforward networks are universal approximators," Neural Networks, vol. 2 (1989) p. 359
- 4) M. A. Sartori et al. : "A Simple Method to Derive Bounds on the Size and to Train Multilayer Neural Networks," IEEE Trans. Neural Networks, vol. 2 (1991), p. 467
- 5) S. C. Huang and Y. F. Huang : "Bounds on number of hidden neurons in multilayer perceptrons," IEEE Trans. Neural Networks, vol. 2 (1991), p. 47
- 6) S. Tamura : "Capabilities of a Three Layer Feedforward Neural Network," Proc. of IJCNN91 (1991), p. 2757

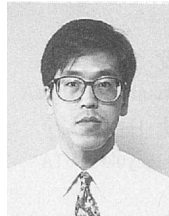


〈筆 者〉



田村 震一 (たむら しんいち)

基礎研 3 部
ニューラルネットワークの理論、
応用研究に従事。



立石 雅彦 (たていしまさひこ)

基礎研 3 部
ニューラルネットワークの応用研
究に従事。