

多層ニューラルネットワークの
情報集約機能に関する基礎的研究

米 倉 達 広

①

多層ニューラルネットワークの
情報集約機能に関する基礎的研究

米倉達広

目次

まえがき

第1章. 序論

- 1. 1 統計的パターン認識における次元数削減の現状および諸問題 ... 6
- 1. 2 多層ニューラルネットワークの背景と現状 ... 8
- 1. 3 本研究の位置付け、目的ならびに構成 ... 10

第2章. 多層ニューラルネットモデルの定義と機能

- 2. 1 多層ニューラルネットによる判別分析機能と情報集約機能 ... 15
- 2. 2 恒等写像を近似する砂時計モデル ... 18
- 2. 3 パルス入力・パターン出力写像を近似するPPNモデル ... 20
- 2. 4 まとめ ... 23

第3章. 3層PPNの幾何学的性質

- 3. 1 パターンの成分分解 ... 26
- 3. 2 出力層素子が線形特性をもつ場合 ... 30
- 3. 3 出力層素子がsigmoid特性をもつ場合 ... 30
- 3. 4 シミュレーション ... 33
- 3. 5 まとめ ... 37

第4章. 4層PPNと区分線形因子分析

4. 1	区分線形関数 F_{pl} による sigmoid 関数の近似	... 40
4. 2	4 層 P P N の集約表現の解析	... 42
4. 3	4 層 P P N による区分線形因子分析法とクラスタリング	... 48
4. 4	シミュレーション	... 50
4. 4. 1	多様体上に分布した標本の近似	... 51
4. 4. 2	アイリスデータの次元削減と区分線形回帰	... 52
4. 4. 3	考察	... 53
4. 5	まとめ	... 55

第 5 章. 多層ニューラルネットによる
データ圧縮機能、因子分析機能

5. 1	多層ニューラルネットによるデータ圧縮と因子分析	... 61
5. 1. 1	3 層砂時計モデルと 3 層 P P N	... 62
5. 1. 2	5 層砂時計モデルと 4 層 P P N	... 63
5. 2	シミュレーション	... 64
5. 3	まとめ	... 66

第 6 章. あとがき

6. 1	総括	... 73
6. 2	今後の課題	... 76

謝辞

付録 A.		... 82
付録 B.		... 83
付録 C.		... 84
付録 D.		... 85
付録 E.		... 86

まえがき

統計的パターン認識は、従来、我々の身のまわりで見たり聴いたりすることのできる事物、森羅万象（物理的現象）を類別したり識別したりするための統計的手法である。従って、その目指す機能は人間の視覚や聴覚等の感覚を基準とする必要があり、実際、それがパターン認識能力の工学的な評価尺度になっている場合が多い。当然のことながら我々人間には、この類別や識別を行う能力が備わっており（もっと正確には無意識のうちにそれらの能力を身につけると言うべきであろうが）、いとも簡単に他人の顔や声、更には食べ物の臭いや味を認識することができる。しかし、残念ながらそれを行う具体的手順またはアルゴリズムを、我々が自ら自覚し、定式化しているとはいえない。我々人間が、効率の良いパターン認識アルゴリズムやパターン分類器をうまい具合に設計できないのもこのためであろう。従って、その意味で、我々人間の脳がおこなっている情報処理をなんらかの形で定式化し、これを実現する数理モデルを体系化することは、パターン認識の大きな目標の一つであると言ってよい。

一方、現在自然科学や人文、社会科学等種々の分野で幅広く用いられている多変量解析をはじめとする統計手法は、その殆どが線形あるいはそれに準ずる準線形の確率分布モデルを仮定しており、高等動物の脳がおこなっている複雑で柔軟で効率のよい情報処理の形式をモデル化する道具としてはあまりにも非力な点が多い。すなわち、パターン認識の方法論の進展のために、生体の脳の動作の有様を探ること、そしてこれを説明するための数理的な道具（統計的手法）を整備すること（非線形数学の更なる充実、特に多変量解析手法の非線形への拡張等）、はきわめて重要な課題の一つといえる。

言い替えば、我々高等動物の脳の機能とその形態とを解析し、これを模倣することでこの工学的モデルを実現しようとする試みは、パターン認識の分野だけでなく、数理統計学に対しても全く新しい方法論を提起することが可能となるように思われる。その動きがまさに、ここ数年の間おおきな成果を挙げつつある、ニューラ

ルネットワークである。

本文はこの観点に基づいて、線形性を仮定する多変量解析のいくつかの手法と、パターン認識の分野で理論、応用の両側面から現在注目されている多層ニューラルネットワークとの基本的な対応付けを目的とする。なお、多変量解析の枠組みは、外的基準の有無により大きく2つに分割することができる。本文の目指すところは、それらのうちでも特に柔軟な手法を要すると思われる、外的基準無しの解析法と多層ニューラルネットワークの関連、言い替えれば多層ニューラルネットワークの“類別”（パターン認識における教師無し学習に対応する）の能力を考察し、その能力を多種多様な問題に活用するための理論的基盤の足掛りを確立することにある。本文では、筆者らが独自に提案した構成の多層ニューラルネットワークを用いて、これらの理論的考察と実験的検証を進める。

次頁からの序論では、統計的パターン認識における教師無し学習法の現状と多層ニューラルネットワークの現状について説明し、本文の目的の学術的意義を示すと同時に本文全体の構成を簡潔に示す。

第1章

序論

1. 1 統計的パターン認識における次元削減機能の現状および諸問題

パターン認識とは、物理的現象や心理的現象により生ずる観測データであるパターン（ベクトル）の集まりを有限個の類（クラス）に分けることを目的とする工学的分野である。従って、数学的には、パターン空間（パターンベクトルの集合）から、カテゴリー空間（カテゴリーの集合）への写像ととらえることができる(1)-(4)。

そこで、そのカテゴリーの情報が明示的に与えられているか否かにより、パターン認識の手法は教師付き（supervised）分類と教師無し（unsupervised）分類の2種類に大別される。前者は、多変量解析における外的基準有り(2)の場合に相当し、既知のカテゴリー情報に基づいて、入力される未知パターンを既知カテゴリーのいずれかに対応させる（これを識別と呼ぶ(5)）問題である。この手法の研究は、文字、音声、画像といった実際の産業応用、医学応用と結び付いてパターン認識のいわば主流となっており、具体的な手法も多種多様に提案されている。一方、後者は多変量解析における外的基準無しの場合に相当し、類似性を表す何等かの尺度により、類似したもののどうしを個々の類（クラス）に入れる手法と解釈されている。これは、類別の手法(5)、あるいはクラスタリングともいわれる。この手法は、識別の手法と比較して、カテゴリー構造の情報がより少ないかまたは欠落している場合に用いられ、従って個々の手法は、識別の手法と比較すれば、むしろad hoc（個々の問題に応じて使用すべき方法が異なる）な性格をもっているといえよう。しかしながら、クラスタリングは、本質的次元（intrinsic dimensionality）(3)や、次元数削減の概念、ならびに多次元データの潜在的な構造情報の抽出、データ圧縮または集約表現等の用語と関連しており、未分化な状態(6)（本質的なカテゴリー情報が未だ得られ

ていない状態)のパターン集合を対象とする場合の問題として、実際上きわめて身近で重要な問題であるということもできる(クラスタリングの諸手法および応用例については、例えば文献(3)(4)(6)を参照されたい)。

一方、識別、すなわち教師付き分類の問題においても、与えられたパターンを、何等かの単一的な変換を通して直接にカテゴリ空間に結び付ける手法は、個別の応用分野に依らず殆どの場合に非能率であるため、不変量抽出や統計的特徴抽出のための変換によるパターン次元数削減のプロセス(必要な情報の損失を最小限度に抑えながら、より少ない次元に落とす)が不可欠であり、またその効率が、パターン認識システム全体の能力に与える影響の大きさはよく知られている(1)。

これらの意味で、パターン認識、特に統計的パターン認識における効率的な次元削減の手法は、類別の問題にあっては本質的次元の空間における表現の獲得やデータ圧縮の効率を決定し、識別の問題ではそれ以降の識別能力を左右する、基本的な問題といえる。

こうした状況にあって、実際の統計的パターン認識で用いられる次元削減の方法は、識別の問題においてはクラス間の分離を強調する線形判別分析法が、類別においては近似誤差を最小にするK-L展開などの線形近似手法がよく用いられる。実際、多次元ガウス分布等で近似し得る線形な確率モデル(2)の場合、これらの手法の有効性は明かであろう。しかしながら、この仮定の成り立たない場合において、これら線形の手法(多変量解析手法と言い替えてもよい)には限界があることは言うまでもない。

一つの例として、無作為に選んだ、(例えば)日本人1万人の、身長 X と体重 Y と手のひらの表面積 Z を計測した3変数データのサンプル集合を考える。容易に想像し得るように、1万個の3次元データ (X, Y, Z) を3次元ユークリッド空間にプロットすれば、それらのサンプルは、幾分ばらつきを持ちながらもおそらく一つの曲線でよく近似し得る分布となろう。しかもその曲線は、 Z の値が X の2次式で、 Y の値が X の3次式で近似表現されるであろうことが予想される。これは、身長と体重と手のひらの表面積の各次元の関係を知っていれば誰でも予想し得ることであり(パターン認識問題の多くの場合、これらの関係を事前に知ることは不可能であるが)、 X, Y, Z を線形に1変数で近似表現する場合(すなわちK-L展開1項近似を意味する)と比較して近似誤差は必ず減少するはずである。また、与えられた3変数につ

いてこのような関係が成り立つことを知る道具があれば、逆にそれら変数の間の潜在的な構造を解析することも可能であろう（非線形因子分析と呼ぶべき概念）。

上の例では、与えられた多次元の測定データが、連続かつ非線形の分布モデル（分布モデルについては未知）に従う場合を想定し、そこでの非線形データ回帰手法の有用性を強調した。実際、パターン認識の諸問題で測定量の間にはこのような非線形かつ潜在的な解析的關係が成り立つ場合に、現在の線形手法による次元削減が有効であるとは考えられない（現状では、おそらくこのような場合が考慮されることは極めて希であり、変数どうしの潜在的な関係を推し量る尺度として用いられているものの殆どは、線形な近似からの逸脱の度合を数量化したものである）。

この意味で、多変量解析の諸手法を、数学的には統一的な形で非線形に拡張した、非線形因子分析や非線形判別分析による次元削減手法の確立は、現在の統計的パターン認識における極めて有効な道具として、大きな関心の一つになっている。

1. 2 多層ニューラルネットワークの背景と現状

本節では、多層ニューラルネットワークが今日盛んに研究されている歴史的背景および、同モデルの理論研究の面での方向性について触れる。なお、以下に述べる歴史的背景は、甘利の文献(7)等で詳細に示されているものの概略である。

人間など高等動物の神経回路網の動作原理を情報処理の観点から捉えようとする試みは、McCulloch & Pitts による線形しきい値素子（多入力の線形荷重和があるしきい値を越えるか否かで1または0の出力を出す素子）に始まるとされる(7)。この後、心理学者 Hebbによる学習規則、すなわちニューロン（素子）間の結合強度の変化で学習を実現できることが明かにされた(7)。これらはいずれも1940年代の研究成果である。それ以降、この学習規則を実現する工学モデルが数多く提出された。なかでも、Rosenblatt(8)らにより提案されたパーセプトロンは、学習の収束定理が証明されたことから1960年代のブームとなっている。しかしながら、1960年代末に Minsky & Papert(9)によって示されたパーセプトロン限界説（入力層と中間層の間の結合強度修正を含まない、本質的に2層の多層ニューラルネットワークでは、線形分離可能でないクラス判別問題を解くことができない）に呼応して、そのブームが終わる。そこで得た大きな課題は、解かれるべき問題の表現と実際のネッ

トワークモデルの数理的關係を明かにすること、すなわち、問題に対して適正なニューラルネットモデルを対象とすべきであるということである。代表的な例として、線形分離不可能な判別問題に対しては、3層以上の多層パーセプトロン型ニューラルネットに対する汎用な学習アルゴリズムが実現されるべきである、等である。

その後、約15年を経過して、より強力な機能をもつニューラルネットのモデルが幾つか提案された。その中でも現在最も多く引用され研究されているものの一つが、Rumelhart & McClelland(10)により考案された、3層以上の多層パーセプトロンに対する汎用で具体的な学習アルゴリズムである誤差逆伝搬法 (Error Back Propagation) である。この詳細については次章で述べるとし、以下では多層パーセプトロン型ニューラルネット (以下、単に多層ニューラルネットと略記) の機能と多変量解析の関連について考えることにする。

前節で述べた、体系的な枠組みをもつ非線形多変量解析の概念は、多層ニューラルネットによるパターンの変換を通して実現できる可能性があることを以下に示す。

多層ニューラルネット (またはより広義に階層型ニューラルネットと言い替えてもよい) を、図1-1に示す。ここで、ネットワークの構成要素である素子は全体として層状に構成される。信号 (通常アナログ値) は素子間の結線を通して左側の層から右側の層へと一定方向に伝搬し、一つの層内の素子間に結線はない。また、各素子は多入力-出力の単一特性 (一般には非線形特性) を持つものとする。

このモデルの果たす機能は、入力層 (原パターン信号が入力される層) にはパターンデータを、出力層 (変換後のパターンが出力される層) には望ましい教師信号を与えることで、それらの対間の写像を何等かの基準に沿って学習により近似させることである。入力層と出力層の間には幾つかの層 (これらを中間層と呼ぶ) があり、それら各層内には複数個の素子があるため、これらの素子の状態を数値で表せば、各層の状態を、層内素子の値を成分としてもつベクトルに対応させて考えることができる。更に、各素子のもつ非線形特性が連続で滑らか、すなわち解析的であるとすれば、それらベクトル間の非線形写像もまた解析的である。

従って、出力層に与えられる教師信号をカテゴリー空間とすれば、これは、パターンベクトル集合からカテゴリーベクトル集合への非線形連続写像であり、多変量解析における外的基準有りの諸問題 (または、統計的パターン認識の識別: 教師付き分類) に相当する、非線形求解法となっていることは明かである。

この対応の理論的研究を体系的に行うことが、前節で提起した非線形多変量解析法の有力な候補になると同時に、多層ニューラルネットの実用上の性質の理論的基盤を確立することにつながる。筆者はここに、ニューラルネット研究の学術的意義を見いだすものである。

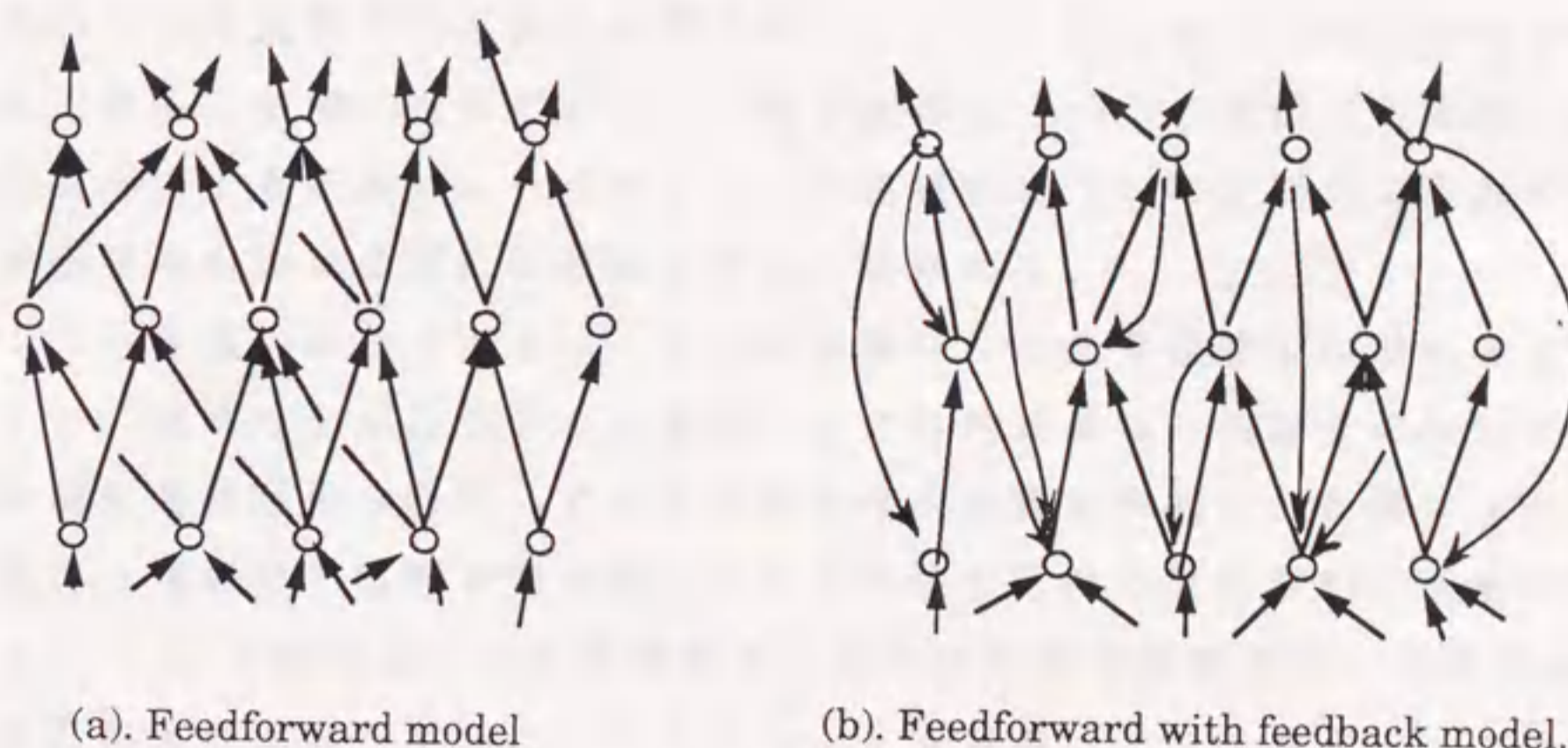


Fig.1.1 Examples of multi-layer neural network.

図1.1 多層ニューラルネットワークの例

1. 3 本研究の位置付け、目的ならびに構成

ニューラルネット研究の全貌は、認知科学、心理学、生理学、解剖学、医学、数学、物理学、情報工学、電子工学等、種々の分野間の相互協力の上に成り立っており、その目的は生体の行う高度で柔軟な情報処理の数理モデルや工学モデルの実現、またそのための各種知見の統合化など極めて広範である(11)。従って、それらは単なるニューラルネットモデルの数理を探ることによって達成されるものではない。しかしながら、現時点(1990年秋)で、そのアーキテクチャならびに学習メカ

ニズムが最もシンプルでありながら、その処理効率、パターン変換能力の柔軟性ゆえに理論、応用、ハードウェア技術ともに最も数多く研究対象とされている多層ニューラルネットを、多変量解析と同じ土俵で対比させることは、今後のニューラルネットモデル研究の方向性を指し示す手がかりにもなると考えられる。また、多層ニューラルネットと多変量解析手法の対応のうち、最近では判別分析に関する対応が、Gallinari et. al. (12), Asoh & Otsu (13) (14)らにより序々に明かにされてきており、上述の理論的基盤は部分的に着手されているところである。しかしながら、それ以外の多変量解析手法との関連については、最も簡単な構成の場合 (15), (16)に限ってのみしか解明されておらず、同手法のもう一つの範疇である外的基準なしの問題に相当する多層ニューラルネットの諸特性については明かにされていない。

本研究の目的はまさにこの点にある。すなわち、

- (1) . 多層ニューラルネットを外的基準なしの多変量解析に対応させる。
- (2) . 同ネットが行うデータ集約（主に外的基準なしの場合の次元削減）のための非線形連続写像の特性、すなわち幾何学的性質を解明し、多層ニューラルネットの構造（層の数と各層の素子数）とその非線形写像の幾何学的関連を明かにする。
- (3) . 1.1節で述べた外的基準なしの非線形多変量解析法、具体的には非線形因子分析およびこれを用いたクラスタリングを多層ニューラルネットで行う具体的手順を示唆する。
- (4) . 外的基準なしの場合の多層ニューラルネットを更に、恒等写像モデル (17), (18)により得られるデータ圧縮写像 T_1 と、同モデルかまたは、筆者らが提案するパルス入力・パターン出力写像モデル（これらのモデルの詳細は2章で述べる）により得られるデータ回帰写像 T_2 の2つの部分に分け、両写像間を関係付ける。

本論文では、このうちの(1)を第2章、(2)を第3章と第4章、(3)を第4章の後半で扱い、(4)を第5章で扱うことにする。

これらの具体的な内容は以下のようなものである。

2章では、多層ニューラルネットを数学的に定式化し、外的基準付きの手法である、判別分析に当てはめる場合のカテゴリー分類の機能（判別分析機能）および外的基準なしの手法に相当する情報圧縮の機能（情報集約機能）を与える。特に後者については、データ圧縮・データ回帰のモデルである砂時計モデルと、筆者らが独自に提案した、データ回帰機能（因子分析の機能に相当する）のみを行う P P N モ

デルを導入し、その数理モデルを規定する。また、現在までにそれらについて行われた内外の研究報告を概観し、本研究の位置付けを確認する。

3章では3層砂時計モデルおよび3層PPNによる写像の幾何学的性質を、出力層素子が線形特性を持つ場合と、Sigmoid特性を持つ場合とに分けて解析し、この両者にK-L展開を加えた3者の情報集約能力の実験的比較を行う。

この結果、以下の事柄が明かにされる。

- ・3層ネットワークによる情報集約は、いずれのモデルの場合も、出力層素子が線形特性の場合にK-L展開の能力を越えられない。
- ・出力層素子がSigmoid特性の場合に、それらの情報集約能力は、標本分布の幾何学的性質に依存して、K-L展開を越えることが有り得る。
- ・3層砂時計モデルと3層PPNとを比較すれば、後者の能力は理論的に前者のそれを上回る。

4章ではこれを4層以上の場合に拡張する。そこで、5層砂時計モデルのデータ回帰写像の能力は、4層PPNのそれと等価になり得ることを示し、続いて両モデルについて3章と同様の理論的考察と実験的検証を行う。ここでの考察は、議論の見通しをよくするため、Sigmoid関数を適当な区分的線形関数で近似して行う。その結果、以下を得る。

- ・これら両モデルによる情報集約の性質は、幾何学的にはある種のPL多様体による近似、すなわち、区分的に線形なデータ回帰問題の求解手法とほぼ等価であると説明できる。
- ・従って、K-L展開と比較してはるかに優れた集約能力を持つ。
- ・5層砂時計モデルと4層PPNとを比較すれば、3層モデルのそれらの比較同様、後者の能力は理論的に前者のそれを上回る。
- ・これらのモデルを用いた具体的な同データ回帰問題の解法、およびこれを用いたクラスタリング、すなわち“類別”の一手法を提示する。

5章では、多層ニューラルネットの諸機能を整理し、個々の機能の相互関係を明らかにする。すなわち、外的基準無しの2手法である、データ圧縮（砂時計モデル）と因子分析（PPN）の幾何学的関連を考察し、これら2手法と、外的基準の有る場合の代表的手法である判別分析（カテゴリー信号を教師データとする多層パーセプトロン）との実験的な比較を行う。

参考文献

- (1). 大津展之: "パターン認識における特徴抽出に関する数理的研究", 電総研研究報告 第818号(1981).
- (2). Anderson, T.W. : "An introduction to multivariate statistical analysis" 2nd Edition, John Wiley & Sons, New York(1984).
- (3). Fukunaga, K. : "Introduction to Statistical Pattern Recognition", Academic Press, New York(1972).
- (4). Duda, R.O. and Hart, P.E. : "Pattern Classification and Scene Analysis", Wiley, New York(1973).
- (5). 大津展之: "パターン認識における統計的学習", 平成2年電気・情報関連学会連合大会論文集, S31-1, 5, pp.75-78(1990).
- (6). Anderberg, M.R. : "Cluster Analysis for Applications", Academic Press, New York(1973).
- (7). 甘利俊一 : "ニューラルネット研究の過去、現在、将来", 人工知能学会誌, Vol.4, No.2, pp.120-127(1989).
- (8). Rosenblatt, F. : "Principles of Neurodynamics", Spartan(1961).
- (9). Minsky, M and Papert, S : "Perceptron", MIT Press(1969).
- (10). Rumelhart, D. E. and McClelland, J. L. : "Parallel distributed processing", Vol.1, MIT Press(1986).
- (11). 麻生英樹 : "ニューラルネットワーク情報処理", 産業図書(1988).
- (12). Gallinari, P., Thiria, S. and Fogelman Soulie, F. : "Multilayer Perceptrons and Data Analysis", Proc. ICNN-88, 1, pp.391-399(1988).
- (13). Asoh H. and Otsu N. : "Nonlinear Data Analysis and Multilayer Perceptrons", Proc. IJCNN-89, 2, pp.411-415(1989).
- (14). Asoh H. and Otsu N. : "An Approximation of Nonlinear Discriminant Analysis by Multilayer Neural Networks", Proc. IJCNN-90-S.D., 3, pp.211-216(1990).
- (15). 船橋賢一: "三層ニューラルネットワークによる恒等写像の近似的表現について

の理論的考察”, 信学技報, MBE88-174, pp.109-114(1989).

(16).Bourlard, H. and Kamp, Y.: " Auto-Association by Multilayer Perceptrons and Singular Value Decomposition", Biol. Cybern. 59, pp.291-294(1988).

(17).Elman, J.L. and Zipser, D.: "Learning the hidden structure of speech", Tech.Rep. 8701, Univ.of California, San Diego, Institute for Cognitive Science(1987).

(18).Cottrell, G.W., Munro, P. and Zipser, D. : "Image Compression by Back Propagation: An Example of Extensional Programming", Tech.Rep. 8702, Univ. of California, San Diego, Institute for Cognitive Science(1987).

第2章

多層ニューラルネットの定義と機能

以下の各節では、次章以降の考察の対象となる多層ニューラルネットモデルの概念と機能に関する数学的準備を与える。

2.1 多層ニューラルネットによる判別分析機能と情報集約機能

多層ニューラルネット（層の数を n とする）は、入力層（第1層）、出力層（第 n 層）、および、一つ以上のかくれ層（第2, … 第 $n-1$ 層）から成り、各層は a_i ($i=1, 2, \dots, n$) 個の入出力素子を持つような層状のネットワークとして定義される(1)(2)。また、一つの層の内にある素子の数をその層のサイズとよぶことにする。ここで、各層内の素子は各々前段層、後段層にある素子のみと接続され、信号は前段より後段へと伝搬する。また個々の素子は多入力-出力の形をとり、素子間の各結合は各々重みパラメータ W を、また入力層を除いて各層のすべての素子は個々にしきい値 θ を持つ。このとき、各素子の入出力特性は各入力値を x_i 、出力値を y として式(2-1)で表される。ここで関数 F は単調増加で微分可能な一価関数である必要がある(2)。そこで、以下本文では式(2-2)で表される sigmoid 特性もしくは式(2-3)の線形特性を用いる。

$$y = F(\sum_i W_i x_i - \theta) \quad (2-1).$$

$$F(x) = 1 / (1 + \exp(-x)) \quad (2-2).$$

$$F(x) = x \quad (2-3).$$

ここで、誤差逆伝搬（学習）法とは、このネットワークの入力層に第 n サンプル ($n=1, \dots, N$) の入力信号 I_n を、出力層（第 M 層とする）に第 n 教師信号 A_n を例示して、その入力 I_n を受けた際の出力層からの出力信号 O_n と与えられた教師信号 A_n との差の（ノルムの）二乗誤差 e_n の総和 $e = \sum_n e_n = \sum_n |O_n - A_n|^2$ を最急

降下法により最小化するように、各 W および θ を修正していく方法である(2)。

以下これを具体的に示す。いま第 n サンプル提示の際、図2-1において第 $m-1$ 層第 i 素子から第 m 層第 k 素子への結線の重みパラメータを $W^{m,i,k}$ 、第 m 層第 i 素子のしきい値を $\theta^{m,i}$ で表すとき、それらの修正量 $\Delta W^{m,i,k}$ 、 $\Delta \theta^{m,i}$ を、

$$\Delta W^{m,i,k} = -\alpha \left(\partial e_n / \partial W^{m,i,k} \right) \quad (2-4).$$

$$\Delta \theta^{m,i} = -\alpha \left(\partial e_n / \partial \theta^{m,i} \right) \quad (2-5).$$

により決定する(但し α は小さな正数で修正係数と呼ばれる)。ここで、第 m 層第 i 素子への入力総和を $S^{m,i}$ 、そこからの出力値を $S'^{m,i}$ とすると、

$$S'^{m,i} = F(S^{m,i}) \quad (2-6).$$

$$S^{m,i} = \sum_j W^{m,j,i} S'^{m-1,j} \quad (2-7).$$

が成り立つ。一方、同素子への入力 $S^{m,i}$ の変化に伴う誤差 e_n の変化分を

$$g^{m,i} = \partial e_n / \partial S^{m,i} \quad (2-8).$$

と置けば、これを用いて式(2-4)の括弧内の項は

$$\partial e_n / \partial W^{m,i,k} = \left(\partial e_n / \partial S^{m,k} \right) S'^{m-1,i} = g^{m,k} S'^{m-1,i} \quad (2-9).$$

と表せるから、

$$\Delta W^{m,i,k} = -\alpha g^{m,k} S'^{m-1,i} \quad (2-10).$$

但し、 $g^{m,k}$ と $g^{m+1,j}$ との間には次式が成り立つ。

$$g^{m,k} = \left(\sum_j W^{m+1,k,j} g^{m+1,j} \right) f(S^{m,k}) \quad (2-11).$$

ここで、 $f(x) = dF(x) / dx$ である。

すなわち、逆伝搬学習法とは、式(2-10)、(2-11)に従って、各素子における誤差への寄与分 $g^{m,k}$ を後段から前段へと伝搬し、その過程で重みパラメータ W を修正していく方法である(しきい値 θ についても同様の操作を行う)。

なお、入力信号 I_n 、出力信号 O_n および教師信号 A_n は通常各素子の出力値を変数とする多次元ベクトルで表記されるが、本文ではこれらをパターンベクトルあるいは単にパターンとよぶことにする。

この、多層ニューラルネットおよび誤差逆伝搬学習法の本来の使用目的は、入力信号にサンプルパターンを、教師信号にそのカテゴリー情報を何等かの方法で数値化したベクトルを与えて、パターン空間とカテゴリー空間の間の連続写像(隣合う層の間の写像は、入出力特性が式(2-2)または(2-3)で表されるような各素子の出力値を成分とするベクトル間の写像であることから連続写像であり、これを結合した

写像もまた連続写像となる)を学習により獲得すること、換言すれば、判別関数(Discriminant Function)の学習にある。これが多層ニューラルネットが本来、多層パーセプトロン(Perceptron)と呼ばれるゆえんである。この問題に関しては、Asoh and Otsu(3)(4)、Gallinari, et al.(5)により考察されており、それらによれば、多層ニューラルネットにより得られる判別関数は一般に非線形判別関数となる(2層ニューラルネット、および中間層が線形な3層ニューラルネットで線形判別、それ以外では非線形判別)ことが明かとなっている。これは多変量解析における外的基準を教師信号として与える場合の問題に対応している(6)。

一方、本論文で対象とする、外的基準無しの問題を同様の多層モデルで扱う試みもこのところ増えてきており、データ圧縮ならびにパターンの集約表現の獲得を目的とした、入力と教師信号の両方に同一のパターンデータを例示する、砂時計モデル(パターン自身の恒等写像をおこなう)といわれるものが知られている(7)-(15)。以下にこの方式を簡単に説明し、加えて、これとは別に、集約表現の獲得に関して有効な計算モデルであるパルス入力パターン出力回路(Pulse-in Pattern-out Network: PPN)を導入する。

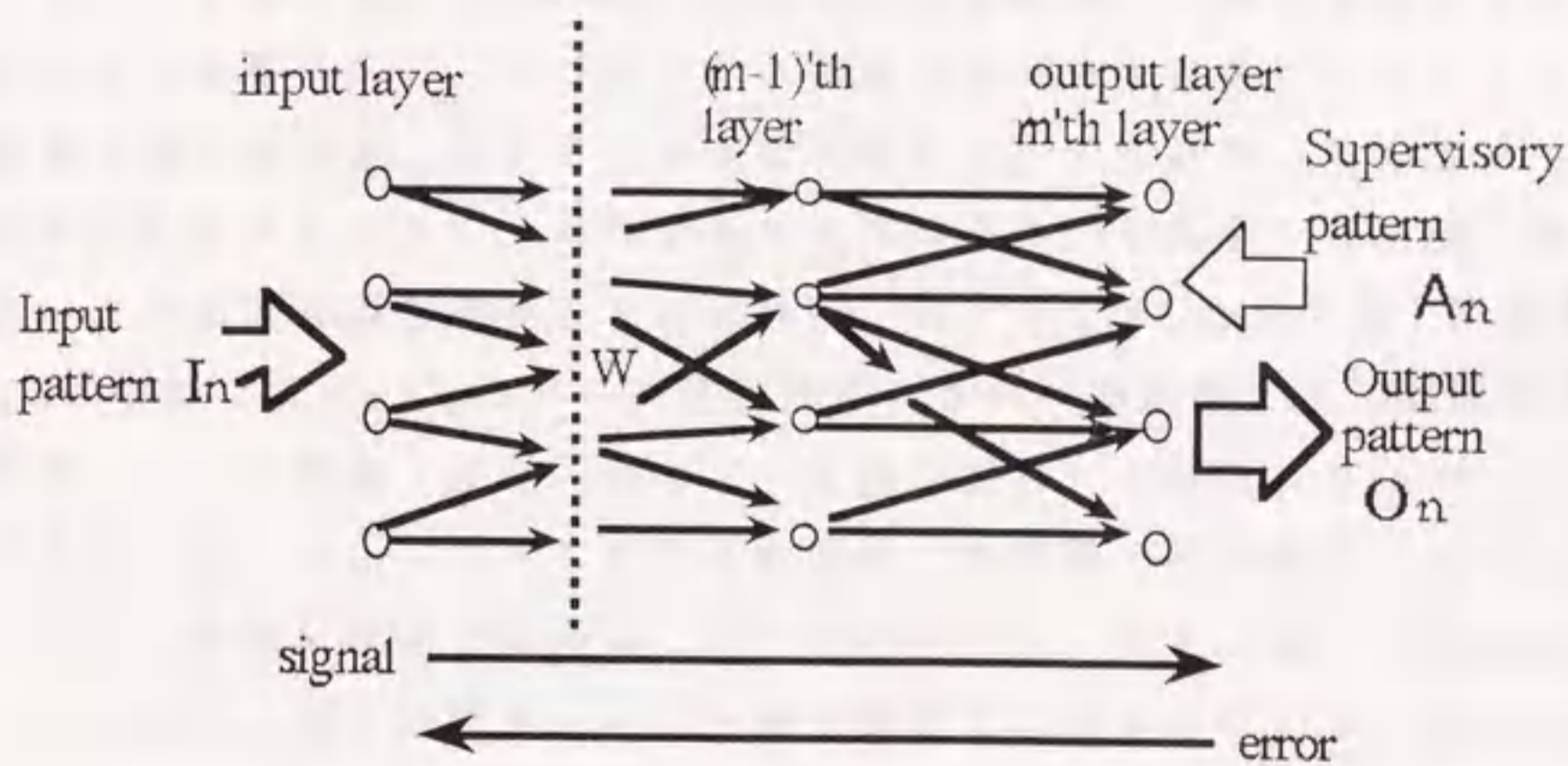


Fig.2.1 Multi layer neural networks and error back propagation

図2.1 多層ニューラルネットと誤差逆伝搬法

2. 2 恒等写像を近似する砂時計モデル

多次元データの圧縮に数理的な定義を与えれば、以下のようなになる（図 2-2）。すなわち、多次元データ圧縮とは、有限個のベクトル $\mathbf{x}_k \in R^M$ （但し $k = 1, \dots, K$ ）を一組の多変量データ、有限個のベクトル $\mathbf{y}_k \in R^D$ （ $M > D$ ）をそれより少ない変数からなる多変量データとすると、ある正数 ρ に対し、

$$T1: \mathbf{x}_k \rightarrow \mathbf{y}_k \quad (2-12)$$

$$T2: \mathbf{y}_k \rightarrow \mathbf{x}_k' \quad (2-13)$$

$$\sum_k |\mathbf{x}_k - \mathbf{x}_k'|^2 < \rho^2 \quad (2-14)$$

を満たす写像 $T1$ 、 $T2$ が存在する仮定のもとで、 $T1$ を指して言う。ここで、 \mathbf{y}_k を圧縮表現（あるいは集約表現）、写像 $T2$ を \mathbf{y}_k から原データ \mathbf{x}_k への回帰と呼ぶ。また、次式の ε^2 （2乗誤差総和）をデータ圧縮の際の情報損失量を表す指標とする。

$$\varepsilon^2 = \sum_k |\mathbf{x}_k - T2(T1(\mathbf{x}_k))|^2 \quad (2-15)$$

そこで、 $\mathbf{x}_k, \mathbf{y}_k, \mathbf{x}_k'$ を（第 k 標本を提示する際の）各々、入力層、いずれかの中間層、出力層の各素子の出力値を成分とするベクトルとみなした場合の多層ニューラルネットワークを考えれば、これは、入力信号と教師信号の両方に同一の M 次元パターンデータの組を例示し、いずれかの中間層サイズ D を M より小さくとして恒等写像を学習（逆伝搬学習）させるモデルに対応する。すなわち、入力層サイズ（＝出力層サイズ）よりも小さく、かつ最もサイズの小さい中間層を集約層、そこでの出力パターンを集約表現と呼ぶことにすれば、原データ \mathbf{x}_k は入力層・集約層間で集約表現 \mathbf{y}_k に変換（式(2-12)のデータ圧縮操作）され、集約層・出力層間で再び原データの近似 \mathbf{x}_k' へと変換（式(2-13)のデータ回帰操作）される。そこで、入力層・集約層間の部分的なネットワークを符号化回路、集約層・出力層間のそれを復号化回路と呼ぶ。この種の計算モデルは、自己想起モデル、あるいはその形状から砂時計モデルと呼ばれ、既に各種データの圧縮や解析に応用されている。以下にその具体的な背景を示す。

砂時計モデルは、その構造が最も単純な場合の3層モデルについて Cottrell et al. (7)による画像データ圧縮の試みで具体的に提案されて以来、Elman et al. (8)による音声データ解析の試みや、曾根原らの画像データの解析(9)、長坂らの心電図の解析(10)で実験的検証が行われており、Bourlard and Kamp(11)、船橋(12)に

より、出力層素子が線形特性を持つと仮定したときの圧縮効率の限界（最小解に収束したときに得られる）が主成分分析のそれに一致することが示唆されている。

また、主成分分析を上回る圧縮能力が期待される5層砂時計モデルのデータ圧縮については、入江ら(13)（多様体上の標本分布の次元削減）、大山ら(14)（多様体上の標本分布の次元削減）、臼井ら(15)（マンセル色空間の特徴解析）、によりシミュレーション実験の結果が報告されており、線形の主成分分析に対する優位性が示されているが、いずれもその詳細な数理的性質の解析には至っていない。

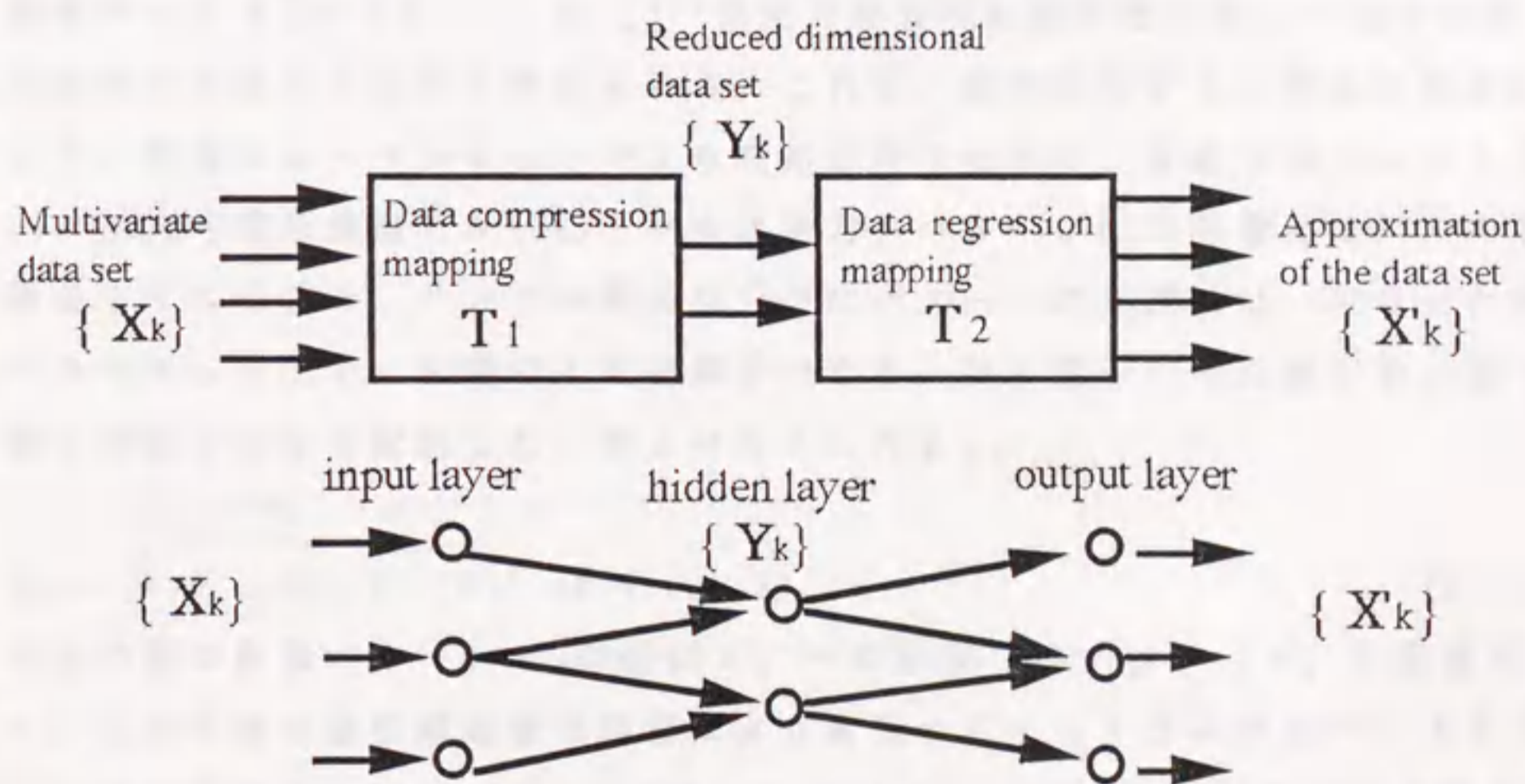


Fig.2.2 Data compression and regression mapping T_1, T_2 (above) and auto association networks (below).

図2.2 データ圧縮, データ回帰写像 T_1, T_2 (上) と砂時計モデル (下)

2. 3 パルス入力・パターン出力写像を近似する P P N モデル

与えられた一組の多変量データの統計的性質に着目し、これらを比較的少数の共通な潜在変数（因子とも呼ぶ）の線形結合 ϕ で集約的に表現する操作は、外的基準のない場合の多変量解析手法の一つで因子分析モデルと呼ばれる(6)。すなわち、個体 k についての変数 m の観測値を x_{km} 、潜在変数（因子スコア）を y_{k1}, \dots, y_{kD} 、これらにかかる因子負荷を $\phi_{m1}, \dots, \phi_{mD}$ として、線形回帰式

$$x_{km} = \sum_d (\phi_{md} y_{kd}) + \varepsilon_{km} \quad (2-16).$$

で表記される。但し ε_{km} は近似誤差を表す。

式(2-13)のデータ回帰写像 T_2 は、潜在変数 $\mathbf{y}_k = (y_{k1}, \dots, y_{kD})^t$ を集約層表現、多変量データ $\mathbf{x}_k = (x_{k1}, \dots, x_{km})^t$ を出力層表現に置き換えてこの因子分析モデルを非線形に拡張したものと考えられる。これを、砂時計モデルと同様の表現形式を用いて、多層ニューラルネットにより簡略に行うために、著者らはパルス入力・パターン出力写像を提案した(16)。パルス入力・パターン出力写像とは、入力層、出力層のサイズを各々、サンプル数 K ならびにパターン次元数 M に一致させた多層ニューラルネット上で、 K 個の入力層素子のうち、第 k 素子のみに値 1 を、残り $K-1$ 個の全素子に 0 を配置した、第 k パルス入力 \mathbf{p}_k ;

$$\mathbf{p}_k = \underbrace{(0, 0, \dots, 0, 1, 0, \dots, 0)}_{(k-1)\text{個}}^t \quad (k=1, 2, \dots, K) \quad (2-17).$$

から出力層の教師パターン \mathbf{x}_k の近似 \mathbf{x}_k' への写像 $T_3: \mathbf{p}_k \rightarrow \mathbf{x}_k'$ と定義される。また、この写像を逆伝搬法等の学習により実現するネットワークをパルス入力・パターン出力回路または Pulse-in Pattern-out Network の意味で P P N と呼ぶ。砂時計モデル同様、最もサイズの小さい中間層（これは入力層サイズ K ならびに出力層サイズ M より小さい）を集約層と呼び、そこでの出力パターンを集約パターンと呼ぶことにすれば、集約層と教師パターンの提示される出力層の間で式(2-13)の因子分析モデルを実現する（図 2-3）。すなわち、P P N は各々が直交しあう疑似入力パターンであるパルス入力を、素子数の少ない第 2 層を介して、出力層の教師パターン（を近似したもの）へ変換する能力 T_2 を学習により獲得し、結果的には、

$$\varepsilon^2 = \sum_k |\mathbf{x}_k - T_2'(\mathbf{y}_k)|^2 \quad (2-18).$$

を最小化するかたちで、 $\mathbf{x}_k \in R^n$ を $\mathbf{y}_k \in R^D$ ($D \leq M$) で集約表現している。

これらのことから、砂時計モデルがデータ圧縮写像 T_1 とデータ回帰写像 T_2 の両方を学習するニューラルネットであるのに対し、PPN はデータ回帰写像 T_2' のみを学習するという点で、砂時計モデルの部分ネットワークであると考えられる。砂時計モデルと同様、PPN の集約層・出力層間の部分的ネットワークをも復号化回路と呼ぶことにする。

尚、データ圧縮写像 T_1 とデータ回帰写像 T_2 が、近似的には互いに逆写像の関係となることは容易に知られ、これらに線形性を仮定した場合、一方の計算結果から他方を求めることは容易である。実際、線形の変量解析で最小 2 乗誤差を評価関数とするとき、 T_1 を求めることにより T_1, T_2 の両方を獲得する手法は主成分分析法(6)としてよく知られている。

ここで、恒等写像モデルと PPN において、多次元標本データの集約を行う際、層数を一定にして、より小さい集約層サイズで、良い近似精度を得るために、両ネットワークの構成は以下のようにとれば有利であることが予想される。

1)、砂時計モデル：入力層・集約層間のデータ圧縮写像 T_1 と同程度の能力をデータ回帰写像 T_2 に与えるために、入力層・集約層間に含まれる層の数および各層サイズと、集約層・出力層間のそれらを等しくとる。従って、通常、集約層を中心として左右対称の構造を持たせることが望ましい。これは、3層砂時計モデルについては文献(7)-(10)、5層砂時計モデルについては文献(13)-(15)で示されている。

また、既に Irie and Miyake (17)、船橋(18)が明らかにしているように、3層のニューラルネットを用いて中間層サイズ(中間層素子は Sigmoid 特性)を増やすことで、理論的には、任意の連続写像の近似 ($R^n \rightarrow R^n$) を任意の精度で実現し得る。

このことは、最小解への収束結果が得られれば、5層ニューラルネット(入力層-集約層間、集約層-出力層間がともに3層から成るモデル)を用いて、第2層、第4層のサイズを増やすことにより、任意の(有限個の標本からなる)多次元標本集合を、任意の精度で(近似誤差をいくらでも少なく)任意の次元に削減することが可能であることを示唆する。

2)、PPN：集約層-出力層の間は上と同様のデータ回帰写像を近似する回路と考えられるが、PPN では各標本に対応する個々のパルス入力が集約表現に直接に写像される、いわゆるテーブルルックアップ方式(従ってデータ圧縮の操作を行っていない)となっているため、第2層出力は各標本に対して独立に任意の値に設

定できる（これを、入力パターンどうしが干渉しないという）。従って、第2層において任意の集約表現 y_k を生成することができることから、通常 P P N においては、ネットワーク全体の層数によらずに、集約層を第2層目に設定することができる。

また、これらのことから、復号化回路が層数 n の同一構成をもつような、全体層数 $(2n - 1)$ の砂時計モデルと、全体層数 $(n + 1)$ の P P N は原理的には（極小解の存在に対する考慮を別にすれば）ほぼ同程度のデータ回帰能力をもつと予想される(19)。

しかしながら、次章で明らかにされるように $n = 2$ の場合の3層砂時計回路と3層 P P N の比較では、学習過程の効率の差によって、一般に後者の集約能力が前者のそれを越える。これに加えて、出力層にシグモイド特性を付加した3層 P P N では、標本分布の幾何学的性質に依存して主成分分析の集約能力をも越える結果を与える(16)。

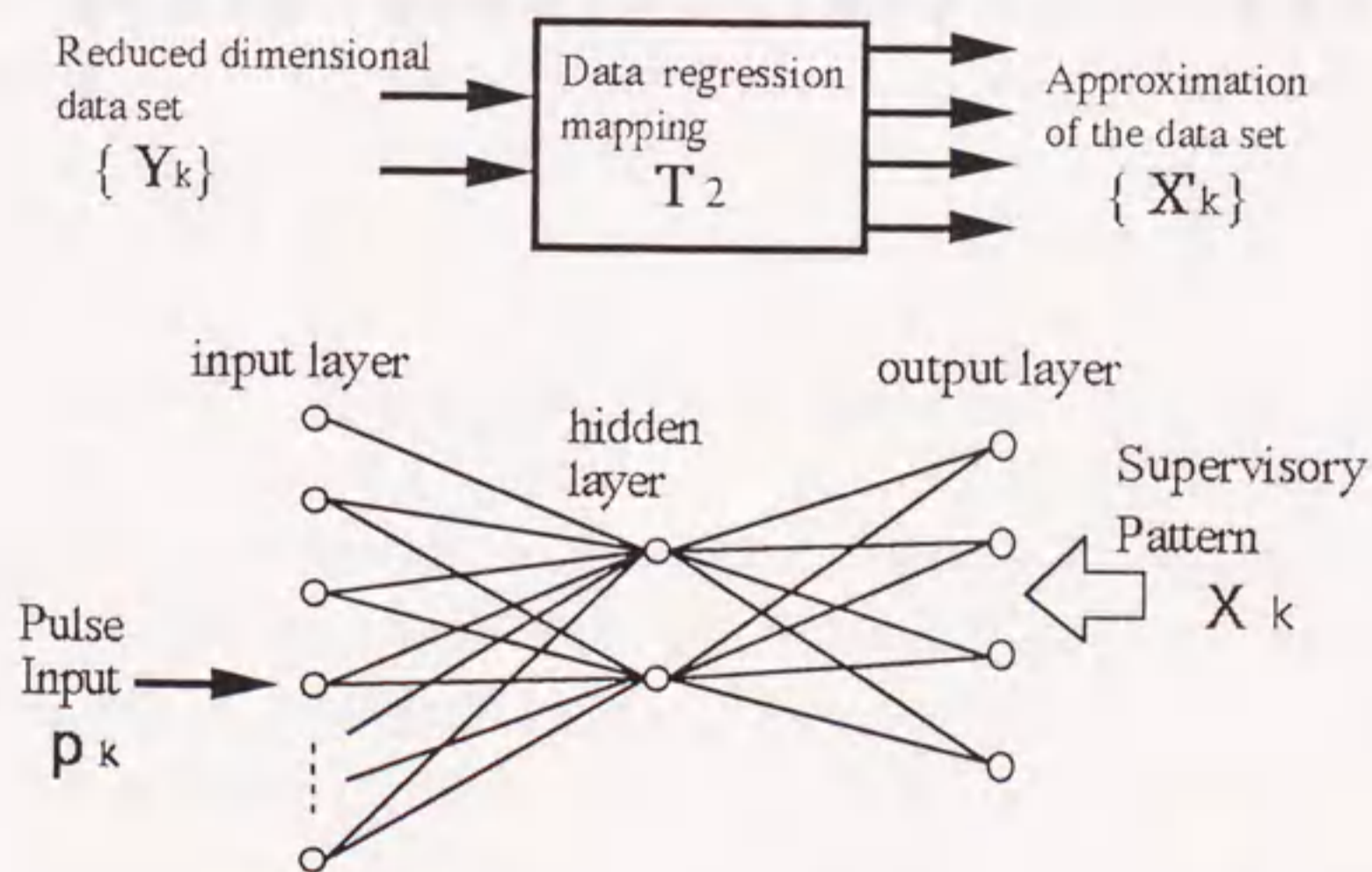


Fig. 2.3 Factor analysis model (above) and Pulse-in Pattern-out Networks (below)

図2.3 因子分析モデル(上)とパルス入力パターン出力型ネットワーク(下)

2. 4 まとめ

本章では、多層ニューラルネットおよびその学習法として広く知られる誤差逆伝搬 (Error-Back-Propagation) 法について紹介し、パターン認識における、同モデルの諸機能と実際の利用法について述べた。すなわち、以下に示す3機能である。

- ・従来からの、パターン空間からカテゴリ空間への写像であるクラス判別機能。
- ・パターン空間をより低次元の空間に写像し、これを元のパターン空間へ戻すデータ圧縮およびパターン回帰の機能 (線形手法の主成分分析に対応)。
- ・互いに直交する疑似入力である、パルス入力を集約空間へ射影し、そこからパターンデータへの回帰写像のみを獲得する因子分析 (パターン回帰) 機能。

こうして、多変量解析の幾つかの代表的手法とそれに相当する多層ニューラルネットの機能面での対応が本章の議論により明かにされた。

そこで、次に明かにすべき事項としては、これらの数理的対応がある。次章では、まず、最も簡単な3層砂時計モデルと3層PPNについて多変量解析との関連を明かにする。

参考文献

- (1). 甘利俊一 : "神経回路網の数理", 産業図書(1978).
- (2). Rumelhart, D. E. and McClelland, J. L. : "Parallel Distributed Processing", Vol. 1, MIT Press(1986)
- (3). Asoh H. and Otsu N. : "Nonlinear Data Analysis and Multilayer Perceptrons", Proc. IJCNN-89, 2, pp.411-415(1989).
- (4). Asoh H. and Otsu N. : "An Approximation of Nonlinear Discriminant Analysis by Multilayer Neural Networks", Proc. IJCNN-90-S.D., 3, pp211-216 (1990).
- (5). Gallinari, P., Thiria, S. and Fogelman Soulie, F. : "Multilayer Perceptrons and Data Analysis", Proc. ICNN-88, 1, pp.391-399(1988).
- (6). Anderson, T. W. : "An Introduction to Multivariate Statistical Analysis" 2nd Edition, John Wiley & Sons, New York(1984).
- (7). Cottrell, G. W., Munro, P. and Zipser, D. : "Image Compression by Back Propagation: An Example of Extensional Programming", Tech. Rep. 8702, Univ. of California, San Diego, Institute for Cognitive Science(1987).
- (8). Elman, J. L. and Zipser, D. : "Learning the Hidden Structure of Speech", Tech. Rep. 8701, Univ. of California, San Diego, Institute for Cognitive Science(1987).
- (9). 曾根原登、川入光男、三宅誠、中根一成 : "ニューラルネットによる画像データ圧縮(Neuro-CODEC)の検討", 信学技報, IE88-62, pp. 57-64(1988).
- (10). 長坂保典、岩田彰、鈴木宣夫 : "ニューラルネットによる長時間心電図の特徴抽出と情報圧縮", 信学技報, MBE88-91, pp. 73-80(1988).
- (11). Bourlard, H. and Kamp, Y. : "Auto-Association by Multilayer Perceptrons and Singular Value Decomposition", Biol. Cybern. 59, pp.291-294(1988).
- (12). 船橋賢一 : "三層ニューラルネットワークによる恒等写像の近似的表現についての理論的考察", 信学技報, MBE88-174, pp.109-114(1989).
- (13). 入江文平、川入光男 : "多層パーセプトロンによる内部表現の獲得 - テーブ

ルルックアップ法と違うのか？ -", 信学技報, NC89-15, pp. 33-40 (1989).

(14). 片山泰男、大山公一: "自己組織逆伝搬ニューラルネットの諸特性", 平成元年信学春大, 7, pp. 309-310 (1989).

(15). 臼井支朗, 中内茂樹, 中野正恵、: "多層ニューラルネット恒等写像モデルによるマンセル色空間の特徴解析", 信学技報, NC89-40, pp. 45-50 (1989).

(16). 米倉達広、横井茂樹、鳥脇純一郎: "多層ニューラルネットによる情報集約の一手法とその理論的考察", 信学論(D-II), J73-D-II, 8, pp. 1205-1212. (1990).

(17). Irie, B. and Miyake, S. : "Capabilities of Three-layered Perceptrons", Proc. IEEE ICNN-88, pp. 1641-647 (1988).

(18). 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, pp. 127-134 (1988).

(19). 宮崎慎也、米倉達広、鳥脇純一郎: "サンプル分布の幾何学的構造の解析における自己想起型回路と PPN の関連について", 信学技報, NC89-102, pp. 73-78 (1990).

第 3 章

3 層 P P N の幾何学的性質

前章で与えた砂時計モデルおよび P P N は、教師信号としてパターンデータを例示し、その想起 (Recollection) を逆伝搬法で学習するという点で共通である。以下では、この種のネットワークの内部表現を、最も簡単な 3 層の場合について考察し、多変量データを、少ない数の成分で効率よく表すというデータ集約のための成分分解の共通的性質について解析する。なお、この種の問題を理論的に考察した例として、Bourlard and Kamp(1)、船橋(2)、Baldi and Hornik(3)らによるものが挙げられるが、これらはいずれも素子特性 (特に出力層) を線形なものに限定した場合の考察である。ここではこれを線形の場合とシグモイドの場合に分けて議論を進める。

3. 1 パターンの成分分解

まず、出力層でのパターンがネットワーク内でいかに表現されうるかを調べるため、前章で定義した 3 層砂時計モデルあるいは 3 層 P P N について、サイズ v から成る出力層 (第 3 層) と、サイズ u から成る第 2 層に現れるパターンの関係に着目する (図 3-1)。

いま、第 2 層の第 p 素子 ($p=1, 2, \dots, u$) のみが発火 (出力値 1 を出し)、残り $u-1$ 個の素子がすべて出力値 0 をとる場合に出力層に現れるパターンを \mathbf{h}_p 、第 $n-1$ 層の素子が全く発火していない場合に出力層に現れるパターンを \mathbf{h}_0 とすれば、素子特性として式 (2-1) の F を用いて次式を得る。

$$\mathbf{h}_0 = (h_{01}, \dots, h_{0v})^t ; \text{ただし } h_{0q} = F(-\theta_q) \quad (3-1)$$

$$\mathbf{h}_p = (h_{p1}, \dots, h_{pv})^t ; \text{ただし } h_{pq} = F(W_{pq} - \theta_q) \quad (3-2)$$

ここで、 W_{pq} は第 2 層第 p 素子と出力層第 q 素子の間の結合荷重である。

また、一般に出力層におけるパターンを $f = (f_1, \dots, f_v)^t$ 、第2層の集約表現を $g = (g_1, \dots, g_u)^t$ とすれば、

$$f_q = F \left(\sum p (W_{pq} g_p) - \theta_q \right) \quad (3-3)$$

この式は、式(3-1), (3-2)の $h_{\theta q}$ 、 h_{pq} の各要素を用いて、

$$f_q = F \left[\sum p \{ g_p (F^{-1}(h_{pq}) - F^{-1}(h_{\theta q})) \} + F^{-1}(h_{\theta q}) \right] \quad (3-4)$$

あるいは、個々の要素に対称な形

$$F^{-1}(f_q) - F^{-1}(h_{\theta q}) = \sum p [g_p \{ F^{-1}(h_{pq}) - F^{-1}(h_{\theta q}) \}] \quad (3-5)$$

に書き直すことができる(ただし F^{-1} は F の逆関数)。ここで、関数 F^{-1} は、 F が式(2-2)に代表される飽和型関数(有界な値域をもつ単調関数)の場合に有界な定義域をもつ可微分一価関数となり、式(2-3)の線形関数の場合には同じく線形関数となることが知られる(その場合には式(3-4), (3-5)中の F, F^{-1} は削除される)。

この式は、出力パターン f の各要素が、前層でのパターン g の各要素と、そこでの素子個々の出力層への寄与 h_{pq} および個々の出力素子のバイアス $h_{\theta q}$ のみによってきまることを意味する。

いま、 K 個の教師パターン $\{ f^k : (f^k_1, \dots, f^k_v)^t \mid k=1, \dots, K \}$ について行う逆伝搬学習の過程を考える。このときネットワークに対しては、逆伝搬法の性質から、各々のサンプルについての、反復回数 t の関数としての出力パターン $\{ f_k(t) : (f_{k1}(t), \dots, f_{kv}(t))^t \mid k=1, \dots, K \}$ と教師パターンの二乗誤差の総和

$$E(t) = \sum k \{ e_k(t) \}^2 = \sum k \sum q \{ f^k_{q} - f_{kq}(t) \}^2 \quad (3-6)$$

またはこれに式(3-4)を代入したもの(この過程における第 $n-1$ 層の出力の各パターン $\{ g_k : (g_{k1}, \dots, g_{ku}) \mid k=1, \dots, K \}$, 同層の各素子単独での出力層への寄与 h_{pq} , バイアス $h_{\theta q}$ を全て反復回数 t の関数とみなす)

$$E(t) = \sum k \sum q \left[f^k_{q} - F \left[\sum p \{ g_{kp}(t) (F^{-1}(h_{pq}(t)) - F^{-1}(h_{\theta q}(t))) \} + F^{-1}(h_{\theta q}(t)) \right] \right]^2 \quad (3-7)$$

を減少するように、各 w, θ が反復修正される。それ故、与えられた K 個の教師パターン f_1, f_2, \dots, f_K のネットワークによる学習は、二つの過程

1). 第2層と出力層の間で行われる、基底パターン $(F^{-1}(h_{pq}) - F^{-1}(h_{\theta q}))$ の張りかたの最適化:

2). 入力層と第2層の間で行われる、各教師パターンを最適に近似するような、上記の基底系における係数 g_{kp} の計算:

の並列したプロセスと解釈することができる。

ここで、 $\alpha_{pq} = F^{-1}(h_{pq}) - F^{-1}(h_{0q})$, $O(f_{kq}) = F^{-1}(f_{kq}) - F^{-1}(h_{0q})$ とおくと、式(3-5)は、

$$O(f_{kq}) = \sum_k g_{kp} \alpha_{pq} \quad (3-8)$$

の形に書かれる。そこで、サンプルに依存しない共通項である α_{pq} を以降、表現素 (primitive) と呼び、各パターン $f_k = (f_{k1}, \dots, f_{kv})$ の第 $n-1$ 層での表現 $g_k = (g_{k1}, \dots, g_{ku})$ をその層における成分パターンと呼ぶ。更に、各教師パターンを、それを合成する基になる各層成分パターンへ変換する写像を、その層への成分分解と呼ぶ。以下、成分パターンに関して砂時計モデルと PPN を比較する。

砂時計モデルと PPN の各々において形成される成分パターンの数理的性質を調べる。まず、 n 層から成る砂時計モデルおよび PPN において、任意の層 (第 a 層) の成分パターン $g^a_k = (g^a_{k1}, \dots, g^a_{kp})$ と、その前段層の成分パターン $g^{a-1}_k = (g^{a-1}_{k1}, \dots, g^{a-1}_{kq})$ との間には、式(3-3)と同様の関係が成り立つ。ここで、砂時計モデルと PPN との相違は、入力層へのパターンの与え方である。すなわち、両回路での入力層における、第 k 番目のサンプルについての成分パターン $g^1_k = (g^1_{k1}, \dots, g^1_{kR})$ (R は入力層サイズ) は、前者では教師パターンと同じ $f^1_k = (f^1_{k1}, \dots, f^1_{kR})$ 、後者ではパルス入力 $p_k = (p_{k1}, \dots, p_{kK}; p_{kr} = 1 \dots (r=k), p_{kr} = 0 \dots (r \neq k))$ (ただし入力層サイズ = K) となる。従って、両回路における第 2 層での成分パターンの各要素 $g^2_{kr} = F [\sum_p \{g^1_{kp} W^2_{pr}\} - \theta^2_r]$ は、

$$g^2_{kr} = F [\sum_p \{f^1_{kp} W^2_{pr}\} - \theta^2_r] \\ : (f^1_k, \{W^2\}, \theta^2_r) \text{ の関数} \quad (\text{砂時計モデル}) \quad (3-9)$$

$$g^2_{kr} = F [\sum_p \{p_{kp} W^2_{pr}\} - \theta^2_r] = F [W^2_{kr} - \theta^2_r] \\ : (W^2_{kr}, \theta^2_r) \text{ の関数} \quad (\text{PPN}) \quad (3-10)$$

ただし、 W^2_{ab} は、入力層第 a 素子と第 2 層第 b 素子の間の結合係数、 θ^2_c は第 2 層第 c 素子のしきい値を表す。

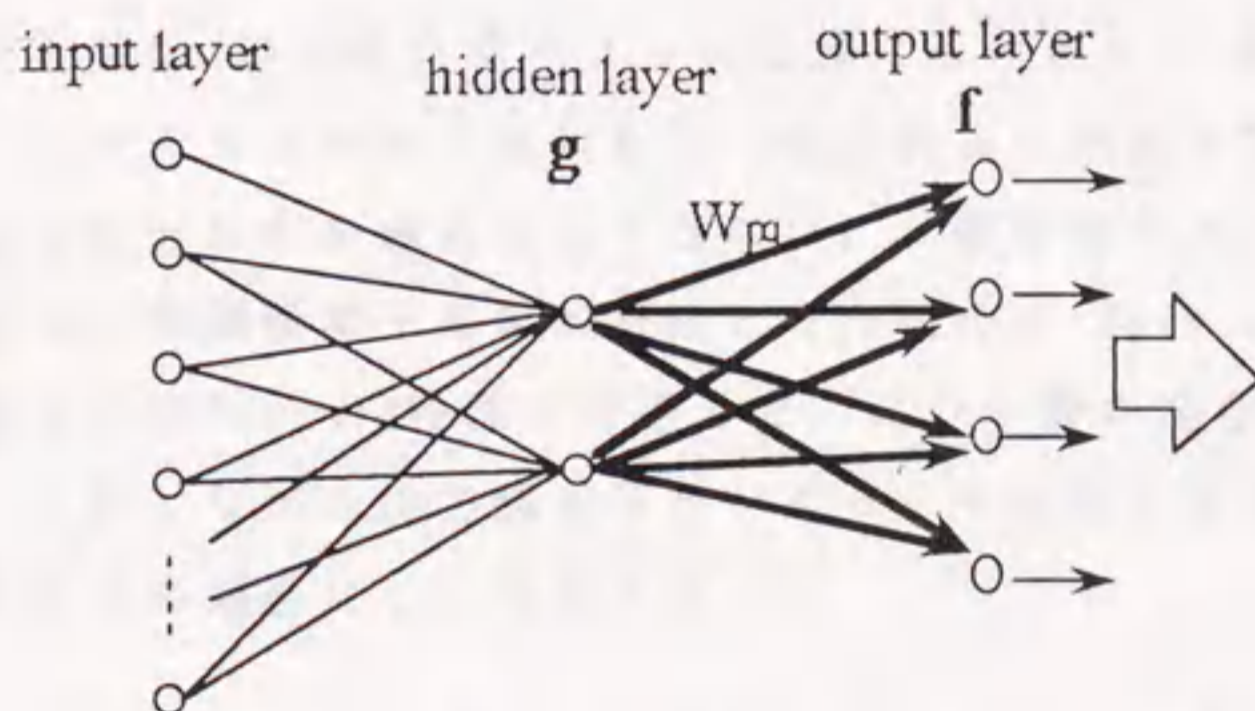
従って、砂時計モデルの場合、第 2 層成分パターンの各要素は、その素子に結合する全ての結合係数および教師パターンの関数となるのに対し、PPN の場合、同要素はその素子と入力層の第 k 素子 (サンプル番号に対応する素子) の間の結合係数 1 個のみから決まることが知られ、以下を得る。

(1) 入力層・第 2 層間の結線 1 本に注目した場合、その重みの修正は、PPN で

はサンプルのどれか1個のみの学習の際に生じ、それ以外のサンプルの学習時には生じ得ない（他のサンプルの学習による束縛を受けない）のに対し、砂時計モデルではその修正は全サンプルについて生ずるので、各結合の強化の是非についてサンプル間で競合が生じ得る。

(2) PPNでの入力パターンどうしは直交する。このため、第2層での成分パターンは、PPNでは後段の層から伝搬される誤差の吸収のためにのみ形成され、入力層からの拘束を受けない（従って、より望ましい成分パターンが組織され易い）のに対し、砂時計モデルでは後段の層からと同時に入力層からの拘束をも受ける。

PPNのこの性質は、層数が4以上の場合においても成り立つ。これらより、一般には同一数の層、同一（入力層サイズを除いて）の各層サイズからなる両回路において、砂時計モデルはPPNに比べて集約効率の点で不利となる。この傾向は、画像データ圧縮等の実験例でも報告されている(4)。



pattern g of the hidden layer is mapped to pattern f of the output layer, where $f = F(W_{gf}g - \theta)$

Fig.3.1 Three layer PPN and its component decomposition.

図3.1 3層PPNとその成分分解の説明

3. 2 出力層素子が線形特性をもつ場合

素子の入出力特性 F を式(2-3)で表される線形特性とすると、第 k 標本の出力パターン(第 q 成分) f_{kq} は、式(3-3)から

$$f_{kq} = \sum p (W_{pq} g_{kp}) - \theta_q \quad (3-11)$$

あるいは、 $\mathbf{f}_k = (f_{k1}, \dots, f_{kv})^t$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_v)^t$, $\mathbf{g}_k = (g_{k1}, \dots, g_{ku})^t$ のようにベクトル表記し、2、3層間の重み行列を W (u 行 v 列ベクトル) で表せば、

$$\mathbf{f}_k = \mathbf{W}^t \mathbf{g}_k - \boldsymbol{\theta} \quad (k=1, \dots, K) \quad (3-12)$$

を得る。このとき、教師パターン \mathbf{f}_k' との差ベクトルのノルムの2乗を標本全体について総和をとったもの $\sum_k |\mathbf{f}_k' - \mathbf{f}_k|^2$ を最小にするような重み行列 W 、集約表現 \mathbf{g}_k 、しきい値ベクトル $\boldsymbol{\theta}$ を求める問題は、 $\{\mathbf{f}_k\}$ ($k=1, \dots, K$) の平均値ベクトル、分散共分散行列の固有ベクトルを固有値の大きい順から u 個横に並べた行列を各々、 $\boldsymbol{\mu}$ 、 $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_u)$ として次式で与えられる。

$$\mathbf{f}_k = \mathbf{A} \mathbf{g}_k' + \boldsymbol{\mu} \quad (k=1, 2, \dots, K) \quad (3-13)$$

これは主成分分析(5)(統計的パターン認識における $K-L$ 展開)の問題に他ならない。従ってこのモデルの情報集約能力(成分数または素子数を一定にした場合の近似精度)は主成分分析を越えることはない。3層砂時計モデルに対するこの厳密な数学的証明は、中間層素子を線形特性とした場合に Bourlard et al. (1)により、同素子を線形またはSigmoid特性で考慮した、より一般の場合に船橋(2)により与えられている。しかしながら出力層素子をシグモイド特性とする場合には言及していない。以下ではその場合について述べる。

3. 3 出力層素子がシグモイド特性をもつ場合

Sigmoid関数(6)には表3-1に示すものが挙げられるが、ここでは関数 F として最もよく用いられる式(2-2)を例にとって考える。このとき、式(3-3)は、

$$f_{kq} = [1 + \exp \{-\sum p (g_{kp} W_{pq}) + \theta_q\}]^{-1} \quad (3-14)$$

のように書き直される。ここで逆関数 F^{-1} は、 $F^{-1}(x) = \ln(x/(1-x))$ ($0 < x < 1$) となる。そこで、例として \mathbf{g} を中間層の(1変数)成分パターン: $\mathbf{g} = g$ 、 \mathbf{f} を2変数出

カパターン: $f = (f_1, f_2)^t$ 、中間層・出力層間の重みを W_i 、出力層の各素子のしきい値を各々 θ_i ($i=1, 2$) とすれば、以下の式を得る。

$$f_i = 1 / [1 + \exp(-g W_i + \theta_i)] \quad (i=1, 2) \quad (3-15).$$

f_1, f_2 を、 g を媒介変数として関係付ければ、

$$\left(\frac{f_1 \exp(\theta_1)}{(1 - f_1)} \right)^{1/W_1} = \left(\frac{f_2 \exp(\theta_2)}{(1 - f_2)} \right)^{1/W_2} \quad (3-16).$$

f_1, f_2 を各々横軸、縦軸にとったとき式(3-15)を満たす点の軌跡を、 (W_2/W_1) および $C = \theta_2 - (W_2/W_1)\theta_1$ をパラメータとして描いたものを図3-2. b に示す。

これらの曲線群は線形回帰分析法(6) (主成分分析、重回帰分析等)における主成分軸、回帰直線に対応する、回帰曲線とも呼ぶべきものであり、2次元空間 ($0 < f_1 < 1, 0 < f_2 < 1$) 上に分布した多数のサンプルベクトル全体を近似するための軸(曲線軸)の候補を表す。

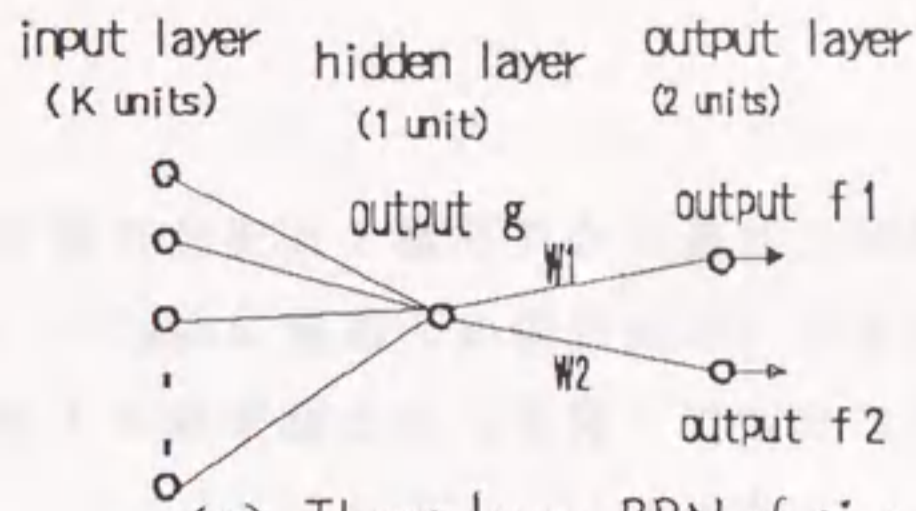
こうして、図3-2. a に示す3層ネットワークにより、 K 個の2次元サンプル集合 $\Omega = ((X_1, Y_1), \dots, (X_K, Y_K))$ を最適に近似する曲線 ($W_1, W_2, \theta_1, \theta_2$ で決まる) および、各サンプルからのユークリッド距離の2乗の総和

$$e = \sum_k |e_k|^2 = \sum_k ([X_k - \{1 + \exp(-g_k W_1 + \theta_1)\}]^{-1})^2 + [Y_k - \{1 + \exp(-g_k W_2 + \theta_2)\}]^{-1})^2 \quad (3-17)$$

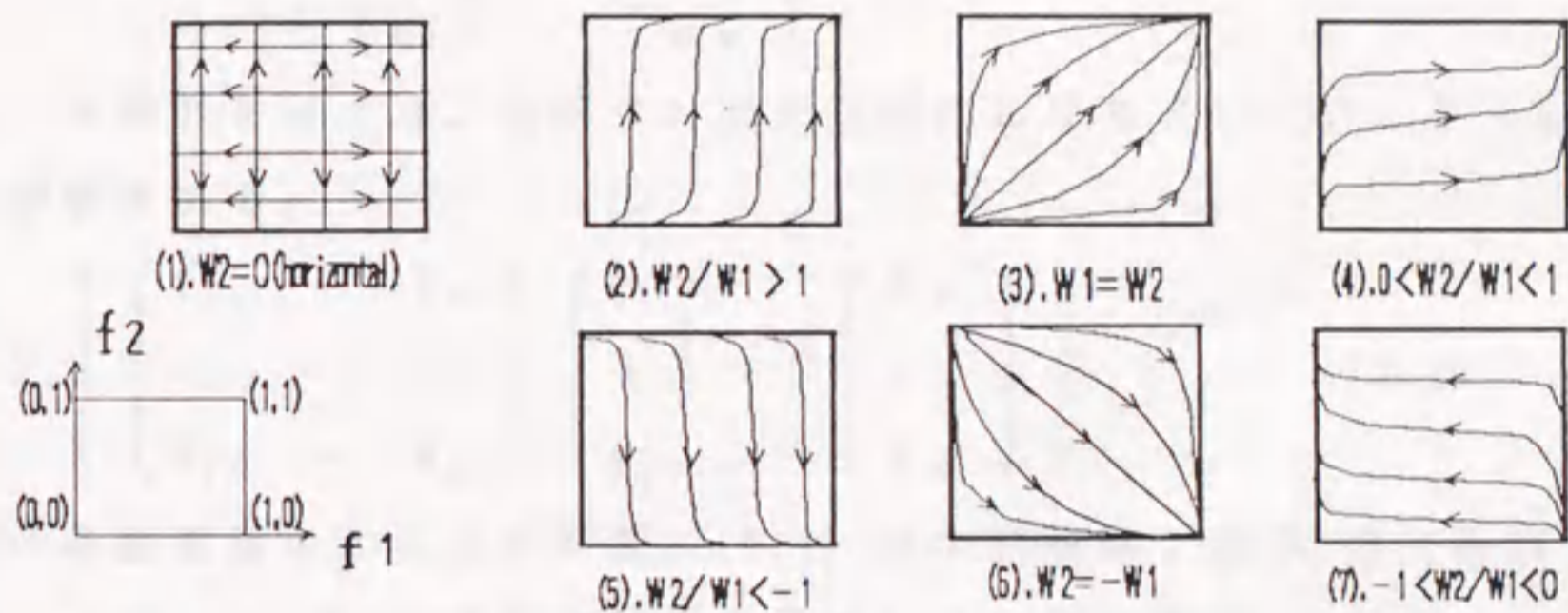
が最小となるその曲線上の値 g_k (その基底系におけるサンプル毎の係数) を決定し得る。

出力層素子が Sigmoid 特性をもつ場合での3層 P P N の幾何学的な解釈を与えれば、その学習は正規化された N 次元空間内 ($(X_1, \dots, X_N) | 0 < X_i < 1, i=1, \dots, N$) に分布するサンプル全体を、データ損失量の尺度としての2乗誤差を最小化する意味で最適に近似する曲線(あるいは曲面)を探索しながら、その推定曲線(曲面)上で各サンプルからのユークリッド距離の総和((3-17)式)が最小となるような点の位置を計算する過程である(図3-2. c)。

具体的には、最適曲線(表現素)を決定するために中間層・出力層間の重み(式(3-17)中の W_i, θ_i) を反復修正する。ここで、各サンプルを代表する曲線上の点を表す成分パターン g_k は、該当サンプルに関連付けられた入力層の第 k 素子と各中間層素子の結線1本のみ関数となる。



(a). Three layer PPN of size $(k-1-2)$.



(b). Point set of (f_1, f_2) satisfying Eq. (3-16)

(Seven types of curves categorized by the value of W_2/W_1 , and the arrows indicate the direction of increasing $(\theta_2 - \theta_1)(W_2/W_1)$.)

(c). Geometrical correspondence between the component decomposition by three layer PPN and K-L expansion

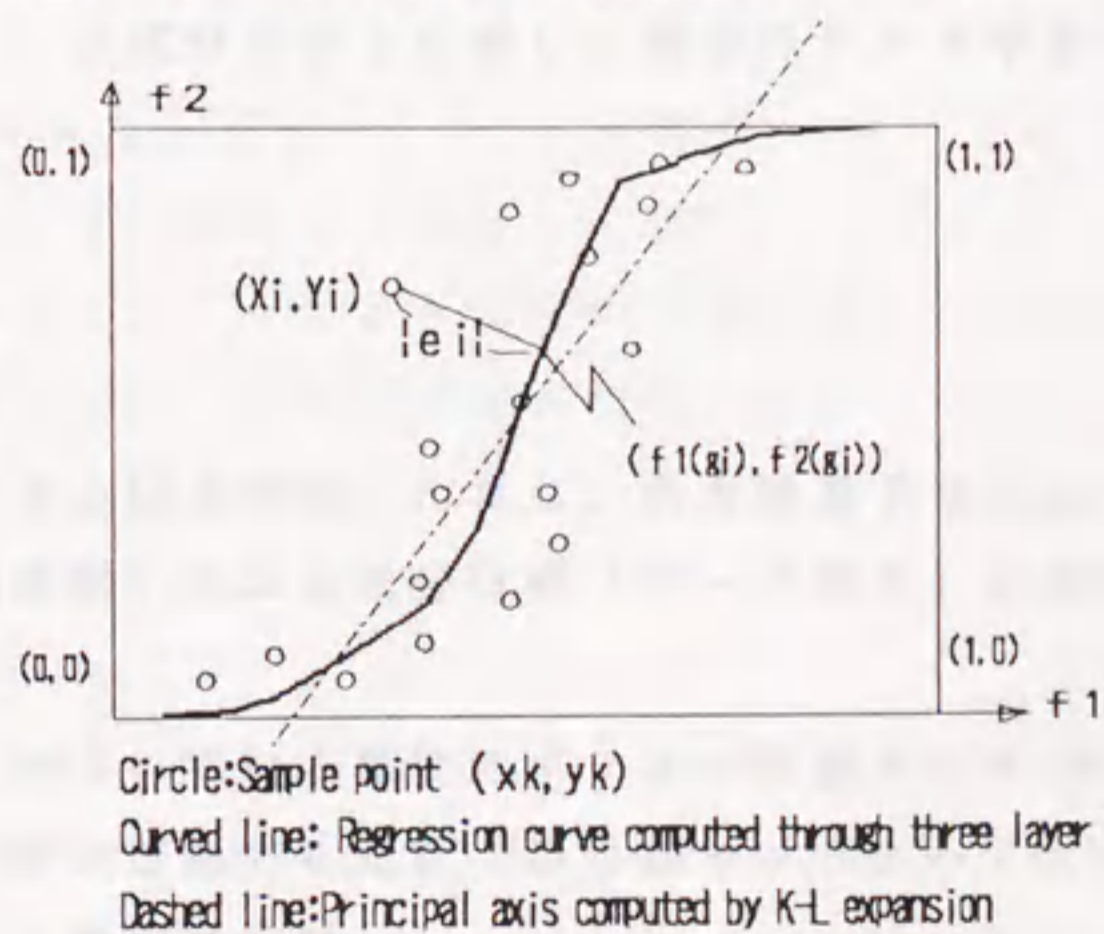


Fig 3.2 Geometrical interpretation of three layer PPN.

図 3 . 2 3 層 P P N の幾何学的解釈

従って、一般に主成分分析と3層PPNの差は、N次元(X_1, X_2, \dots, X_N)空間内のサンプルをd成分(g_1, \dots, g_d)に集約(d成分分解)する場合を例にとれば、主成分分析では、式(3-13)による超平面近似(展開してかけば)、

$$\begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} = (\mathbf{a}_1, \dots, \mathbf{a}_d) \begin{pmatrix} g_1 \\ \vdots \\ g_d \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix} \quad (3-18).$$

であるのに対し、3層PPNでは、有界なN次元空間における式(3-19)による超曲面近似であると解釈される。

$$\begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} = F \left[\begin{pmatrix} W_{11} & \dots & W_{d1} \\ \vdots & & \vdots \\ W_{1N} & \dots & W_{dN} \end{pmatrix} \begin{pmatrix} g_1 \\ \vdots \\ g_d \end{pmatrix} - \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_N \end{pmatrix} \right] \quad (3-19).$$

ただし、 $F(X)$ は X の各要素毎の F による写像、 $\{W_{ij}\}$ は中間層第 i 素子、出力層第 j 素子間の結合行列、 $(\theta_1, \dots, \theta_N)$ は出力層各素子のしきい値を表す。

この結果として、対象とするサンプル集合の特徴空間内での幾何学的性質(超平面のまわりの分布とした場合と、式(3-19)による超曲面のまわりの分布とした場合の近似精度の良否)に依存して、主成分分析と比較した場合のPPN学習の優位性が決まるといってよい。

3.4 シミュレーション

以下では、誤差逆伝搬法による上記2手法(ただし、出力層素子はSigmoid特性)と主成分分析(Karhunen-Loeve展開)による成分分解(データ集約)の効率を簡単な予備実験により評価する。

実験に用いたデータは、よく知られている植物アイリスの計測データ(5)で、ガク片の長さおよび幅、花卉の長さおよび幅の4変数(単位はセンチメートル)からなる150個体のデータである(これらは種類Iris setosa, Iris versicolor および Iris virginica 各々50個体ずつ3クラスからなる)。実験の目的は、これら150個の4変数データ(4次元特徴量空間におけるサンプル点150個とみなす)をいかに情報損失を抑えつつより少ない成分に集約するかを、(a). K-L展開(分

散共分散行列による主成分分析), (b). 3層 P P N, (c). 3層砂時計モデル, の3手法について比較することであり、結果の評価は成分数 ((b), (c) の場合は中間層のサイズ) P と、それらの成分軸に射影した点で各サンプルを代表させた場合の復元時の平均二乗誤差率 δ :

$$\delta(p) = \frac{\sum_k \sum_q (X_{kq} - X'_{kq}(p))^2}{\sum_k \sum_q (X_{kq})^2} \quad (3-20).$$

ただし X_{kq} は第 k サンプル第 q 変数の値、 $X'_{kq}(p)$ は集約成分数 p 個の場合の同変数の近似値。

により行った (直観的には、この値、すなわちデータ損失量、が小さい程、成分毎の情報への寄与が大きいといえる)。なお、原データの単位はセンチメートル (最小値 0.1 から最大値 7.9 までの値がある) であるが、特徴空間内での区間の取り方への依存度を調べるため、手法 (b), (c) の場合の入力 / 教師データへの与え方は、

- 1). 原データの区間 [0 , 1 0] を素子出力値 0 , 1 間の連続値に線形に写像
- 2). 同 [0 , 4 0] を 1 と同様に線形に写像

の 2 種類の方法で与えた。

それらの結果を表 3-2 に、また、各手法 ((b), (c) は区間 [0 , 4 0] によるもの) で 2 成分に射影した際の全データの散布図を付録図 A に示す。

なお、ここで行った実験はいずれも、標本の提示順序は無作為に行い、重みとしきい値の初期値は乱数で与えた。また、逆伝搬学習の学習パラメータ修正の過程 (いわゆる Simulated Annealing) は、誤差関数 e_n を減らす重み修正式 (2-4) (t 回目の重みを $W(t)$ 、その修正分を $\Delta W(t)$ と表す) に慣性項を加えたもの

$$\Delta W(t+1) = \beta \cdot \Delta W(t) - \alpha \cdot (\partial e_n / \partial W(t)) \quad (3-21).$$

における、修正係数 α と慣性係数 β を各々、順に、(0.2, 0.9), (0.15, 0.85), (0.1, 0.8), (0.06, 0.5) で 5000 回ずつ (計 20000 回) の反復学習という要領で行った。

これらの結果の考察として、

- (1). 全ての場合について P P N のデータ損失は砂時計モデルのそれより多少少ない。またその傾向は中間層サイズ (削減成分数) が大であるほど強く現れる。
- (2). K-L 展開と P P N の比較では、ともに同程度のデータ損失となり一概に

どちらが優位であるとは言えない（1割程度PPNが優位となる場合もある）。

(3). PPN、砂時計モデルの場合、原データの値と素子の出力値の対応付けの尺度を変えること（回帰曲線候補群を囲む枠の尺度の、特徴空間に対する相対的变化）により情報損失に差異が生ずる。

(4). 砂時計モデルの場合、中間層サイズを入力層サイズと同一としたまま学習してもかなり誤差が大きいまま収束することがある（極小点への収束を意味する）。

などが挙げられ、前節まで述べた議論の妥当性が示されたと言えよう。また砂時計モデルに対しPPNの不利な点として、一般にサンプル数が大であるほど入力層サイズ、重みの数がともに増大することから学習に要する時間的コストの問題が考えられるが、1サンプルの反復学習1回に要する両回路での入力層・中間層間についての計算量（中間層・出力層間の計算量は同一）は、3.1節からも知られるように、砂時計モデルが（サンプル次元数×中間層サイズ）本の結線修正となるのに対しPPNではパルスを与える入力層素子からの結線のみを学習対象とするため、実際は中間層サイズの本数の修正で十分となる。更に、収束までの反復回数に関してもPPNのそれが砂時計モデルより少なく済むことは二三の実験例で確認されており(4)、実際にシミュレーション実験を行う際のPPNの有用性を裏付けている。

3層PPNと3層砂時計モデルの学習効率に関する、より詳細な実験的検証は文献(7)に与えられているが、PPNが優位であることの原因として中間層・出力層間の重み（基底ベクトルに相当）の最適性に関してPPNが有利であることが知られている。また、Baldi and Hornik (3)は全素子が線形の場合、極小解は誤差空間で、鞍部点(Saddle pointすなわち不安定な平衡点)と呼ばれる点に対応し、その実体は分散共分散行列の固有値の小さい固有ベクトルにより張られる空間を意味するものであることを示している。

なお、前章の議論から、4層以上の層のPPNでは、第2層を集約層とし、第3層以降のサイズを適当に大きくとることにより、3層PPNおよびK-L展開等と比較してより効率のよい情報集約が可能となることは容易に予想される。

表 3-1. シグモイド関数の例 (ln は自然対数を表す)

関数形 $F(x)$ と値域	逆関数 $F^{-1}(x)$
$(1 + \exp(-x))^{-1}$ $[0, 1]$	$\ln(x / (1 - x))$
$(1 - \exp(-x)) / (1 + \exp(-x))$ $[-1, 1]$	$\ln((1 + x) / (1 - x))$
$\arctan(x)$ $[-\pi/2, \pi/2]$	$\tan(x)$

表 3-2. アイリスデータの集約実験 (枠内は平均 2 乗誤差率 δ : 単位は%で表示)

項目 \ 成分数	1 成分	2 成分	3 成分	4 成分
(a). K-L 展開	0.553	0.160	0.0378	0.0
(b). PPN 区間(0,40)	0.599	0.147	0.0349	0.0003
(c). 砂時計モデル 区間(0,40)	0.604	0.167	0.0431	0.0130
(b). PPN 区間(0,10)	0.607	0.158	0.0377	0.0002
(c). 砂時計モデル 区間(0,10)	0.613	0.168	0.0440	0.0067

3. 5 まとめ

本章では、多層ニューラルネットにおける誤差逆伝搬法による学習機構を多変量データの情報集約（成分分解）の点から考察し、その数理的性質を明かにした。すなわち、2層の復号化回路をもつ3層砂時計モデルと3層PPNの数理的な対比を行い、次に出力層素子の特性が線形の場合とシグモイドの場合とに分けて幾何学的性質を考察し、線形の場合には高々主成分分析と同程度の集約能力のみしか期待できないこと、シグモイドの場合は、標本分布の幾何学的性質に応じて主成分分析の能力を上回ることがあること等を理論的に示した。そして最後に多変量解析の分野でよく引用されるアイリスデータを用いたこれら3手法によるシミュレーション実験の結果を示し、これらの考察を裏付けた。

逆伝搬法を用いたパターン回帰の学習については、種々のデータに対するデータ圧縮等の実用化が検討されているが、それらは単に性能面での評価ならびに実験による回路網の形状の検討の範囲に留まるもの(8)-(14)か、もしくは最も簡単な構成の場合についてのみ考察したもの(1)-(3)のいずれかであった。これに対し、本章ではその理論的根拠を、新たに導入したPPNによるパターンの成分分解の概念を用いて論じた。

なお、本章の内容の延長として、集約効率改善のための、原データの値域と素子の出力値の対応までを含めた最適性の考察とそれを実現する学習規則の定式化、および、4層PPNへの拡張の問題等がある。

次章では、その中でも特に重要と思われる4層PPNについて、本章と同様の幾何学的解釈を与える。

参考文献

- (1). Bourlard, H. and Kamp, Y.: "Auto-Association by Multilayer Perceptrons and Singular Value Decomposition", Biol. Cybern. 59, pp. 291-294 (1988).
- (2). 船橋賢一: "三層ニューラルネットワークによる恒等写像の近似的表現について"

- の理論的考察”, 信学技報, MBE88-174, pp.109-114(1989).
- (3). Baldi, P. and Hornik, K. : "Neural Networks and Principal Component Analysis: Learning from Examples Without Local Minima", Neural Networks, 2, pp. 53-58(1989).
- (4). 米倉達広、横井茂樹、鳥脇純一郎 : "ニューラル・ネットを用いた濃淡画像の表現素展開による画像表現", 信学会 画像理解の高度化と高速化シンポジウム論文集, pp.127-132(Apr. 1989).
- (5). Anderson, T.W. : "An Introduction to Multivariate Statistical Analysis" 2nd Edition, John Wiley & Sons, New York(1984).
- (6). Rumelhart, D.E. and McClelland, J.L. : "Parallel Distributed Processing", Vol.1, MIT Press(1986)
- (7). 星野俊仁, 米倉達広, 松本哲也, 鳥脇純一郎 : "3層恒等写像ニューラルネットワークによる主成分分析の検討", 信学技報, PRU90-54, pp.7-14(1990).
- (8). Elman, J.L. and Zipser, D. : "Learning the Hidden Structure of Speech", Tech.Rep. 8701, Univ. of California, San Diego, Institute for Cognitive Science (1987).
- (9). Cottrell, G.W., Munro, P. and Zipser, D. : "Image Compression by Back Propagation: An Example of Extensional Programming", Tech.Rep. 8702, Univ. of California, San Diego, Institute for Cognitive Science(1987).
- (10). 曾根原登、川人光男、三宅誠、中根一成 : "ニューラルネットによる画像データ圧縮(Neuro-CODEC)の検討", 信学技報, IE88-62, pp.57-64(1988).
- (11). 長坂保典、岩田彰、鈴木宣夫 : "ニューラルネットによる長時間心電図の特徴抽出と情報圧縮", 信学技報, MBE88-91, pp.73-80(1988).
- (12). 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 - テーブルルックアップ法と違うのか? -", 信学技報, NC89-15, pp.33-40(1989).
- (13). 臼井支朗, 中内茂樹, 中野正恵、 : "多層ニューラルネット恒等写像モデルによるマンセル色空間の特徴解析", 信学技報, NC89-40, pp.45-50(1989).
- (14). 片山泰男、大山公一 : "自己組織逆伝搬ニューラルネットの諸特性", 平成元年信学春大, 7, pp.309-310(1989).

第4章

4層PPNと区分線形因子分析

前章では、2章の復号化回路をもつ3層砂時計モデルと3層PPNの幾何学的性質について述べ、それらとK-L展開をシミュレーション実験により比較した。そこで、次に考察すべき問題点として、3層の復号化回路をもつ5層砂時計モデルならびに4層PPNについての同様な考察が挙げられる。5層砂時計モデルについての考察ならびに検討は入江ら(1)、臼井ら(2)、片山ら(3)、宮崎ら(4)により試みられているが、その幾何学的解釈を得るには至っていない。

本章ではこの問題、すなわち両モデルでの素子数とデータ回帰能力(情報集約能力)の幾何学的関係について議論する。

4層ネットワークの数理的性質は、パターン分類、情報集約のいずれの問題に用いる場合も、3層の場合と比較してはるかに複雑であり、その内部表現の解析はかなり困難である。このことはPPNについてもいえ、3層PPNについて前節で行った代数的記述による線形モデルとの比較をそのまま4層に拡張することはできない。

しかしながら4層PPNはその復号化回路部分が3層ネットワークどなっている(図4-1)ことから、その集約能力は本質的には3層ニューラルネットの写像能力の問題に帰着できる。従って集約層の後段層、すなわち第3層のサイズを適当に大きくとることにより、理論的に3層PPNあるいは線形回帰モデルをはるかに上回る情報集約能力を持つと予想される(5)(6)。この意味で、4層PPNに何等かの幾何学的解釈を与えることは極めて重要である。しかしながら、いずれの報告も、中間層の素子数 n を有限として、3層ニューラルネットの写像能力を n の関数として定式化するまでには至っていない。従って、この問題を解明し、4層PPNの第3層サイズと集約能力の関係を定性的、定量的に評価することが急務となる。

そこで、以下では、各素子の入出力特性を簡単な関数で近似することにより、4

層 P P N に対するこの解析を試みる。

4. 1 シグモイド関数の区分線形関数近似

素子の入出力特性を F とする多層ネットワークにおける入力パターンから出力パターンへの写像は、隣合う 2 層間での前層の各素子の出力値 $F(X)$ と、層間での重み行列 W の積が後層の入力として与えられるという計算過程が次々と後の層へと伝搬する、いわゆる写像 F の反復計算の形式をとる。この F を Sigmoid 関数とするネットワークモデルの解析は、一般に式 (2-2) のような非線形関数の再帰形式となり、代数的な記述による取扱いが難しい (7)。

そこで、以下では、同モデルの内部状態の解析に区分線形近似の考え方を導入し、素子の入出力特性である非線形関数を区分的に線形な関数で置き換えることにより、ネットワーク全体の写像能力を考察する。

いま、式 (2-2) が、点 $(0, 0.5)$ に関して対称であることから、 $X=0$ の回りで Taylor 展開すれば、

$$F(X) = 1/2 + X/4 - X^3/48 + \dots \quad (4-1).$$

ここで、3 次以上の項を無視し、次式を得る。

$$F(X) \approx 1/2 + X/4 \quad (4-2).$$

更に、式 (2-2) が、 $0 < F(X) < 1$ の形の飽和型関数でかつ単調増加であることを考慮すれば、これを式 (4-2) および $(F(X)=0)$, $(F(X)=1)$ の 3 つの直線を用いた区分的線形関数で良く近似できる (図 4-2)。

$$F_{pl}(X) = \begin{cases} 1 & \dots \quad (X > 2) \\ 1/2 + X/4 & \dots \quad (-2 \leq X \leq 2) \\ 0 & \dots \quad (X < -2) \end{cases} \quad (4-3).$$

以下、説明の都合上、 F_{pl} の定義域を各々、第 1 区間 $(X < -2)$ 、第 2 区間 $(-2 \leq X \leq 2)$ 、第 3 区間 $(X > 2)$ と呼ぶ。

この関数近似により、隣合う 2 つの層におけるパターンの写像 (前層の出力パターン y から後層の出力パターン u への写像 T) は、次式で定式化される区分的線形写像とみなすことができる。

$$\begin{pmatrix} u_{11} \cdots u_{1K} \\ \vdots \quad \cdots \quad \vdots \\ u_{M1} \cdots u_{MK} \end{pmatrix} = F_{pl} \left(\begin{pmatrix} W_{11} \cdots W_{N1} \\ \vdots \quad \cdots \quad \vdots \\ W_{1M} \cdots W_{NM} \end{pmatrix} \begin{pmatrix} X_{11} \cdots X_{1K} \\ \vdots \quad \quad \quad \vdots \\ X_{N1} \cdots X_{NK} \end{pmatrix} - \begin{pmatrix} \theta_1 \cdots \theta_1 \\ \vdots \quad \quad \quad \vdots \\ \theta_M \cdots \theta_M \end{pmatrix} \right) \quad (4-4).$$

ただし、 $F_{pl}(A)$ は、行列（またはベクトル） A の各要素を式(4-3)の関数 F_{pl} で写像したものを要素とする行列（またはベクトル）を表す。

多層ネットワークの内部表現の解析のツールとしては、しきい値関数近似の考え方をを用いる試みが有用であり、パターン分類問題に関して種々の成果をあげている(7)。以下では、しきい値関数に比べ、Sigmoid関数の近似誤差がより少ない上記手法により一般の3層ネットワークのパターン変換の性質を幾何学的に示し、そのうえで4層PPNによる情報集約問題を考察する。

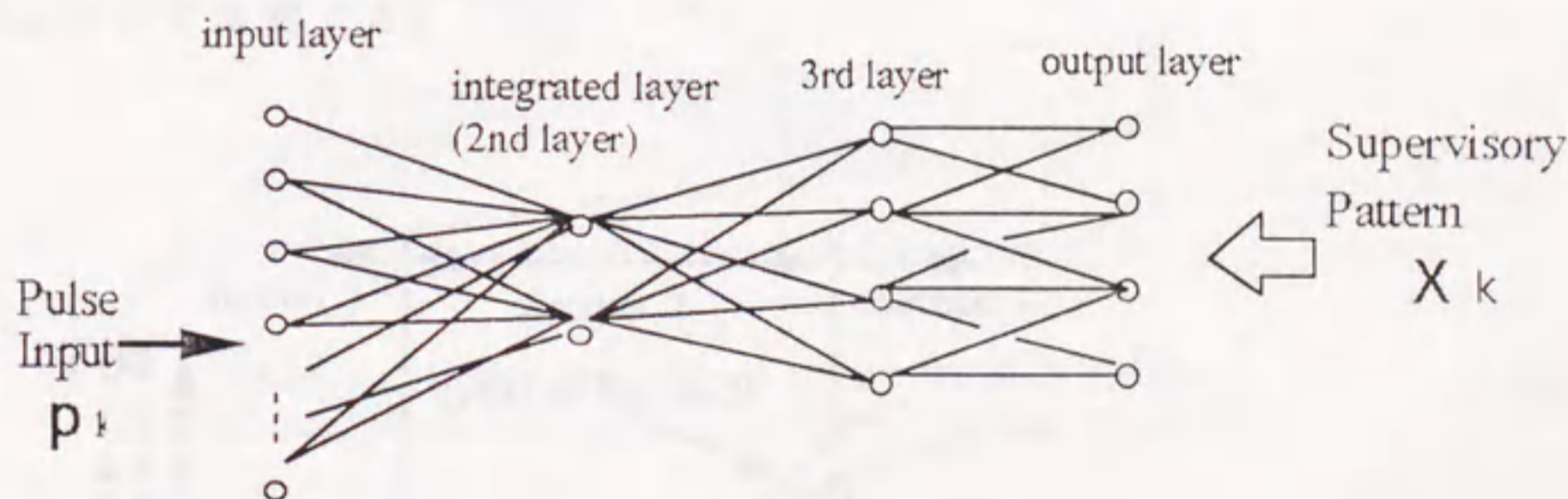


Fig.4.1 4layer pulse-in pattern-out network

図4.1 4層パルス入力パターン出力型ネットワーク

4. 2 4層 P P N の集約表現の幾何学的解釈

いま、簡単のために4層 P P N で、第3層素子を除く全素子が線形特性、第3層素子が F_{pl} による区分線形特性をもつものを考え、集約層すなわち第2層でのパターンがいかに出力層のパターンに写像されるかを調べる。集約層でのパターンを y (N次元ベクトル)、第3層でのパターンを u (H次元ベクトル)、出力層でのパターンを x (M次元ベクトル、ただし $M > N$) とすれば、これらの間に、

$$u = F_{pl}(P \cdot y - \theta) \tag{4-5}$$

$$x = Q \cdot u - \phi \tag{4-6}$$

が成り立つ。ただし P 、 Q は各々、第2層(集約層) - 第3層間、および第3層 - 出力層間の、重み行列 ($H \times N$ 行列、 $M \times H$ 行列) を表し、 θ 、 ϕ は各々、第3層、第4層の各素子のしきい値を並べた列ベクトル (H-ベクトル、M-ベクトル) を表す。ここで、この第2層から第4層までを一つの3層ネットワークと考え、その写像能力について考察する。

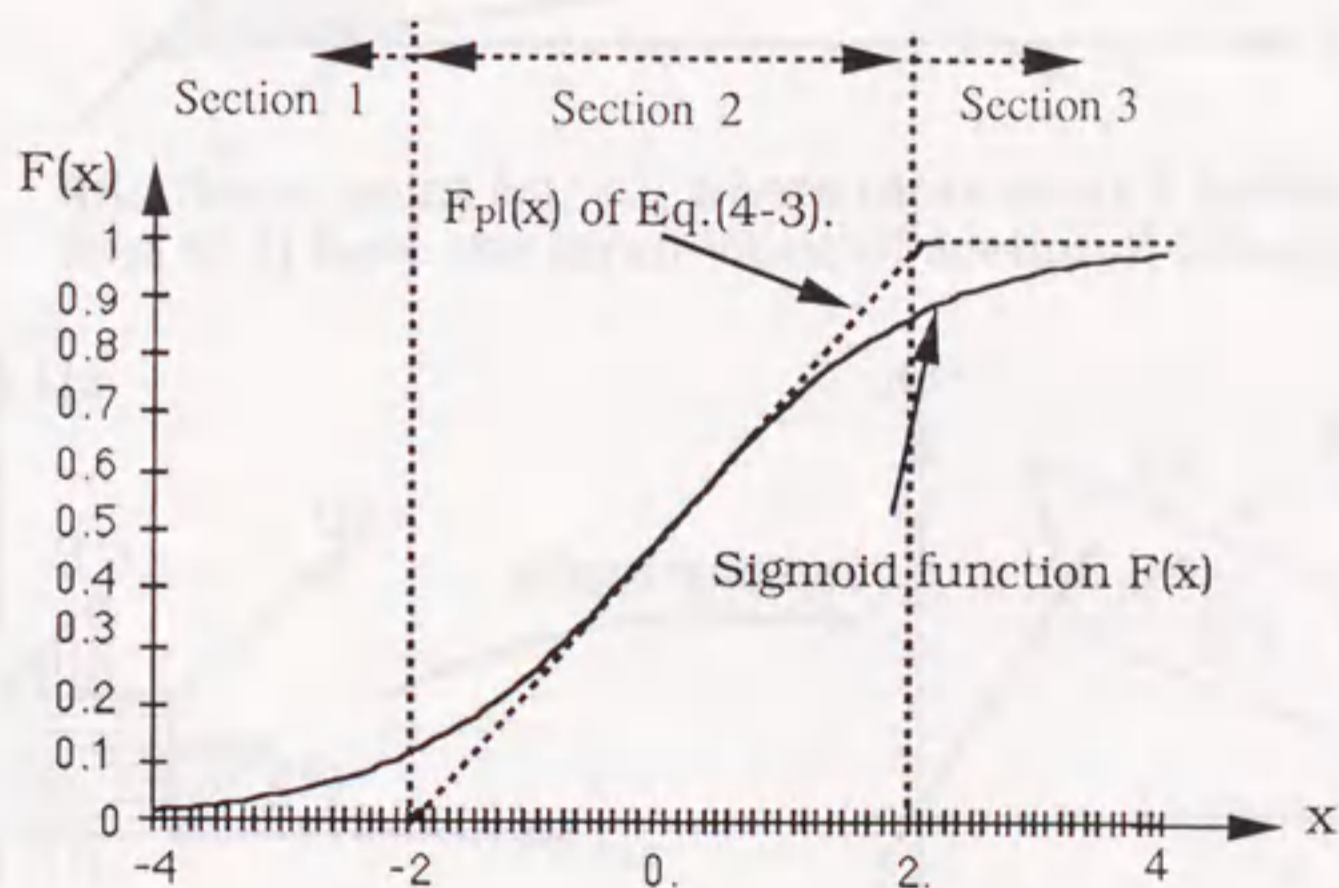
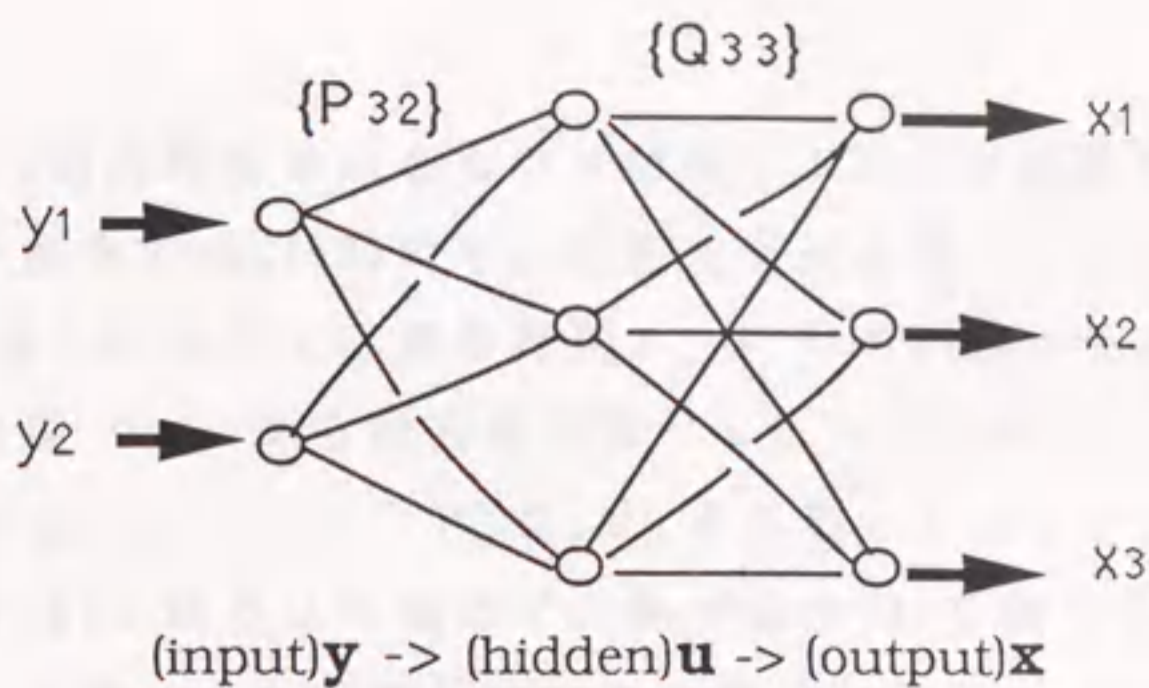
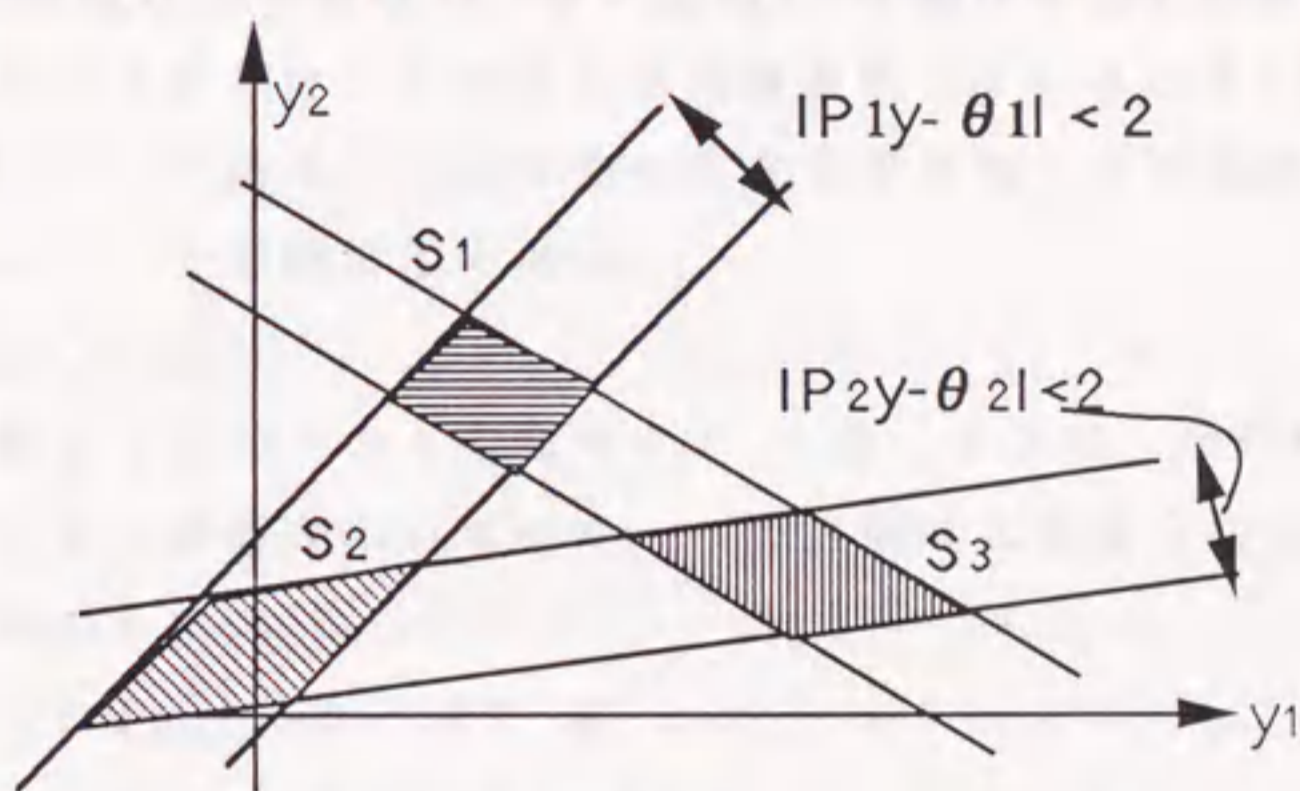


Fig.4.2 Piecewise linear function $F_{pl}(x)$ as an approximation of Sigmoid function.

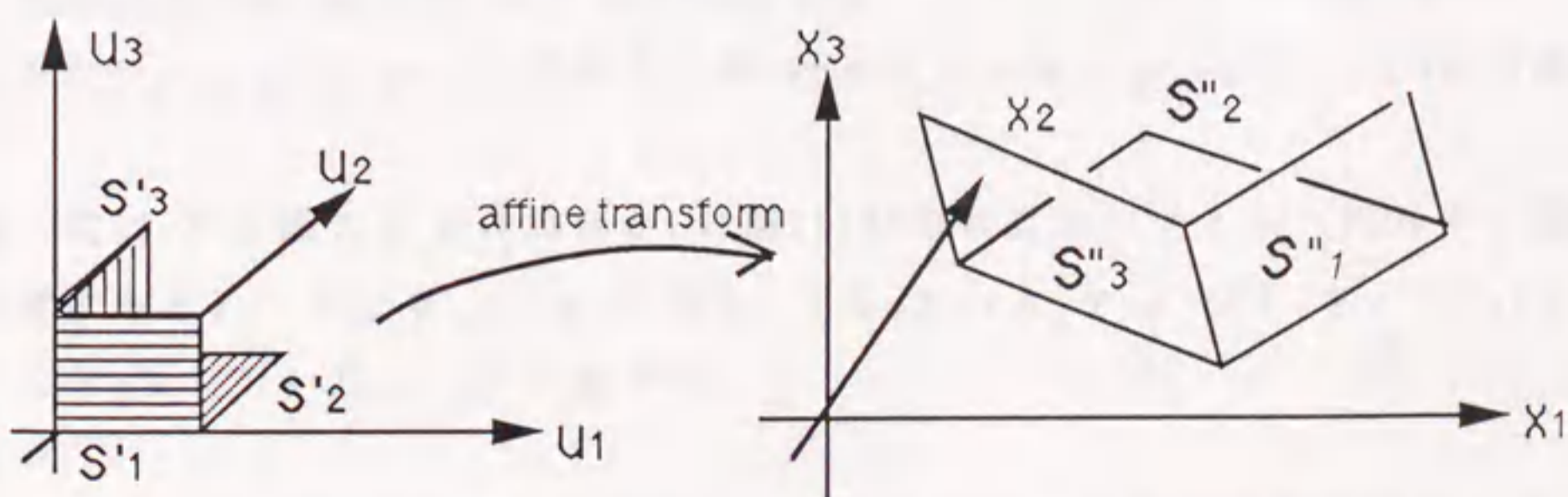
図4.2 Sigmoid関数を近似するための区分線形関数 $F_{pl}(x)$



(a). Three layer sub-network (2-3-3) of the integrated layer through the output layer of the four layer PPN.



(b). Set of point (y_1, y_2) where more than 2 hidden units (out of 3) have the input value of section 2 (shaded area).



(c). The mapping from U space of the hidden layer into X space of the output layer (affine transform).

Fig.4.3 Explanation of PL mapping by the network shown in (a).

図4.3 4層 PPN による 区分線形写像 T2 の幾何学的解釈

[補題1] (入出力特性 F_{pl} をもつ3層ネットワークの幾何学的性質) (図4-3)

中間層素子が全て式(4-3)の F_{pl} 特性をもつ3層ネットワーク (素子数 $N-H-M$, ただし $N \leq M$) において、重み行列 P, Q , しきい値ベクトル θ, ϕ が与えられたとき、入力層表現 $\{y\}$ から出力層表現 $\{x\}$ への写像 ε は、これらで決まる区分的線形写像となる。このとき、 P, Q, θ, ϕ に応じて以下が成り立つ。

- (a). 下記条件(1)が成り立つ場合で、かつ条件(2)を満たすような N 次元ベクトル y の集合 $\{y\}$ を ε により写像してできる出力層表現 $\{x\}$ は、 R^n で N 次元半平面を張る。
- (b). 条件(1)が成り立つ場合で、かつ条件(3)を満たすような N 次元ベクトル y の集合 $\{y\}$ を ε により写像してできる出力層表現 $\{x\}$ は、 R^n で N 次元凸閉包を張る。
- (c). (a)および(b)で張ることの可能な N 次元半平面と N 次元凸閉包の総数は、中間層サイズ H について単調増加である。

条件(1). $\text{rank}(Q) \geq N$

条件(2). 中間層各素子への入力値の中で、 f 個 ($f \geq N$) が同時に第2区間に入り、その他が第1または第3のいずれか一つの区間に入るような N 次元ベクトル y が存在する。すなわち、

$$\begin{aligned} \{y \mid -2 \leq (p_i \cdot y - \theta_i) \leq 2 \quad (\forall i = a_1, \dots, a_f) \\ \text{かつ,} \quad (p_i \cdot y - \theta_i) < -2 \quad (\forall i = a_{f+1}, \dots, a_{f+e-1}) \\ \text{かつ,} \quad (p_i \cdot y - \theta_i) > 2 \quad (\forall i = a_{f+e+2}, \dots, a_H), \\ \text{但し, } f \geq N, a_s = 1, \dots, H; a_s \neq a_r \quad (\forall s \neq r)\} \end{aligned} \quad (4-7).$$

が空でない。ここで、 p_i^t は行列 P の第 i 行ベクトルを、 $p_i \cdot y$ は p_i と y の内積を表す。

条件(3). 式(4-7)を満たす y の集合が、 L 個 ($L > N$) の N 次元ベクトル (内 $N+1$ 個は一般の位置にある) : y_0, y_1, \dots, y_L に対し、 $\{\lambda_0 y_0 + \lambda_1 y_1 + \dots + \lambda_L y_L \mid \lambda_0 + \lambda_1 + \dots + \lambda_L = 1, \lambda_i \geq 0, i = 1, \dots, L\}$ と書ける。

[証明] 入力(出力)ベクトル $y(x)$ の集合は $N(M)$ 個の実数要素からなる、 $R^N(R^M)$ の点集合である。ここで、(4-5)式を(4-6)式に代入し、

$$x = Q \cdot F_{pl}(P \cdot y - \theta) - \phi \quad (4-8).$$

Fplは区分的線形関数であるから、写像 $\mathcal{E} : \{y\} \rightarrow \{x\}$ は明かに区分的線形写像である。従ってこの式は、AをM行N列の行列、cをM次元ベクトルとし、yのある範囲（例えば式(4-7)）について

$$x = A \cdot y + c \tag{4-9}$$

のように書くことができる。ここで、この範囲内のyの変化 $(y_1, \dots, y_N) \rightarrow (y_1 + \Delta y_1, \dots, y_N + \Delta y_N)$ によって写される出力層表現 R^M の像: $\mathcal{E}(y)$ の次元は、明らかに $\text{rank}(A)$ に等しく、高々Nである。いまAを求めるために式(4-9)を展開し、次式を得る。

$$\begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix} = \begin{pmatrix} Q_{11} \cdots Q_{1H} \\ \vdots \\ Q_{M1} \cdots Q_{MH} \end{pmatrix} \begin{pmatrix} F_{pl}(p_1 \cdot y - \theta_1) \\ \vdots \\ F_{pl}(p_H \cdot y - \theta_H) \end{pmatrix} - \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_M \end{pmatrix} \tag{4-10}$$

(a). 条件(1) ($\text{rank}(Q) \geq N$) が成り立つとする。一般性を損なわずに、列ベクトル $(P \cdot y - \theta)$ の成分中第1成分から第f成分までのf個のみが第2区間、第f成分からe (≥ 0)個は第1区間、残りは全て第3区間をとるようなyの範囲が存在するとすれば、式(4-3)のFplの項は次のように変形される。

$$\begin{pmatrix} F_{pl}(p_1 \cdot y - \theta_1) \\ F_{pl}(p_2 \cdot y - \theta_2) \\ \vdots \\ \vdots \\ F_{pl}(p_H \cdot y - \theta_H) \end{pmatrix} = \begin{pmatrix} 1/2 + (p_1 \cdot y - \theta_1) / 4 \\ \vdots \\ 1/2 + (p_f \cdot y - \theta_f) / 4 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \tag{4-11}$$

但し、 $\{y \mid -2 \leq (p_i \cdot y - \theta_i) \leq 2 \quad (\forall i=1, \dots, f),$

かつ、 $(p_i \cdot y - \theta_i) < -2 \quad (\forall i=f+1, \dots, f+e+1),$

かつ、 $(p_i \cdot y - \theta_i) > 2 \quad (\forall i=f+e+2, \dots, H)\} \tag{4-12}$

式(4-11)を式(4-10)に代入し式(4-9)の形に表せば、 $\text{rank}(A) = \min(N, f)$ が成り立つ。従って条件(2)すなわち、N個以上の中間層素子が同時に第2区間の入力値

を受け、その他が第1もしくは第3区間のいずれか片方のみを受けるとする y の範囲において、出力層に出力される x の集合は、 R^M で N 次元半平面をなす。

(b). 式(4-12)は、 $\{p_i\}, \{\theta_i\}$ を適当に選べば、条件(3)を満たすようにできる。このような $\{y\}$ はその定義から、 R^M で N 次元凸包(8)をなす。また、この $\{y\}$ をアフィン変換(階数 N の線形変換 A + 平行移動 c)により R^M に写した像も明かに N 次元凸包をなす。

(c). (a)および(b)で張ることの可能な N 次元半平面(および凸包)の総数は、式(4-7)を満たす集合 $\{y\}$ を与える係数集合の3組 $(\{a_1, \dots, a_f\}, \{a_{f+1}, \dots, a_{f+e-1}\}, \{a_{f+e+2}, \dots, a_H\})$ (但し $f=N, N+1, \dots, H$) の組合わせの数の最大に等しい。一方、中間層サイズ H を1増やしたとき、この数が必ず増加するように (p_{H+1}, θ_{H+1}) を設定する方法は(式(4-7)中 $H \rightarrow H+1$ かつ N が固定であることから)無数にある。

<証明終>

上記の性質は、3層ネットワークの、入力空間の点集合から、それより大きな次元の出力空間上の点集合への PL 写像の幾何学的性質を示唆する。すなわち、3層ネットワークで、入力層サイズ以上のサイズをもつ中間層を通した場合、入力空間 R^N は出力空間 R^M で1個以上の N 次元凸包(または超半平面)に写像される。加えて、(中間層サイズ - 入力層サイズ)が大きくなるほど、出力空間上でより多くの面からなる多面体に写像する能力が与えられる。

この性質を利用すれば、入力空間上に適当な点集合を与え、中間層サイズを増やすほど、出力空間 R^M の複雑な形状の標本分布(例えば多数の N 次元凸包からなる多様体、多面体(8)上の点で代表されるような標本分布)を、より少ない誤差で近似し得ることは容易に知られる。

次に、上記の結果をもとに、 Fpl 特性をもつ4層 PPN の情報集約機能について議論する。

いま、第3層の素子が全て特性 Fpl 、それ以外の素子が線形特性をもつ4層 PPN を考える。このとき、以下の性質が成り立つ。

[定理1] (Fpl 特性をもつ4層 PPN の幾何学的性質)

各層サイズが $(K-N-H-M)$ の4層 PPN において、出力層に提示された、 R^M (但し、 $M > N$) 上の標本分布 $\{x_k | k=1, \dots, K\}$ は、

(a). 各層間の重み行列を適当に決めれば、集約層で表現される N 次元ベクトル y を

変数とする区分的線形写像 f により、(第3層サイズを大きくとるほど)いくらでも多くの区分(凸包, 超半平面)に分けて近似できる。

(b). このように近似する際、入力層集約層間の重みを適当に設定すれば、 k 番目のパルス入力時に、 R^M の像 $f(y)$ 上で第 k 標本 x_k へのユークリッド距離の最も小さい点 $f(y_k)$ を出力させることができる。

[証明]

(a). 4層PPNの、集約層(第2層)から第4層までを取りだした3層からなるネットワークを考えれば、これは補題1の性質を全て満たす。従って補題1-(c)より明かに、第3層サイズを大きくするほど、多くの N 次元凸包(超半平面)に分けて R^M (但し、 $M > N$)上の標本分布 $\{x_k | k=1, \dots, K\}$ を近似し得る。

(b). 入力層第 k 番目の素子のみ値1、その他に値0の信号(第 k パルス信号)を与えた際に、 $f(y)$ 上で x_k に最も近い点 $f(y_k)$ を出力させることは、入力層第 k 素子と集約層の間の重みを、期待すべきパターン y_k の各成分に一致させることで直接行うことができる。

<証明終>

上述の前半の議論では、前章で導入した区分線形関数を用いて、非線形ネットワークの内部状態の幾何学的かつ定性的な解釈を試みた。これにより、近似の枠内ではあるが、既に示唆されている3層ネットワークの写像能力(5)(6)と、ネットワークの内部に得られる重み行列との幾何学的関連を具体的に与えた。

また後半では、4層PPNで重み行列を適当に設定することにより、標本分布の集約的表現(次元を削減した空間への射影)を集約層で得ることが可能であることを示した。

これらの考察はいずれも、素子の入出力特性であるSigmoid関数を3個の区間から成る区分的線形関数 F_{pl} で置き換え、ネットワークによる非線形写像を区分的線形写像で近似して行った。従って、Sigmoidを用いた実際の出力と、これを近似する F_{pl} を用いたそれとの間には計算誤差が生じるため、ここでの考察が実際のネットワークの性質を完全に記述するものではない。一方、実際に F_{pl} 特性をもつ素子からなるネットワークでは(F_{pl} が微分不可能のため)誤差逆伝搬(7)による学習を行うことは不可能である。

しかしながら、関数 F_{pl} を用いることは、Sigmoid特性のネットワークで逆伝搬学習を通して得られた重みによる写像の幾何学的性質を比較的容易に解析できるうえ

に、しきい値関数等を用いるよりも計算誤差の点ではるかに有利である。更に、4層 P P N で効率の良い学習結果が得られれば（極小解の回避）集約層での表現が、区分的線形因子分析の解となり得ることは容易に知られる。次節ではこれを具体的に示し、更にその応用としてのクラスタリングの一手法を提案する。

4. 3 4層 P P N による区分的線形因子分析とクラスタリングの概念

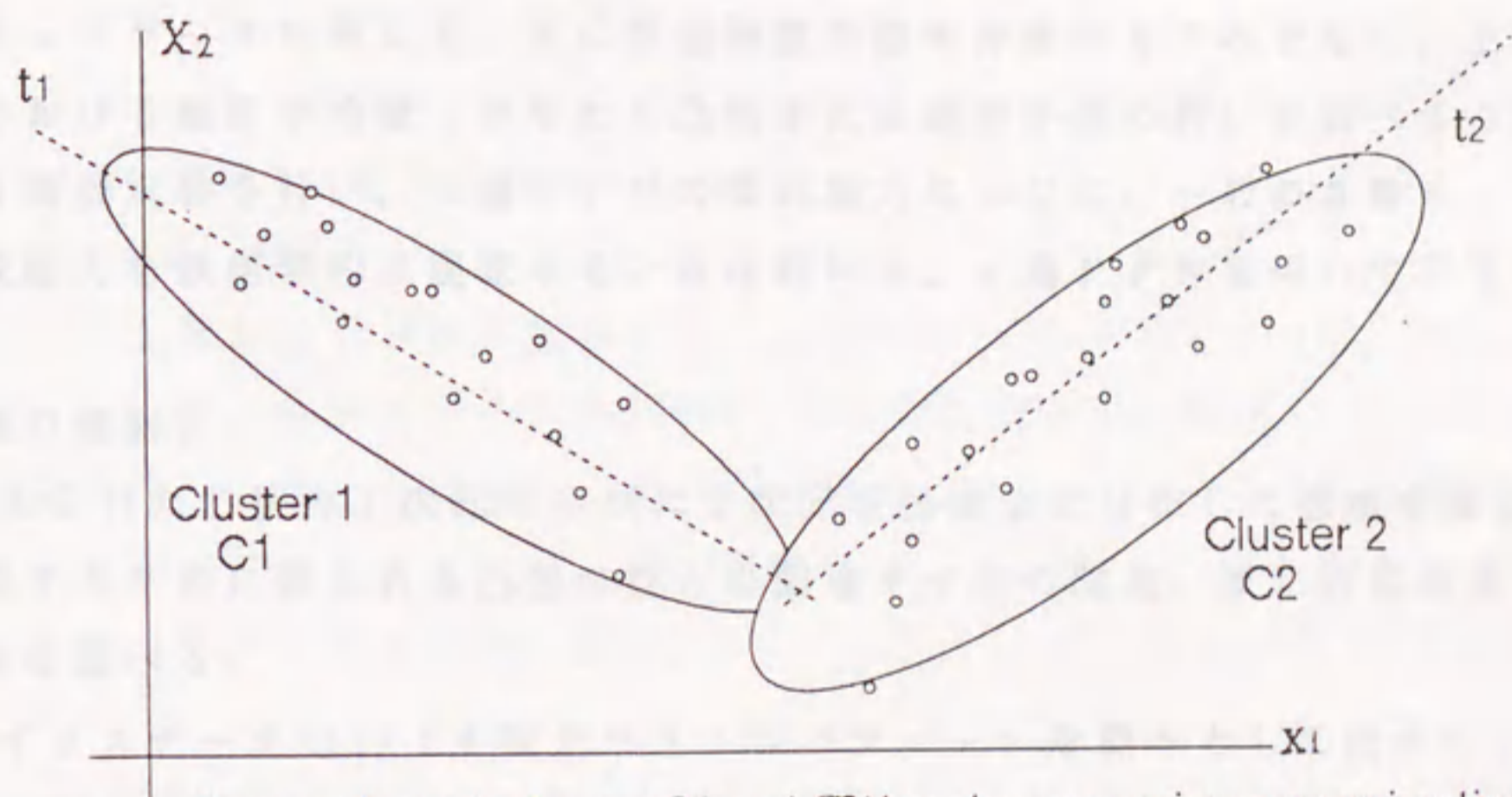
前節で導入した区分的線形関数 F_{pl} を用いた、集約表現 y から出力表現 x への写像 $f_{pl}: y \rightarrow x$ により式(2-16)の因子分析モデルを記述すれば、

$$x = f_{pl}(y) = \begin{cases} f_1 y + c_1 & \dots (y \in C_1) \\ f_2 y + c_2 & \dots (y \in C_2) \\ \vdots & \vdots \\ f_z y + c_z & \dots (y \in C_z) \end{cases} \quad (4-13)$$

のようになる。但し、 f_1, \dots, f_z は各々異なる線形変換（ M 行 N 列行列）、 c_1, \dots, c_z は M 次元ベクトル、 C_1, \dots, C_z は N 次元集約空間における、4. 2 節で定義された半平面もしくは凸閉包を表す。従って、各 C_i に属する点 y に対しては、対応する（ f_i, c_i ）により線形変換と平行移動で x が回帰されることになる。

また、この考え方で更に、同一の C_i に属する集約空間上の点 y に対しては同一の回帰式が適用されることから、これらをひとまとめにして類別するクラスタリング法が考えられる。この手法は、異なる一次回帰式で表されるような、多次元データのかたまり $\{x_1\}, \{x_2\}, \dots, \{x_z\}$ の類別に効果的であると考えられる（図4-4）。尚、このクラスタリング法と従来からのクラスタリングの諸手法（例えば文献(9)を参照）との関連性の考察は残された課題である。

次節では、これまでに考察した区分的線形近似の有効性、および4層 P P N のデータ集約機能の性質をシミュレーション実験により確認し、上で述べた4層 P P N を用いた実際の区分的線形因子分析の簡単な応用例を与える。



Data set in each cluster (C1 and C2) has the appropriate regression line (t1 and t2) respectively.

Fig. 4.4 Piecewise linear data regression and a clustering method by the regression lines.

図4.4 区分線形データ回帰(因子分析)と回帰直線群によるクラスタリングの概念

4. 4 シミュレーション実験

4. 2節における議論では、区分線形な特性をもつ3層ネットワークの写像能力を、“入力空間 R^N を、中間層サイズについて単調増加であるような個数の N 次元凸包（または超半平面）からなる、出力空間 R^M の超多面体に変換する能力”という表現で定式化し、これを包括する4層PPNの情報集約機能も同様な性質をもつことを示した。そこで、第3層を除く全素子が線形特性（第3層はSigmoid特性）をもつ実際のネットワークを用いて、単に評価関数の値を評価するのみでなく、上記の考察結果における幾何学的量（すなわち凸包または超半平面の数）を調べることを目的とする模擬実験を行い、4層PPNの集約能力ならびに、一般の3層ネットワークの写像能力を幾何学的に規定する。具体的には、4層PPNを用いて以下の実験を行う。

（実験の概要）

(1). 与えられた形状の1次元ならびに2次元多様体上に分布した標本を提示し、これを近似するために張られる凸包の数と中間層サイズの関連、また近似誤差とそれらの関連を調べる。

(2). アイリスデータ(11)（4次元ベクトルパターン）を標本として提示し、 R^4 のPL曲線近似、2次元多面体（PL曲面）近似を行うと同時に、母集団分布の区分的な線形回帰分析（以下、区分線形回帰と呼ぶ）を試みる。なお、実験における情報集約の度合いの定量的評価には、次式で表される平均二乗誤差率(Normalized Mean Square Error: NMSE)(10) δ を用いる。但し、標本を $\{X'_{mk}\}$ 、近似出力を $\{X_{mk}\}$ ($m=1, \dots, M$: 次元数) ($k=1, \dots, K$: 標本数)で表す。

$$\delta = \frac{\sum_k \sum_m (X_{mk} - X'_{mk})^2}{\sum_k \sum_m (X'_{mk})^2} \quad (4-14)$$

4. 4. 1 多様体上に分布した標本の近似(1)

(A). 2次元平面の上に多角形上(三角形, 四角形, 五角形)に分布した標本点(各々30, 40および50点)の1変数への次元削減(1次元集約)、および、

(B). 3次元空間中で球面上に分布した標本100点の2変数への次元削減(2次元集約)。

の各々を行ったシミュレーション実験について述べる。なお、本実験はいずれも3章の実験と同様、標本の提示順序は無作為に行い、重みとしきい値の初期値は乱数で与えた。また、逆伝搬学習の学習パラメータ修正の過程(いわゆる Simulated Annealing)は、誤差関数 e_n を減らす重み修正式(2-4)(t 回目の重みを $W(t)$ 、その修正分を $\Delta W(t)$ と表す)に慣性項を加えたもの

$$\Delta W(t+1) = \beta \cdot \Delta W(t) - \alpha \cdot (\partial e_n / \partial W(t)) \quad (4-15).$$

における、修正係数 α と慣性係数 β を各々、順に、(0.2, 0.9), (0.15, 0.85), (0.1, 0.8), (0.06, 0.5) で5000回ずつ(計20000回)の反復学習という要領で行った。

以下、簡単のため個々の実験を、その項目番号とネットワークの各層のサイズの組み合わせで呼ぶことにする(例えば、項目Bで第2、第3、第4層サイズが1、3、3の実験であればB133、等)。

A). 多角形上に分布した標本点分布

以下に、2次元平面において多角形上に分布した標本点集合を1変数 y (集約層サイズ1)に集約する実験例を示す(付録図Bおよび表4-1)。ただし、図中の丸印は標本点の位置を、太線は変数 y の、区間 $[0, 1]$ に対応する近似曲線(回帰曲線) $f(y)$ を、細線は標本点 x_k とそれを代表する近似出力 $f(y_k)$ のそれぞれを結んだものを示す。

ここで、同実験で得られた重みを式(4-7)にあてはめ、張られる区間の数を計算により求めた結果、A114, A124, A134, A144, A154の場合、それぞれ1, 2, 3, 4, 5となり(図Bの近似曲線の形状とほぼ一致)、標本分布はほぼ第3層サイズだけの線分数(1次元凸包数)の結合から成る折線によって近似されると同時に平均二乗誤差率が線分数の増加に伴い減少することが知られる(表4-1)。1変数集約に関する、これとほぼ等価な実験的検証とその解釈の例は、文献(4)(21)等で見られている。

B). 球面上に分布した標本点分布

次に、3次元空間において球面上に分布した標本点集合を2変数 y （集約層サイズ2）に集約する実験例を示す（付録図Cおよび表4-2）。

得られた重みの、上と同様の区間数を計算で求めたところ、B223, B233, B243の場合、 R^3 の球面上の分布は各々1, 3および5を得た（図4-3と同様の図を作成して確認）。これは、多角形分布と同様、近似多面体を構成する面の数に相当し、表4-2の結果と合わせて、この数の増加に伴い二乗誤差率は減少することが知られる。

4. 4. 2 アイリスデータの次元削減と区分線形回帰

以下では、多変量解析で古くからよく知られているアイリスデータ（4変数、個体数150からなる）(11)を、カテゴリ一名および標本分布の幾何学的構造が未知であるとして、1および2変数に次元を削減する(12)実験例を示す（付録図Dおよび表4-3）。なお、個々の実験の名称を、項目名を K として4.4.1と同様に、“ $K(N)(H)(M)$ ”で表す。

ここで、例として、K124の場合について、このデータの区分線形近似表現を定式化する（ただし、Sigmoid関数を式(4-3)の F_{pl} で近似する）。

<例4-1>（アイリスデータの区分線形回帰）

まず、得られた重み行列を R （入力層・集約層間）、 P （集約層・第3層間）および Q （第3層・出力層間）とし、第3層、出力層のしきい値ベクトルを各々 θ, ϕ とする。集約層サイズは1であるから、そこでの表現を y とすれば、 $P = \{p_{hn}\}$ ($h=1, 2; n=1$)となり、中間層表現 u との関係は、 F_{pl} を用いて

$$u_h = F_{pl}(p_{h1} \cdot y - \theta_h) \quad (h=1, 2) \quad (4-16).$$

で表せる。この式に実際の学習で得られた重み（ $p_{11} = -9.98$, $p_{21} = -9.42$, $\theta_1 = -0.516$, $\theta_2 = -7.70$ ）を代入し、

$$(u_1, u_2) = \begin{cases} (-2.49y + 0.629, 1) \cdots & (0 < y \leq 0.252) \\ (0, 1) \cdots & (0.252 < y < 0.605) \\ (0, -2.36y + 2.425) \cdots & (0.605 \leq y < 1) \end{cases} \quad (4-17).$$

を得る（付録図D.c）。更に、学習で得た $Q = \{Q_{mh}\}$ ($m=1, \dots, 4; h=1, 2$)および $\phi = (\phi_1, \dots, \phi_4)$ の値（省略する）を用いて式(4-13)の形にすれば、区分線形近似式

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \mathcal{L}(y) = \begin{cases} (-15.9y+7.6, -5.5y+3.28, -21.4y+6.7, -9.2y+2.48)^t & \dots (0 < y \leq 0.252) \\ (3.6, 1.9, 1.3, 0.2)^t & \dots (0.252 < y < 0.605) \\ (6.1y-0.1, 6.37y-1.95, 0.71y+0.87, 0.47y-0.1)^t & \dots (0.605 \leq y < 1) \end{cases}$$

(4-18).

を得る。各標本の特徴ベクトルに対する近似は、第 k 標本を y_k とし、 $\mathcal{L}(y_k) = \mathcal{L}(F(R_{1k}))$ により求められる。同様な回帰表現は集約層サイズが 2 以上の場合もこれと同等な方法で定式化できる。

なお、この式は実際、従来の非線形因子分析での、観測変数 x_m から潜在変数ベクトル y への非線形回帰式

$$x_m = \phi_m(y) + \text{err}_m \quad (4-19).$$

但し ϕ_m は非線形関数、 err_m は近似誤差。

の ϕ_m を、区分的に線形な表現で置き換えたものとなっている。

4. 4. 3 考察

以下では、上記の実験結果を 4. 2 節の議論との関連において考察する。まず、4. 4. 1 における実験の結果から、以下のことが知られる。

a). 4 層 P P N は、1 次元集約、2 次元集約ともに、第 3 層サイズを適当に設定すれば、区分的に線形な分布モデルあるいは非線形な分布モデルに従う標本集合を、区分的にほぼ線形に近似（折線回帰、P L 曲面回帰）することができる。また、1 次元集約に関する限り、各標本に対する集約層表現 y_k は、その“回帰折線”上で、対応する標本に対してユークリッド距離が最小となる点に写像される値に収束している。従って、集約層でほぼ最適な集約表現を獲得していることが知られる。これは定理 1 の結果と一致する。

b). 4 層 P P N と K L 展開との比較においては、第 3 層サイズが集約層サイズと等しいとき（3 層 P P N と等価）に集約効率はほぼ同等となり、第 3 層サイズ > 集約層サイズであれば 4 層 P P N の集約効率が K L 展開のそれを上回る。これは補

題1の結果と一致する。

c). 第3層サイズを増やしていく際、同層が効率のよい表現を獲得しうるサイズに達するとき(計算による凸包数(図B, C)と、標本が潜在的にもつ区間の数とが一致したとき)に近似誤差の減少分が最大となる(A124, B234)。これは4.2節の結果と一致する。

d). 集約層サイズ N 、第3層サイズ H と、4節における張り得る凸包数 z との関係として、ここで行った実験に関しておおまかには $z \approx {}_H C_N$ が成り立つと予想される。この定性的根拠として以下が予想されるが、現在のところ厳密な証明は得られていない。

d-1). 凸包数 z の問題は式(4-7)中、 H 個のうち同時に N 個以上が第2区間に入るような組み合わせ数の最大化問題に帰着すると考えられる。このとき、

d-2). $\{(p_i, \theta_i) | (i=1, \dots, H)\}$ の H 個の対のうち、どの N 個の組合わせを選んでも、それらが一般の位置にあれば、 $p_i \cdot y - \theta_i = 0$ ($i=a_1, \dots, a_N$)の解が存在する。従って、 H 個のうちの任意の N 個が同時に $-2 \leq (p_i \cdot y - \theta_i) \leq 2$ となる $y \in R^N$ が存在するため、 $z \geq {}_H C_N$ が成り立つ。

d-3). 逆に、ニューラルネットの最急降下法による修正では、各 p_i, θ_i の収束状態において $-2 \leq (p_i \cdot y - \theta_i) \leq 2$ を満たすような (p_i, θ_i) の組が $N+1$ 個以上できるような y が存在する確率は少ないと考えられる。

次に4.4.2の実験からは、

e). 表4-3の中、K124による、区間数2の区分線形近似の場合の誤差は、K114もしくはKL展開1項近似のその約半分であることから、b)およびc)との関連で、アイリスデータは、直線による近似よりはむしろ、2個の線分からなる折線でよく近似されるような構造をもつこと(Versicolor種およびVirsinica種を合わせて1本の線分で、Setosa種をもう一本の線分で近似する)が知られる。例4-1による回帰表現は、これを定式化したものとなっている。

これらのことから、4層PPNが(線形回帰を含めて)3層の場合と比較してはるかに優れた情報集約能力をもつことは明かである。しかもこれらの実験結果は、いずれも本章4.2および4.3節で行った考察を強く裏付けるものである。

4. 5 まとめ

前章と本章では、多層ニューラルネットを用いた情報集約のための有効なツールである P P N の性質を、各々 3 層および 4 層の場合について体系的に論じた。すなわち、前章では 3 層 P P N を含む 3 層ネットワークによる次元削減の問題と、線形回帰の手法の等価性について議論した。次に本章では、一般の 3 層ネットワークの非線形写像能力を区分的に線形近似して考察し、それに基づいて、4 層 P P N が潜在的に区分的線形回帰の能力をもつことを理論的に示し、またそれらの議論の正当性をシミュレーション実験により検証した。

これまで、多層ネットワークをデータ圧縮および次元削減（情報集約）に用いた試みのほとんどは、3 層のそれに限定した実験事例 (10) (13)-(16) ならびに理論的考察 (17)-(20)、もしくは 4 層以上までを考慮してはいるが実験事例の報告のみの枠内にあるもの (1)-(4)、のいずれか一方に分類されるといって良い。本章では理論的な厳密性を欠くことなしに、3 層以上のネットワークの情報集約機能に言及し、3 層と 4 層のネットワーク（具体的にはパルス入力-パターン出力写像を学習するネットワーク）による同機能の幾何学的性質を明かにした。

なお、これらの考察は、極めて直接的な方法で、砂時計モデルなど、P P N 以外の多層モデルに容易に拡張することができるが、いずれの場合も実際には、その最小解以外の極小解への収束が不可避であることが大きな問題であり、収束状態の学習の諸係数に依存した変動は回避できない。そこで筆者らは、4 層 P P N の収束誤差の点での優位性を、これと等価な集約能力をもつと考えられる 5 層砂時計回路との比較実験を通して実際に確認している (4)。

本章ではまた、P P N を含む多層ネットワークの動作の考察に、組み合わせ位相幾何学的なアプローチを用いた。すなわち、Sigmoid 状の S 字型非線形関数は、全て本稿と同様の方法で、3 つの区間からなる区分線形関数で近似され、構成素子がそのような特性をもつ場合、多層ネットワークによる非線形写像 f は、何等かの P L 写像におき替わる。従って、得られた P L 写像の代数的性質との関連において、ネットワークの写像能力を規定することができる。これにより、多層ネットワークの情報集約機能と区分線形回帰との間の幾何学的対応が、本稿ではじめて、具体的に与えられた。これに加えて、4. 3 節で提案した多層ニューラルネットによる区

分線形因子分析とクラスタリングの各手法の有効性を示唆した。

これらの意味で、本研究が多層ネットワークの数理的研究の分野、ならびに統計的パターン認識手法の理論研究の分野にもたらす意義は大きいといえる。

さて、前章および本章で残された課題としては、以下が挙げられよう。

- ・ P P Nにより得られた回帰写像と、砂時計モデルの圧縮写像の対応および砂時計モデルの強化法の実現。
- ・ 非線形性の利点を活かし、収束状態の変動をなくすための、多層ネットワークにおける極小解回避の技法の開発（パターン分類問題、情報集約問題の両面）。
- ・ 4. 2節で述べた4層 P P Nのもつ P L写像の性質の、より詳細な解析（トポロジカルな性質の解明）および4. 3節で触れた区分線形因子分析手法とクラスタリングの、より詳細な性質および実際のパターン認識などへの応用。

そこで、続く5章では、データ圧縮機能を実現する恒等写像と、因子分析のモデルであるパルス入力・パターン出力写像の関連を解明し、砂時計モデルのデータ圧縮能力を強化する手法について述べる。そして更に、前章と本章の手法により得られる集約表現の判別分析への応用に関する有効性を検討する。

表 4. 1 多角形上の標本分布の 1 次元集約例 (枠内は NMSE: 単位は % で表示)

	A 1 2 2	A 1 3 2	A 1 4 2	A 1 5 2	K L 展開 1 項近似
三角形	0.992	0.148	0.147	0.146	5.405
四角形	1.451	0.494	0.0697	0.0670	9.716
五角形	1.368	0.565	0.1601	0.0636	6.129

表 4. 2 球面上の標本分布の 2 次元集約実験例 (枠内は NMSE: 単位は % 表示)

B223	B233	B243	K L 展開 2 項近似
2.970	0.760	0.261	2.966

表 4.3 4層 P P N によるアイリスデータの集約例 (枠内は NMSE: 単位は%)

1次元集約	K 1 1 4	K 1 2 4	K 1 3 4	K 1 4 4	K L 展開 1 項近似
2乗誤差率	0.549	0.279	0.277	0.273	0.538
2次元集約	(K 2 1 4)	K 2 2 4	K 2 3 4	K 2 4 4	K L 展開 2 項近似
2乗誤差率	0.540	0.164	0.125	0.115	0.160

参考文献

- (1). 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 - テーブル
ルックアップと違うのか? -", 信学技報, NC89-15, pp. 33-40(1989).
- (2). 臼井支朗, 中内茂樹, 中野正恵: "多層ニューラルネット恒等写像モデルによる
マンセル色空間の特徴解析", 信学技報, NC89-40, pp. 45-50(1989).
- (3). 片山泰男, 大山公一: "自己組織逆伝搬ニューラルネットの諸特性", 平成元年信
学春大, 7, pp. 309-310(1989).
- (4). 宮崎慎也, 米倉達広, 鳥脇純一郎: "サンプル分布の幾何学的構造の解析におけ
る自己想起型回路と P P N の関連について", 信学技報, NC89-102, pp. 73-78(1990).
- (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, MBE
88-52, pp. 127-134(1988).
- (6). Irie, B. and Miyake, S. : "Capabilities of Three-layered Perceptrons",
Proc. IEEE ICNN-88, pp. 1641-647(1988).
- (7). Rumelhart, D. E. and McClelland, J. L. : "Parallel Distributed Processing",
Vol. 1 and Vol. 2, MIT Press(1986)
- (8). 本間龍雄: "組合せ位相幾何学", 共立全書(1980).
- (9). Anderberg, M. R. : "Cluster Analysis for Applications", Academic Press, New
York (1973).
- (10). Cottrell, G. W., Munro, P. and Zipser, D. : "Image Compression by Back
Propagation", Tech. Rep. 8702, Univ. of California, San Diego, Institute for
Cognitive Science(1987).
- (11). Anderson, T. W. : "An Introduction to Multivariate Statistical Analysis"
2nd Edition, John Wiley & Sons, New York(1984).
- (12). Fukunaga, K. : "Introduction to Statistical Pattern Recognition",
Academic Press, New York(1972).
- (13). Elman, J. L. and Zipser, D. : "Learning the Hidden Structure of Speech",
Tech. Rep. 8701, Univ. of California, San Diego, Institute for Cognitive
Science(1987).

- (14). 曾根原登, 川人光男, 三宅誠, 中根一成: "ニューラルネットによる画像データ圧縮 (Neuro-CODEC) の検討", 信学技報, IE88-62, pp. 57-64 (1988).
- (15). 長坂保典, 岩田彰, 鈴木宣夫: "ニューラルネットによる長時間心電図の特徴抽出と情報圧縮", 信学技報, MBE88-91, pp. 73-80 (1988).
- (16). 米倉達広, 横井茂樹, 鳥脇純一郎: "ニューラル・ネットを用いた濃淡画像の表現素展開による画像表現", 信学会 画像理解の高度化と高速化シンポジウム論文集, pp. 127-132 (1989-04).
- (17). 船橋賢一: "3層ニューラルネットワークによる恒等写像の近似的表現についての理論的考察", 信学論 (A), J73-A, 1, pp. 139-145 (平02-01).
- (18). Bourlard, H. and Kamp, Y.: "Auto-Association by Multilayer Perceptrons and Singular Value Decomposition", Biol. Cybern. 59, pp. 291-294 (1988).
- (19). Baldi, P. and Hornik, K.: "Neural Networks and Principal Component Analysis: Learning from Examples Without Local Minima", Neural Networks, 2, pp. 53-58 (1989).
- (20). 米倉達広, 横井茂樹, 鳥脇純一郎: "多層ニューラルネットによる情報集約の一手法とその理論的考察", 信学論 (D-II), J73-D-II, 8, pp. 1205-1212 (平02-08).
- (21). 米倉達広, 横井茂樹, 鳥脇純一郎: "ニューラルネットを用いた画像展開の一手法とその応用", 信学技報, PRU89-22, pp. 25-32 (1989).

第5章

多層ニューラルネットによるデータ圧縮機能、因子分析機能

外的規準の無い場合の多変量解析手法としては、主成分分析、因子分析、クラスター分析、潜在構造分析、などが挙げられる(1)。前章までの各章では、これらの諸手法に相当する、多層ニューラルネットの具体的なモデル(主成分分析に対応する砂時計モデル、因子分析に対応するPPN)の幾何学的性質について考察した。すなわち、第3章では2層の復号化回路をもつ3層砂時計と3層PPNを、第4章では3層の同回路をもつ4層PPNを扱った。そこで、本章では砂時計モデルとPPNを数理的に対応付け、両モデルにより得られるデータ圧縮写像 T_1 とデータ回帰写像 T_2 の関連について考察する。

5.1 多層ニューラルネットによるデータ圧縮と因子分析

データ圧縮を行う多層ニューラルネットである砂時計モデルおよび因子分析のモデルであるPPNは、いずれも式(2-13)で定式化されるデータ回帰写像を有する点で共通している。また、第2章で述べたように、PPNにおける集約層の位置は、入力パターンどうしが干渉し合うことがないため第2層にとることができるのに対し砂時計モデルの集約層は、 T_1 と T_2 の対称性を考慮して、通常入出力層間の中央部にとり、各層の素子数は集約層を中心に対称に(例えば、集約層を第3層として各層素子数： $M-N-d-N-M$ のように)とることが望ましいとされている(2)(3)(4)。これらのことから、対応する規模の両モデルでデータ回帰の機能のみを比較した場合、収束速度、収束精度(収束時の誤差の大きさ)、1回の反復に要する計算量のいずれの点においても、PPNは砂時計モデルと比較して有利である(3)(5)。

従って、非線形な効果を生かしたデータ回帰写像 T_2' (すなわち因子分析モデル) を実現する P P N の重みパラメータセットが与えられているという条件のもとで、その近似的な逆写像としてデータ圧縮写像 T_1 を求めることは、砂時計モデルの圧縮能力を越える写像 T_1 が比較的容易に得られることが予想される。そこで、以下では 3 層ならびに 4 層 P P N で得られたデータ回帰写像 T_2 から、これらに対応する 3 層および 5 層砂時計モデル中のデータ圧縮写像 T_1 を求める手法を考察する。

5.1.1 3 層砂時計モデルと 3 層 P P N

3 層砂時計モデルは、第 k 標本を x_k , その近似出力を x_k' , 集約表現を y_k で表し、 F_h, F_o を各々中間層素子と出力層素子の入出力特性 (入力ベクトルの個々の成分を Sigmoid 変換あるいは線形変換したものを成分としてもつベクトル), L_1, L_2 を各々、入力層・中間層間 (d 行 M 列), 中間層・出力層間 (M 行 d 列) の重み行列、 α を出力層の各素子のしきい値を縦に並べた (M 次元) しきい値ベクトル (中間層のしきい値は考慮しなくとも一般性を失わない) として、

$$y_k = T_1(x_k) = F_h(L_1 x_k) \quad (5-1).$$

$$x_k' = T_2(y_k) = F_o(L_2 y_k - \alpha) \quad (5-2).$$

で定式化される。

これに対し、3 層 P P N は、 p_k を第 k パルス入力, L_0, L_2' を各々、入力層・中間層間 (d 行 K 列), 中間層・出力層間 (M 行 d 列) の重み行列、 α' を出力層のしきい値ベクトルとして、同様に

$$y_k = F_h(L_0 p_k) \quad (5-3).$$

$$x_k' = T_2'(y_k) = F_o(L_2' y_k - \alpha') \quad (5-4).$$

で表される。

いま、3 層 P P N で得たデータ回帰写像 T_2' を実現する L_0, L_2', α' から、このデータ回帰に対応するデータ圧縮写像を実現する L_1 を求めるために、 $L_2 = L_2', \alpha$

$= \alpha'$ として式(5-1)に式(5-3)を代入し、全ての k についてまとめて書けば、次式を得る(5)。

$$L1 = L0X' \quad (5-5).$$

但し、 X は M 次元列ベクトル $x_k (k=1, \dots, K)$ を横に K 個並べた M 行 K 列の行列、 X' は行列 X の疑似逆行列(6) ($K \geq M$ と仮定)を表す。

また、特に F_h, F_o を共に線形特性、 $\alpha =$ 零ベクトル、とするとき、 $L1L2' \approx E$ (d 行 d 列の単位行列)と考えられるから、以下の式が成り立つ。

$$L1 = (L2')' \quad (5-6).$$

但し $(L2')'$ は $L2'$ の疑似逆行列 ($M \geq d$ と仮定)を表す。

5.1.2 5層砂時計モデルと4層PPN

4層PPNのデータ回帰写像 $T2'$ は、任意の連続写像を実現できるとされる3層ネット(シグモイド特性を持つ中間層素子数を無限に増やせば任意の連続写像を任意の精度で近似できる(12)(13))の復号化回路により実現される。以下にこれと5層砂時計モデルを対応付ける。

いま、集約層・第3層間の重み行列、第3層・出力層間の重み行列を各々 $B3, B4$ で、第3層、第4層のしきい値ベクトルを各々 θ, ϕ で表せば、これは次式で定式化できる。

$$T2'(y_k) = F_o(B4F_s(B3y_k - \theta) - \phi) \quad (5-7).$$

但し F_s, F_o は、各々第3層、第4層の素子の特性を表す。

そこで、これに対応する5層砂時計モデルのデータ圧縮写像 $T1$ が $T2'$ の近似的逆写像であると考えられることから、前節と同様にして、

$$T1(x_k) \approx B3'(F_s^{-1}(B4'(F_o^{-1}(x_k) + \phi)) + \theta) \quad (5-8)$$

ここで $F_s^{-1}(x)$ は、ベクトル x の個々の要素に対し、入出力特性 F_s の逆関数 F_s^{-1} を施したものを要素としてもつベクトル(F_o^{-1} も同様)を、 $B3', B4'$ は、各々行列 $B3, B4$ の疑似逆行列を表す。

従って、5層砂時計モデルは、この式による圧縮写像 T_1 を実現するように、入力層・第2層および第2層・集約層間の重み行列（各々 B_1, B_2 とする）が設定されれば、式(2-18)で示される4層PPNの集約能力とほぼ等価な圧縮能力を持つことができるかと予想される。しかしながら、実際の5層ニューラルネットの第1, 2, 3層（各層素子数 $M-N-d$ ）では、 $F s^{-1}$, $F o^{-1}$ の操作の実現が困難であるため、式(5-8)を正確には実現できない。

そこで、5層砂時計モデルで効率の良い写像 T_1 を得るための手段として以下が考えられる。すなわち、4層PPNで得られた集約表現 y を教師パターン、多変量データ x を入力パターンとした3層ニューラルネット（各層素子数 $M-N-d$ ）の学習を行うことである。この有効性は、二、三の予備的な実験の結果として報告されている(3)。

5.2 シミュレーション

以下に、前章で述べた方法により得られるデータ圧縮写像 T_1 の圧縮能力を、二、三の予備実験により検証する。

ここで行う実験の目的は、砂時計モデルによる学習で得られる写像 T_1, T_2 と、PPNによる学習で得られる T_0, T_2' から合成される、恒等写像中のデータ圧縮写像 T_1' の効率を、2乗誤差の大小により比較することである。ここで実験用入力パターンとしては、アイリスデータ(7)を用いる。

表5-1に、アイリスデータ（150個体、4変数からなる）の各変数の区間 $[0, 1.0]$ を各素子の出力値 $0, 1$ 間の連続値に線形に変換したものに対する、集約層素子数 d を変えた場合の3層モデルでの圧縮能力（ T_1 と T_1' ）の比較を示す。但し、各々の数字は式(2-15)の平均2乗誤差 ε^2 を各々の標本ベクトルの絶対値の総和で割った正規化平均2乗誤差率（Normalized mean-square error - NMSE）

$$\delta = \varepsilon^2 / \sum_k |x_k|^2 \quad (5-9).$$

を表す。また、表5-2には同データを、砂時計モデル（素子数 $4-N-d-N-4$ ）で学習した場合の情報損失量 δ と、4層PPNモデル（同 $150-d-N-4$ ）で学習し

た結果の T_2' を 5 層砂時計モデルの集約層以降の重みの初期値として移植した上で更に T_1 , T_2 を学習した場合のそれとの比較を挙げておく ($(d, N) = (2, 2), (2, 4), (3, 3), (3, 4)$ の 4 種類)。更に、付録図 E として、4 層 P P N (150-2-4-4) での集約表現 y' と 5 層砂時計モデル (4-4-2-4-4) でのそれ y 、および K-L 展開 2 項近似での集約空間の全標本に対する散布図を示す。

なお、一組のパターンのサンプルと個々のパターンベクトルのカテゴリーが既知である場合、統計的パターン認識手法 (8) として、パターンの次元数を削減後に判別分析処理を行う手法はよく知られている。そこで、本実験でも、上で述べた評価の他に、次元数削減後の判別分析 (3 クラス分類) への有効性の評価をつけ加えた。なお、本文の実験結果はニューラルネットが収束したとみなされる状態まで学習させた結果である。収束の判定は視察によったが、結果的には、5000 回の反復の間の NMSE の全変化が 0.0007 ~ 0.0035% 以下になったとき収束したとみなされた。

以下、表 5-3、5-4 として集約表現 y および y' の空間における、下式 (5-10)、(5-11) による判別基準 (7) $J_1(\Psi)$, $J_2(\Psi)$ (クラス間の分離度の指標) の値、および表 5-5 として、アイリスデータに対して K-L 展開 2 項近似を施した後の集約データに対するそれらの値、および同データに対して直接判別分析を行った際のそれらの値を与える。

$$J_1(\Psi) = \text{tr}(\Sigma_T^{-1} \Sigma_B) \quad (5-10).$$

$$J_2(\Psi) = \ln(|\Sigma_T| / |\Sigma_W|) \quad (5-11).$$

但し、 Σ_T , Σ_B , Σ_W は各々、集約パターン空間 (または標本空間) Y から判別特徴空間 Z への写像 $z = \Psi(y)$ が与えられたときの、分散共分散行列、クラス間共分散行列、およびクラス内共分散行列、 $|\Sigma|$ は行列 Σ の行列式、 $\text{tr}(\Sigma)$ は行列 Σ の対角和を、 $\ln(x)$ は正数 x の自然対数を表す。

これらの実験結果の考察として、以下が挙げられる。

- 1) 3 層砂時計モデルで学習した結果得た T_1 と比較して、式 (5-6) で計算した重み行列による写像 T_1' によるデータ圧縮の誤差は、ほぼ同一であり、主成分分析の結果ともほぼ一致する。
- 2) 5 層砂時計モデルのみで学習した結果得た T_1 と比較して、4 層 P P N の T_2' を

用いて5層砂時計モデルの学習を強化して得たT1による圧縮の誤差は、最大で1割程度少ない。

3) 5層モデルの場合、 $d = N$ としたときは、主成分分析および3層モデルとほぼ同一の結果となる。

4) 集約空間における判別基準の計算からは、4層PPNおよび5層砂時計による次元数削減が判別分析の前処理として特に有効であるという結果は得られなかった。(例えば、表5-3、4の $(d, N) = (2, 2)$ の場合の $J_1(\Psi)$ は表5-5の $J_1(\Psi)$ の値より約1割小さいことに注意)。

ここで、4)の原因としては、図5-1の集約空間における散布図を見れば解るように、ニューラルネットワークによるデータの写像が非線形であること、一方判別分析は線形写像の範囲で評価関数 $J_1(\Psi)$ 、 $J_2(\Psi)$ を最大化するものであること、従ってこの評価関数のみでニューラルネットワークの写像は必ずしも適切には評価できない、という点があげられる。実際、散布図から見る限りではニューラルネットワークによる集約空間におけるクラス分離は、判別特徴空間のそれと比べて遜色なく、部分的にはむしろ良くなっている。

これにより、前章で行った考察の正当性が種々のケースで検証された。

5. 3 まとめ

本章では、多変量データの圧縮機能ならびに因子分析機能に対応する多層ニューラルネットワークのモデルである、砂時計モデルとPPNモデルの数理的対応付けを試み、統計的パターン認識への有効性を示唆した。すなわち、まず5.1.1節でその幾何学的性質を対比した3層砂時計モデルと3層PPNモデルの数理的關係を、次に5.1.2節で5層砂時計モデルと4層PPNモデル(この幾何学的解釈は4章で与えた(15))のそれを明かにし、3層の両モデルの重み行列間の変換式を与えた。そして最後に、5.2節でAndersonらが用いたアイリスデータ(7)に対して、これらの議論を裏付けるシミュレーション実験の結果を示すとともに、次元が削減された空間での判別基準の評価により、従来の線形判別分析との比較も試みた。

多層ニューラルネットワークと多変量解析の諸手法の関連については、文献(9)-(14)等で報告されているが、外的基準のない主成分分析と因子分析を統一的に扱い、多層

表 5.1 3層モデルの圧縮写像 T1, T1' と主成分分析の誤差比較 (単位は%)

圧縮次元数	砂時計モデルで得た T1	PPN のデータ回帰写像 T2 から合成した T1	主成分分析
1次元	0.5387	0.5384	0.5384
2次元	0.1595	0.1594	0.1594
3次元	0.0373	0.0372	0.0372

表 5-2 5層砂時計モデルのデータ圧縮写像 T1 の誤差比較 (単位は%)

(d, N)*	砂時計モデルで得た T1	PPN で得た T2 から合成した T1
(2, 2)	0.1610	0.1607
(2, 4)	0.1096	0.1100
(3, 3)	0.0402	0.0392
(3, 4)	0.0357	0.0319

*注: d = 集約表現の次元数、N = 第4層の素子数

表 5.3 4層 P P N モデルの集約層における判別基準値

(d, N)	$t r (\Sigma_T^{-1} \Sigma_B)$	$\ln (\Sigma_T / \Sigma_W)$
(2, 2)	1. 0 8 8	3. 1 3 5
(2, 4)	0. 9 0 1	2. 1 7 2
(3, 3)	1. 1 7 4	2. 8 8 7
(3, 4)	1. 0 9 7	3. 4 1 0

表 5.4 5層砂時計モデルの集約層における判別基準値

(d, N)	$t r (\Sigma_T^{-1} \Sigma_B)$	$\ln (\Sigma_T / \Sigma_W)$
(2, 2)	1. 0 7 8	3. 8 3 3
(2, 4)	0. 9 9 8	2. 8 1 7
(3, 3)	1. 1 5 3	3. 9 5 8
(3, 4)	1. 0 9 3	3. 9 2 3

表 5.5. 判別分析モデル、およびKL展開の集約表現における判別基準値

	$tr(\Sigma_T^{-1}\Sigma_B)$	$\ln(\Sigma_T / \Sigma_u)$
判別分析(2次元)	1.192	3.753
KL展開2項近似	1.068	3.478

文 献

- (1). 芝祐順, 渡部洋, 石塚智一 編: "統計用語辞典", 新曜社 (1984).
- (2). 臼井支朗, 中内茂樹, 中野正恵: "多層ニューラルネット恒等写像モデルによるマンセル色空間の特徴解析", 信学技報, NC89-40, pp. 45-50(1989).
- (3). 宮崎慎也, 米倉達広, 鳥脇純一郎: "サンプル分布の幾何学的構造の解析における自己想起型回路とPPNの関連について", 信学技報, NC89-102, pp. 73-78(1990).
- (4). 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 - テーブルルックアップと違うのか? -", 信学技報, NC89-15, pp. 33-40(1989).
- (5). 米倉達広, 横井茂樹, 鳥脇純一郎: "ニューラル・ネットを用いた濃淡画像の表現素展開による画像表現", 信学会, 画像理解の高度化と高速化シンポジウム論文集, pp. 127-132(Apr. 1989).
- (6). Rao, C. R. and Mitra, S. K. 著, 渋谷, 田辺訳: "一般逆行列とその応用", 東京図書(1973).
- (7). Anderson, T. W.: "An Introduction to Multivariate Statistical Analysis" 2nd Edition, John Wiley & Sons, New York(1984).
- (8). 大津展之: "パターン認識における特徴抽出に関する数理的研究", 電総研研究報告第818号 (1981).
- (9). Gallinari, P., Thiria, S. and Fogelman Soulie, F.: "Multilayer Perceptrons and Data Analysis", Proc. IEEE ICNN-88, pp. 1391-399(1988).
- (10). Asoh, H. and Otsu, N.: "Nonlinear Data Analysis and Multilayer Perceptrons", Proc. IJCNN-89, pp. 411-415(1989).
- (11). Bourlard, H. and Kamp, Y.: "Auto-Association by Multilayer Perceptrons and Singular Value Decomposition", Biol. Cybern. 59, pp. 291-294(1988).
- (12). 船橋賢一: "3層ニューラルネットワークによる恒等写像の近似的表現についての理論的考察", 信学論(A), J73-A, 1, pp. 139-145(平02-01).
- (13). Irie, B. and Miyake, S.: "Capabilities of Three-layered Perceptrons", Proc. IEEE ICNN-88, pp. 1641-647(1988).
- (14). 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, pp.127-134(1988).

(15).米倉達広, 宮崎慎也, 鳥脇純一郎:”恒等写像モデルおよびPPNに基づく4層ニューラルネットの情報集約機能の解析”,信学論(D-II), (採録決定).

(16).米倉達広, 鳥脇純一郎 :”4層恒等写像型ニューラルネットによる疑似区分線形因子分析”, 信学論 (D-II) , (投稿中)

第6章

あとがき—まとめと今後の課題

6.1 総括

本文では、多層ニューラルネットを用いた、次元数削減の問題、もしくは、情報を外的規準無しで集約的に表現する、情報集約の問題を統一的に扱った。

パターン次元数削減の問題（多次元データ集合を何等かの量を保存して、より低次元の空間に写像する問題）は、従来から、教師なし分類、データ圧縮の問題（多次元データの低次元空間への写像と、その低次元空間から元のデータを復元する写像を求める問題）、および、多次元データの可視化（Visualization）やその会話的解析（Interactive Analysis(1)：低次元に写像したデータを、視察によって解析する実用的手法）と密接に関連している。また、データ圧縮は伝送工学の分野の、可視化や会話的解析はパターン認識の分野の主要な問題の一つである。

多変量解析に転じてみれば、そこでの外的規準無しの問題は、潜在的なカテゴリー構造を浮き彫りにするための因子分析(2)（多変量統計データを、より少ない次元のデータの線形結合で集約的に表現する問題）や、考え方は古くからあるがその裾野の広さ故に、いまなお神秘的で極めて大きな関心の持たれるクラスター分析(3)(4)（何等かの類似性の尺度をもって、類似データどうしを各々のクラスにまとめる問題）に区分される。これらのいずれもが、社会的にも学術的にも意義深い問題である(5)。

これらの問題は、外的規準付き、教師付きの問題とは対照的に、カテゴリーに関する構造情報が全く知られていないケースを想定する。利用できるのは観測データのみであり、その目指すところは、観測データに適合する潜在的な規則を見いだして（因子分析）それにより冗長な部分を削り落としたり（データ圧縮）、その規則を利用してカテゴリーの構造を発見（クラスター分析）したりすることである。

一方、外的規準（教師）付きか無しかの、いわゆる問題設定に関する区分という見方の他に、そこで適用される写像が、連続写像（連続関数的）によるものかどうかという、道具の性質からの観点があろう。すなわち、情報の集約的表現の獲得に関していえば、構文的パターン認識の手法や知識工学の諸手法、および階層的クラスタリングの範疇にある手法等は、“不連続な”写像によって、多次元パターンをより低次元の空間や記号情報（すなわち硬い情報）ないしはカテゴリー空間へと写像する。これに対して多変量解析における手法の多く（例えば判別分析や主成分分析、因子分析等）は明かに、ベクトル空間からベクトル空間への“連続写像”を用いてこれを行っている。そしてこの“連続”であることのメリットは、以下に示すように我々の予想以上に大きい。

パターン情報をおある空間から別の空間へ写す場合、その写像が連続写像であることは、多次元空間でパターンの分布が織りなす、“位相的性質”を保存（近い点どうしは、写像される空間においても近い位置に写像されるような性質）することにつながる。加えて、例えば多層ニューラルネットの入出力写像は連続で、かつ、各層における非線形特性（例えばシグモイド関数）の入れ子構造（層数の増加）と重ね合わせ構造（素子数の増加）が、任意の連続写像の近似的実現を可能とする(6)ことから、多層ニューラルネットは関数の補間を実現する能力を持ち(7)、その情報集約過程においては多次元特徴空間のいかなる位相的性質をも保存し得る能力をもつことが予想できる。従って、そのようなパターン分布の位相的性質を問題にすべき場合において、それが柔軟な非線形連続写像であることは、学習の主要素である“汎化の能力”(8)（与えられていない問題であっても、既に与えられた問題の一般性を利用して、それなりに答えられる能力）を活用できることにつながる。何故なら、連続写像であれば、幾つかの代表的な点についての情報を与えるだけで、それらの間をシステム自身が適当に補間できてしまうからである(8)(9)。

そこで、次の問題は、我々が一般にパターン認識で扱う問題について、特徴空間のパターン分布が“位相的性質”を持つかどうかという問題である。筆者は自信をもって（当然のことながら特徴空間の選び方に依存する部分はあるが）殆どすべての場合にこれが成り立つとすることができる（この位相的性質の解明こそが個々のパターン認識問題の本質を決定する鍵ではないであろうか）。逆のいいかたをすれば、位相的性質を、問題に応じたうまいかたち（綺麗な形）で浮き彫りにすることこそ

特徴量の適正な選び方、すなわち特徴選択の良否であるといえよう(10)。

さて、本論に立ちもどって、多層ニューラルネットは明かに連続写像によるパターン変換を行っている。一方、連続写像によるパターンベクトルの変換という意味で最も類似した性格を持つ分野は、多変量解析に他ならない。そこで、外側すなわち使用する機能の面から多変量解析と多層ニューラルネットを関連付け、更に、内側すなわち数理的性質の面からそれらを対比させるといふ考え方をとることが、本文をまとめる目的であった。しかしながら、実際に生み出すことができた成果といえは、外的規準なしの場合のみについての基礎的な考察であり、しかも扱ったデータが、生データとしては、線形モデルでよく近似し得るアイリスデータのみであったことから、これらの手法の実際上の有効性をうまく提示できなかつた点でかなり課題(あるいは悔い)が残っている。

しかし、本文の内容により、これまでベールに包まれていた多層ニューラルネットの種々の能力や性質のうちの、小さな一片が垣間みえるようになったのでは、というのが筆者の希望的観測である。

本文での多層ニューラルネットの位置付けは、従って、非線形な多変量解析を行う効率的な並列処理マシンということになる(というには本文の内容が竜頭蛇尾に過ぎるかも知れないが)。そして、この並列処理マシンにより(特に4層PPNや5層砂時計モデルを用いれば)非線形性を極めて上手に活用した、多次元データの情報集約(データ圧縮、因子分析、クラスター分析)が可能であるということを示したつもりである。ここで、筆者のこの立脚点が、あまりにも歪小な捉え方ではないか、というコネクショニスト先生方のお叱りを受けることは当然予想している。しかしながら、新たな計算モデルについて、その基本的な数理を、まずは現在体系化されている分野に結び付けて考え、その中から新たな方法論や概念を探索し生み出していくということが、情報科学に限らず種々の分野において研究者の取るべき立場であろうというのが、それに対する筆者の反論である。

これらのことを踏まえ、いずれ近い将来、次節で述べるような課題を明かにすることにより、筆者あるいは他の研究者が、本来の目的に沿って本文を補い、これを完成することを確信している。

6.2 今後の課題

前節で述べたように、本研究は、その目的とするところがかなり大きい反面、実際に本文中で著すことができた成果は未だ小さなものである。

そこで、本研究で明かにすべき事柄で、本文で達成できなかったものを各章毎に列挙すると、以下のようになろう。

第2章では、多層ニューラルネットと逆伝搬学習、およびそれらを用いて恒等写像を近似する砂時計モデルによるデータ圧縮、PPNによる因子分析の機能を規定した。ここでの課題は、他の形式の多層ニューラルネットで同様の機能をもつ種々のモデルを少しでも多く紹介することである。例えば文献(11)のように、砂時計モデルではあるが、複数素子で唯一の数字(Number)を表すように仕組んだ3層砂時計モデルの計算論は、ここで述べた枠組み(いわゆる線形回帰と等価な解釈)には納まらない。また、ここでの議論において、教師なし学習という点で共通なKohonenのFeature Map(12)や福島のネオコグニトロン(13)との計算論的な比較(あるいは幾何学的対応)をおこなえば、ニューラルネット研究全体の中での本研究の位置付けを、より明確にすることができたかもしれない。

第3章では、3層砂時計モデルと3層PPNを取り上げて、3層ニューラルネットの情報集約機能に対して幾何学的解釈を与えた。ここでの最も重要な課題は、出力層をシグモイド特性とした場合の、多次元データの区間のとり方に関する最適性の問題である。この候補としては、出力層素子の先に線形特性をもつ“足”の層(出力層素子と足層の素子の数は等しく、両層素子の間に一对一の結線を持たせる)を置き、教師データは足の層から提示することで区間幅の学習も同時に実現するようなモデルを考案することができる。

なお、この章の内容は文献(14)で明かにされている。

第4章は、本研究全体の中で筆者が最も強調したかった部分である。ここでは、従来の多変量解析で対応するものの無い、(準線形ではなく)非線形データ回帰の概念とそれを行う具体的な計算モデル、すなわち4層PPNの説明をした。また、4層PPNは、定性的には区分線形因子分析と呼ぶべきモデルで定式化されること、を示すと同時に、これを用いた全く新しいクラスター分析法を提案し、それを非常

に簡単なデータで実験的に検証した。

ここでの第1の課題は、この集約写像の位相的性質の解明である。すなわち、 n 次元空間から d 次元空間への4層PPNによる写像で、一体何が保存されるのか？この質問は次に、多層ニューラルネットで何ができないのかという質問(7)につながる。第2の課題は、区分線形、あるいは非線形の確率分布モデルに従う生のデータによる同モデルの能力の検証である。このような生データの例としては、臼井らの研究(15)で用いられたマンセル色空間データ等が挙げられよう。第3の課題は、4層PPNによる因子分析を利用したクラスター分析法の詳細な性質を第1の課題の答えと実験的検証の両面から探ることである。

なお、この章の内容は文献(16)で明かにされている。

第5章では、PPNと砂時計モデルの間の数理的関連を詳細に論じると同時に、それらで実現される写像の間の変換式を与えた。また、統計的パターン認識におけるクラス判別の前処理としての特徴選択の手法をPPN、砂時計モデルで置き換えた場合にどうか、についてシミュレーション実験を行った。ここでの課題は、多層ニューラルネットによるカテゴリー分類(非線形判別分析)と、PPN/砂時計モデルによる情報集約(非線形因子分析)の間の幾何学的対応の有無もしくはその性質である。

この章の内容は文献(17)で明かにされている。

これらについて満足すべき成果が得られた時点で、本文が真に完成されたといえよう。その段階で、本文がかつてRosenblattのパーセプトロン(18)に対してMinsky and Papert(19)が提出したのと同様のスピードリミッターを、今度は多層ニューラルネットおよび誤差逆伝搬学習法(20)に対して突きつけることになるのかどうか？またそれはどのような性質のものか？が筆者の最大の関心事である。

参考文献

- (1). Fukunaga, K. : "Introduction to Statistical Pattern Recognition", Academic Press, New York(1972).
- (2). Anderson, T.W. : "An Introduction to Multivariate Statistical Analysis" 2nd Edition, John Wiley & Sons, New York(1984).
- (3). Duda, R.O. and Hart, P.E. : "Pattern Classification and Scene Analysis", Wiley, New York(1973).
- (4). Young, T.Y. and Fu, K.S. : "Handbook of Pattern Recognition and Image Processing", Academic Press, New York(1986).
- (5). Anderberg, M.R. : "Cluster Analysis for Applications", Academic Press, New York(1973).
- (6). 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, pp. 127-134(1988).
- (7). 上坂吉則: "ニューラルネットワークと学習", 平成2年電気・情報関連学会連合大会論文集, S31-2, 5, pp. 79-82 (1990).
- (8). 上坂吉則: "パターン認識と学習の理論", 総合図書(1971).
- (9). 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(D-11), J73-D-11, 8, pp. 1173-1178(1990).
- (10). 大津展之: "パターン認識における特徴抽出に関する数理的研究", 電総研研究報告 第818号(1981).
- (11). Saund, E. : "Dimensionality-Reduction Using Connectionist Networks", IEEE Trans., PAMI, 11, 3, pp. 304-314(1989).
- (12). T. Kohonen, : "Associative Memory", Springer Verlag(1977).
- (13). 福島邦彦 : "神経回路と自己組織化", 共立出版(1979).
- (14). 米倉達広、横井茂樹、鳥脇純一郎 : "多層ニューラルネットワークによる情報集約の手法とその理論的考察", 信学論(D-11), J73-D-11, 8, pp. 1205-1212. (1990).
- (15). 臼井支朗, 中内茂樹, 中野正恵、 : "多層ニューラルネットワーク恒等写像モデルによるマンセル色空間の特徴解析", 信学技報, NC89-40, pp. 45-50(1989).

- (16). 米倉達広、宮崎慎也、鳥脇純一郎：“恒等写像モデルおよびPPNに基づく4層ニューラルネットの情報集約機能の解析”，信学論(D-II)，(採録決定)
- (17). 米倉達広、鳥脇純一郎：“4層恒等写像型ニューラルネットによる疑似区分線形因子分析”，信学論(D-II)，(投稿中)
- (18). Rosenblatt, F：“Principles of Neurodynamics”，Spartan(1961).
- (19). Minsky, M and Papert, S：“Perceptron”，MIT Press(1969).
- (20). Rumelhart, D.E. and McClelland, J.L.：“Parallel Distributed Processing”，Vol.1 and Vol.2, MIT Press(1986).

謝辞

筆者がこの研究を始めたのは、昭和63年4月、それまで7年間勤めていた株式会社山武ハネウエルを辞めて、名古屋大学博士後期課程に再入学するため、妻とともに名古屋入りしてからである。それから3年足らずの間にここまでの研究成果にまとめることができたのは、その後、数多くの方々の励ましや御援助があったからに他ならない。以下、紙面にてそれらの方々に謝意を表したい。

最初に、本研究をすすめるうえで最も貴重な御指導、御鞭達を賜り、この分野での、百編近い数にのぼる資料、文献を御教示頂いた、筆者の修士時代からの恩師である本学鳥脇純一郎教授には深く感謝するものである。

また、大学の役割やそこに生きる研究者の目標などについて、極めて広い見識で御指導頂いた、恩師であると同時に尊敬する大先輩でもある本学横井茂樹助教授には心から感謝すると共に今後益々御発展されることを祈る次第である。

極めて貴重な資料を御教示頂くと共に、厳しさのなかにも心のこもった御鞭達を賜った本学杉江昇教授には、心から感謝するものである。

ニューロに関する技術的情報や学会の情報、あるいはそこで発表する機会を与えて頂いた豊橋技術科学大学臼井支朗教授には深く感謝する次第である。

修士時代からの大先輩である、中京大学長谷川純一教授には、公私ともに、極めて暖かい御指導、御援助を頂いたことを心から感謝すると共に今後益々の御発展を祈る次第である。

また、本学鈴木秀智講師には、研究ならびにその息抜きをするうえで最高の道具をいろいろと教えて頂いた。これは今でも大いに役だっている。

本学安田孝美助手には、昼食時やその他の時間に、研究やその他の話題についての最高に楽しい会話の時間を与えて頂いた。それらは筆者の心の依りどころとなった。

筆者が本研究を学位論文のテーマにしようと決心した、最初の機会を与えて頂い

たのは、筆者の20年来の親友である九州工業大学八木哲也助教授ならびに昭和63年当時その上司であり、いまは筆者の恩師でもある名古屋工業大学舟橋康行教授によってである。御二人の開かれたニューロコンピュータに関するセミナーに参加させて頂いたことが本研究の第1歩となった。ここに心から感謝の意を表す。

また、国内外の貴重で重要な文献や資料を御教示頂いたATR視聴覚研究室曾根原登主任研究員の御援助がなければ本研究成果は有り得なかった。ここに心から感謝する。

筆者と同じ鳥脇研究室でニューラルネットによるパターン認識を目指す、松本哲也氏、宮崎慎也氏、星野俊仁氏には、それぞれ第5章、第4章、第3章の内容に関して多大な御協力ならびに実際のシミュレーション実験による無数のデータを頂いた。彼らには心から謝意を表するとともに今後の御活躍をお祈りする次第である。また、日頃熱心に御討論頂いた鳥脇研究室諸氏に深く感謝の意を表す。

そして、修士時代、独創性の意義とそのセンスを御教示頂いた、10数年来の恩師である、中京大学福村晃夫教授に心から感謝する次第である。

最後に、本研究をすすめることに関して、心から筆者を勇気付けてくれ、また、生活と研究の両面で幾多の試練や喜びをすべて筆者と共に味わい、分かち合ってくれた妻の睦子、それに、昼夜を問わずその泣き声と笑顔で筆者の論文執筆のペースを完全にコントロールしてくれた娘のまりあに対する謝意は言葉に表し難い。

付録図A KL展開、3層砂時計モデル、3層PPNによるアイリスデータの2次元への写像

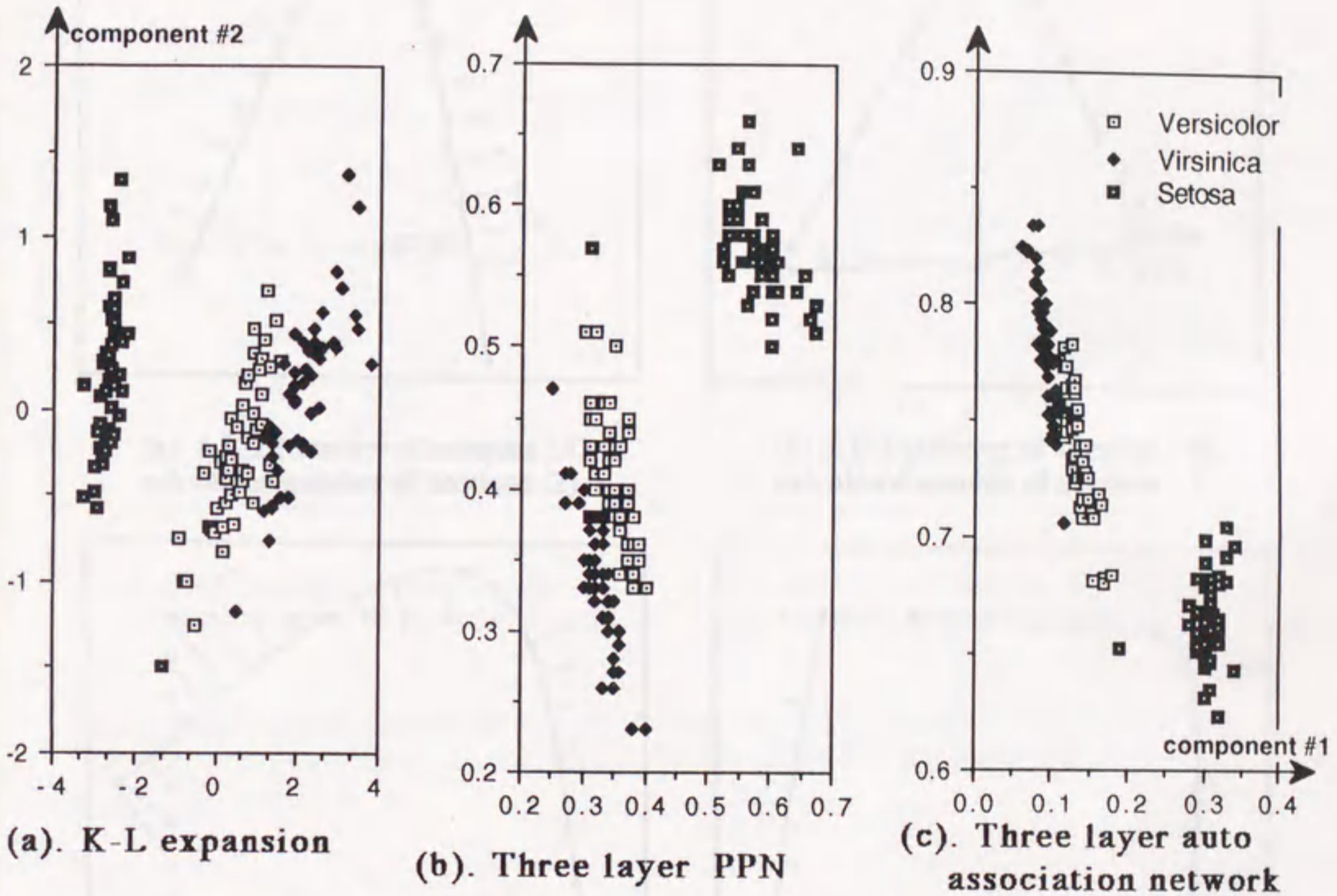
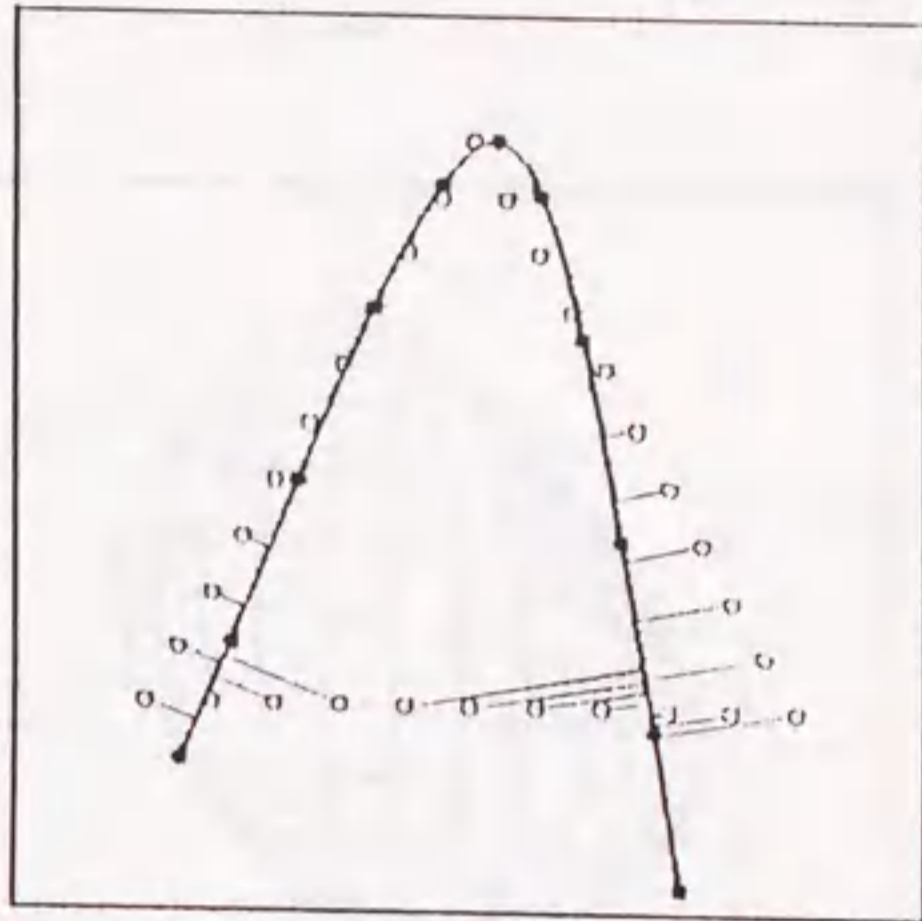
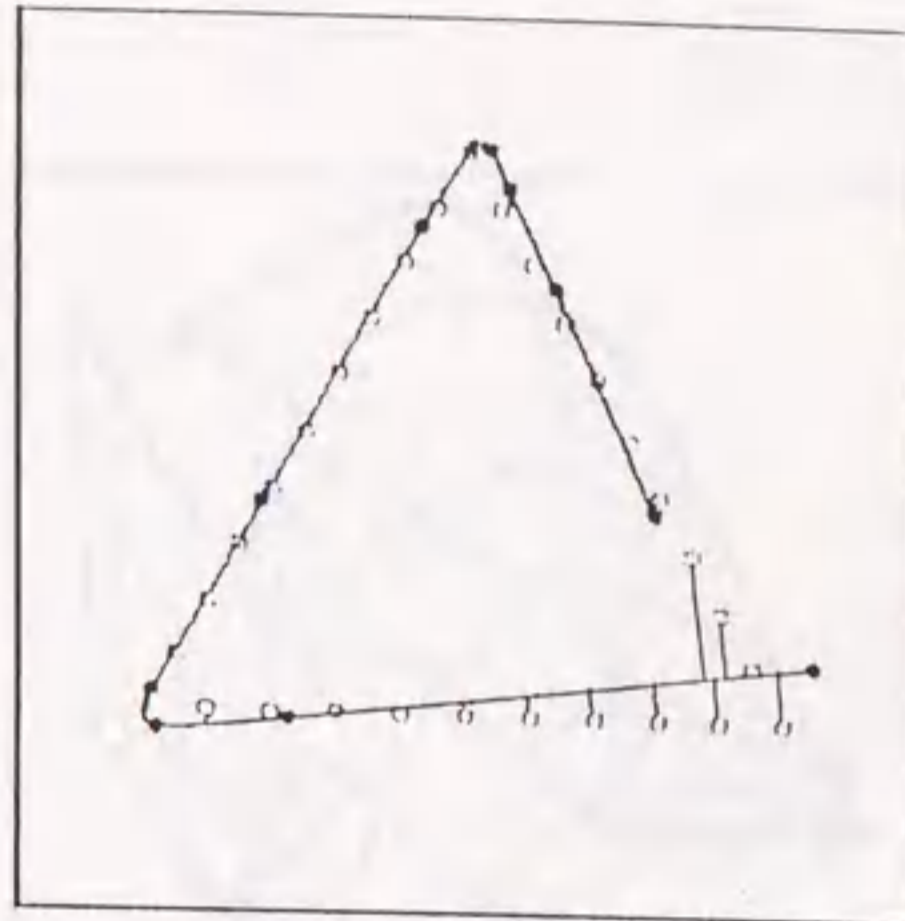


Fig.A Projection of the Iris data onto 2 components.

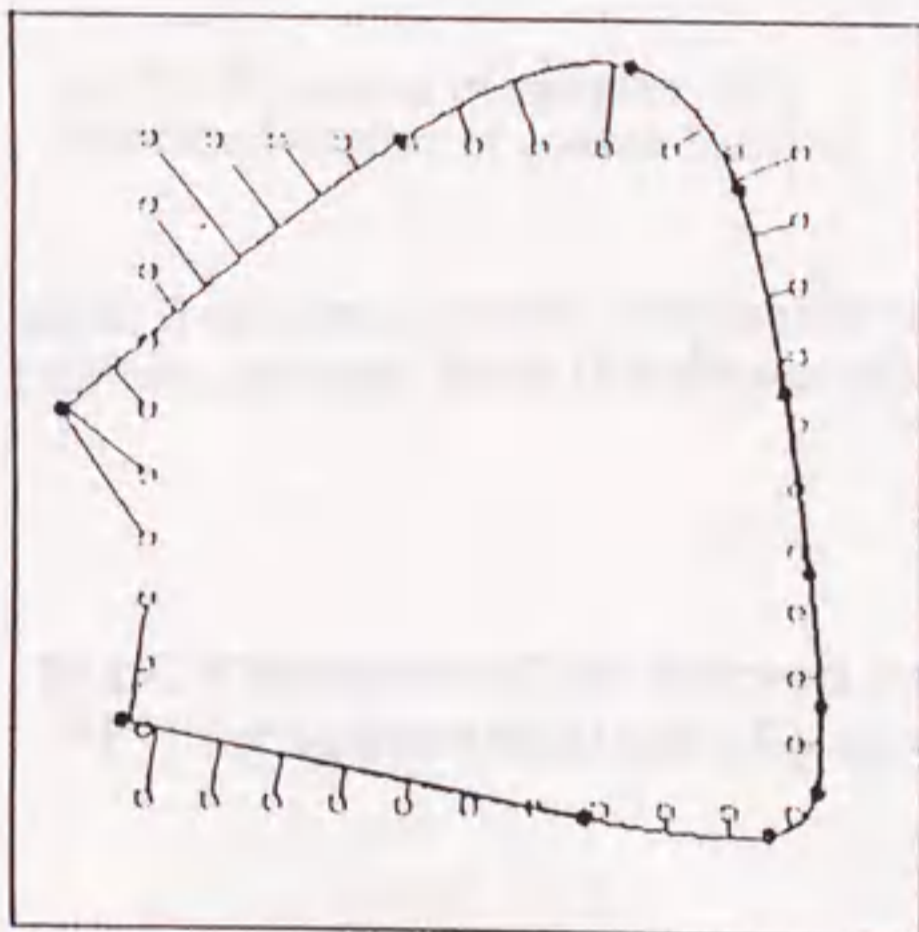
付録図B 多角形上に分布した2次元標本の、4層PPNによる情報集約(近似曲線の獲得例)



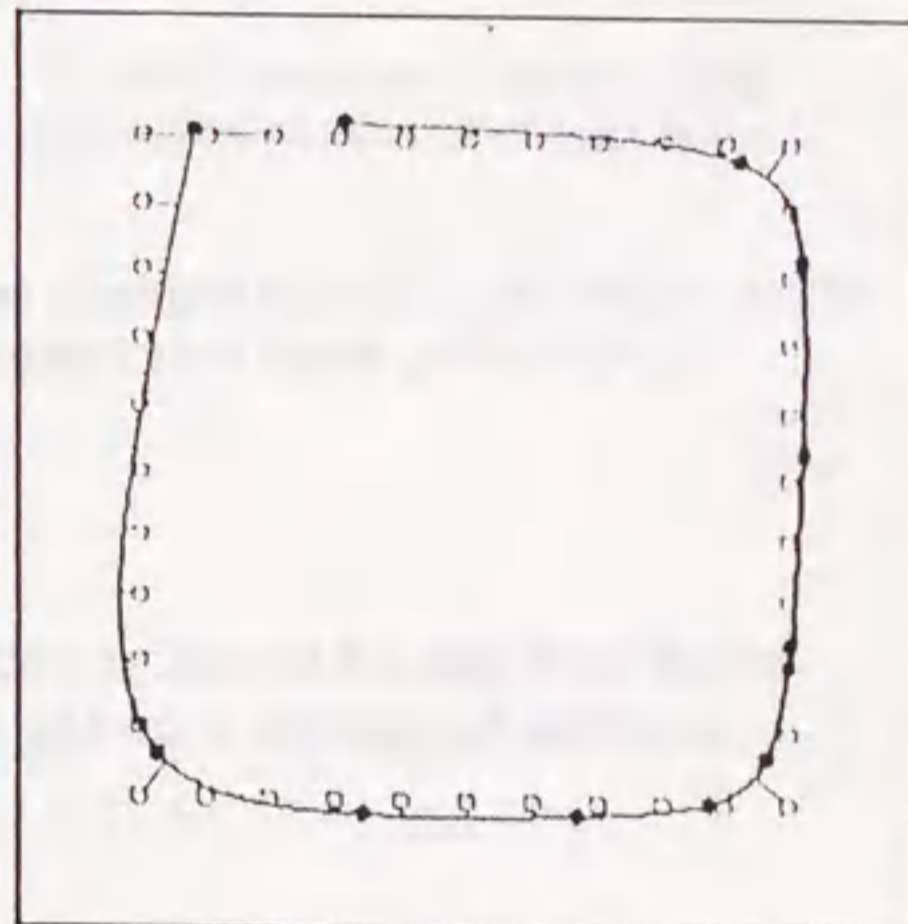
(a) A122 (number of samples :30, calculated number of sections :2)



(b) A132 (number of samples : 30, calculated number of sections : 3)



(c) A122 (number of samples :40, calculated number of sections :3)

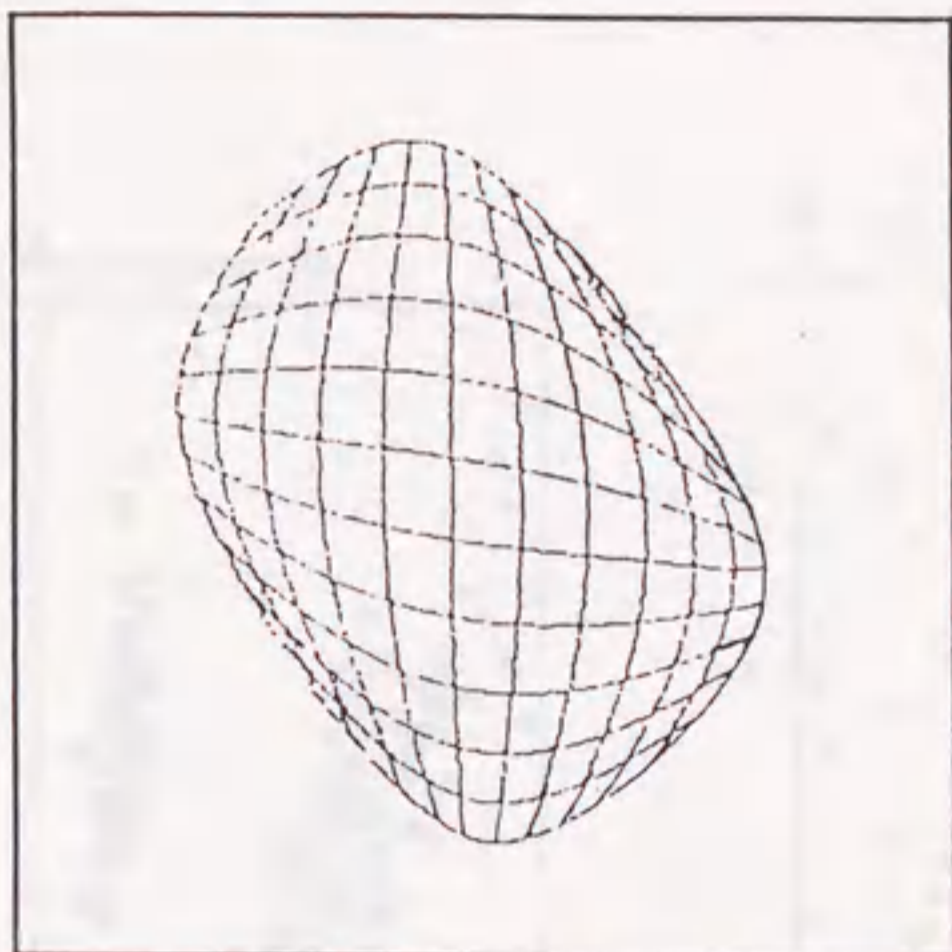


(d) A142 (number of samples :40, calculated number of sections :4)

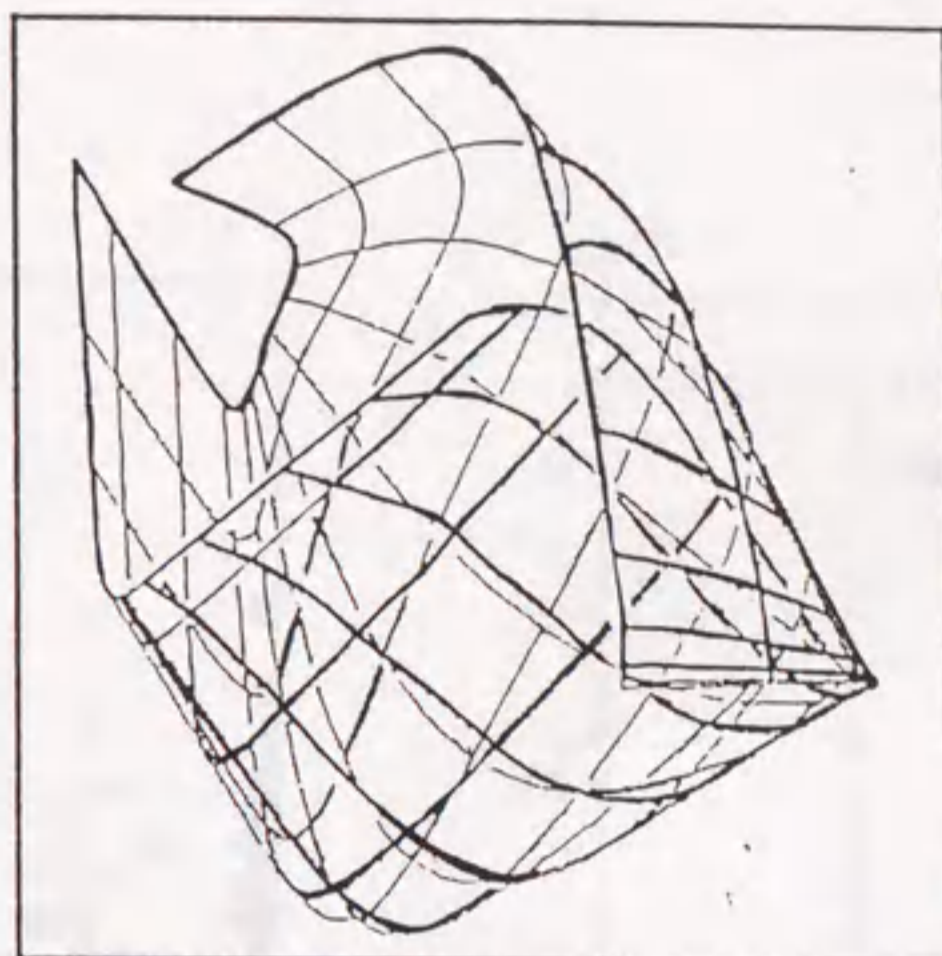
Each (regression) curve represents the set of point (x_1', x_2') , the output layer pattern, mapped from the change of the integrated layer value (y_1, y_2) . Each open dot represents the sample point (x_1, x_2) and the short line are drawn between the sample points and the corresponding approximation points.

Fig.B Examples of the curves spanned by the four layer PPN for approximation of a sample set on polygons (triangle (a),(b) and rectangle (c),(d)).

付録図C 球面上に分布した3次元標本の、4層PPNによる
情報集約(近似曲面の獲得例)



(a) B223 (number of samples :100,
calculated number of convex hulls :1)



(b) B243 (number of samples :100,
calculated number of convex hulls :5)

Each (regression) curve represents the set of point(x_1, x_2, x_3), the output layer pattern, mapped from the change of the integrated layer pattern(y_1, y_2).

Fig.C Examples of the curved surfaces spanned by the four layer PPN for approximation of a sample set on a spherical surface.

付録図D 種々の4層PPNによるアイリスデータの1次元ならびに2次元集約写像

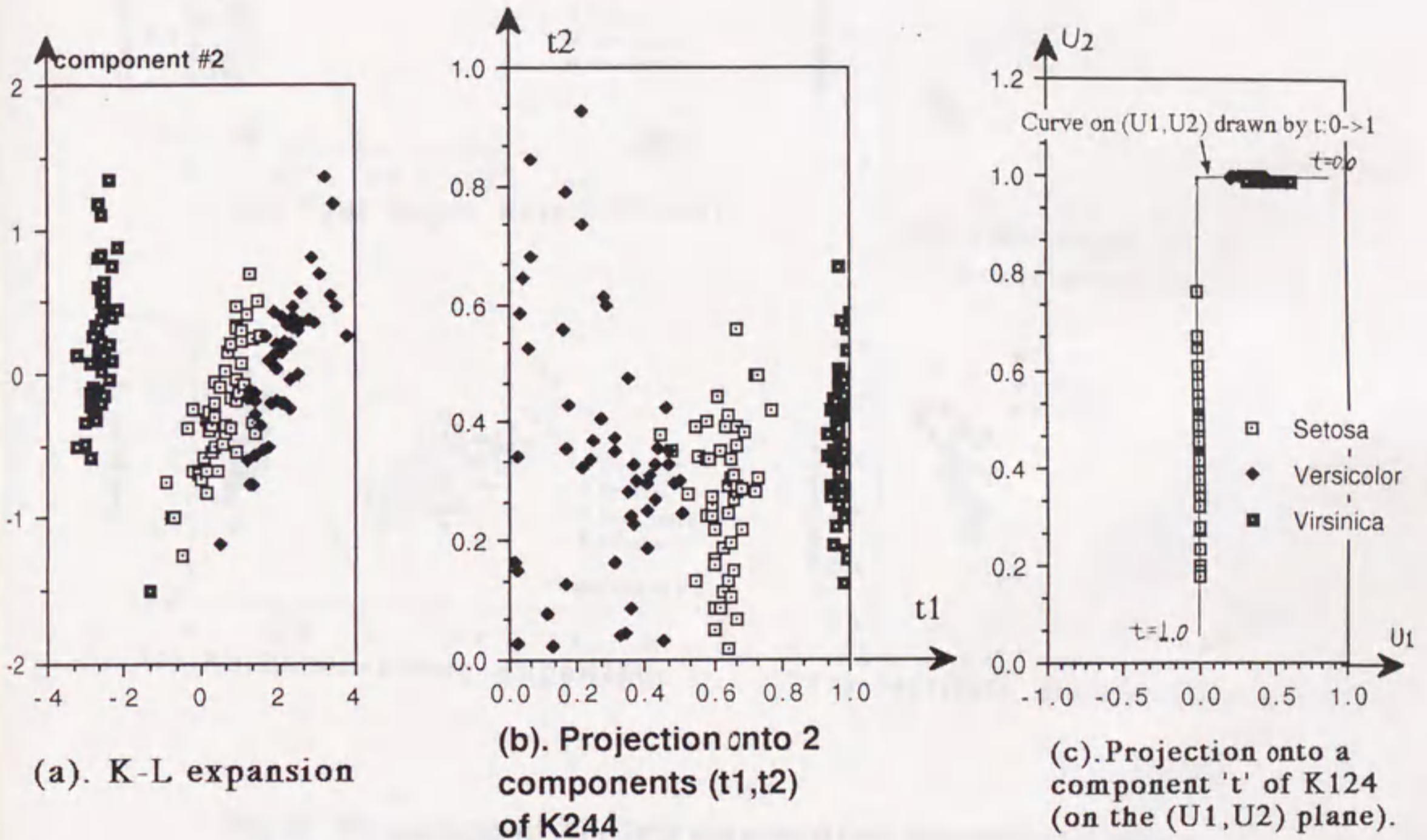


Fig.D Non-linear mapping of the Iris data onto one and two dimensional space by using the four layer PPN.

付録図E 種々の方法によるアイリスデータの2次元集約写像
(4層PPN, 5層砂時計モデル、KL展開ならびに判別分析)

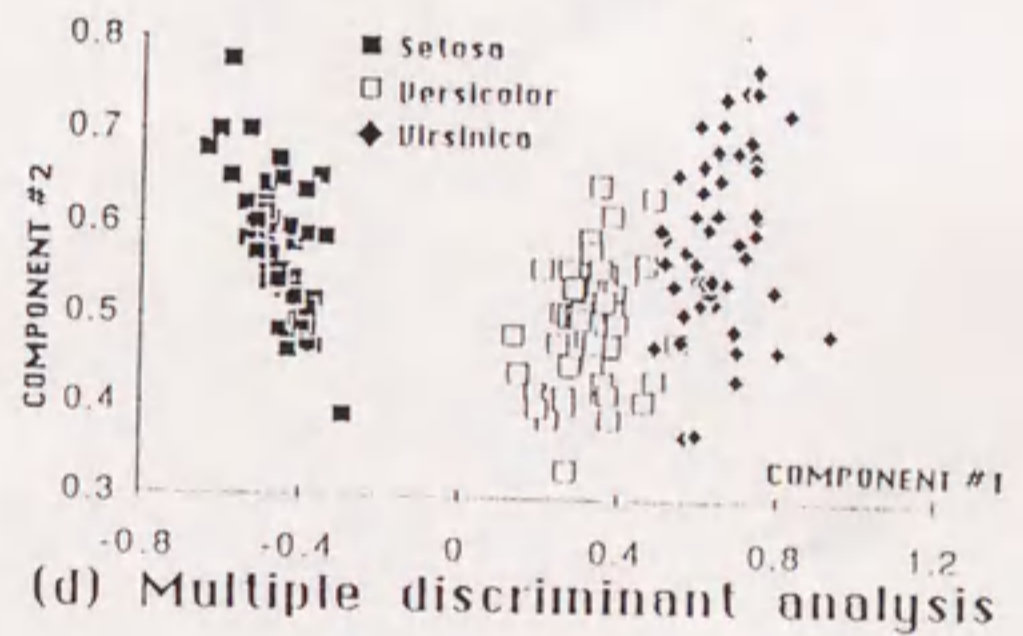
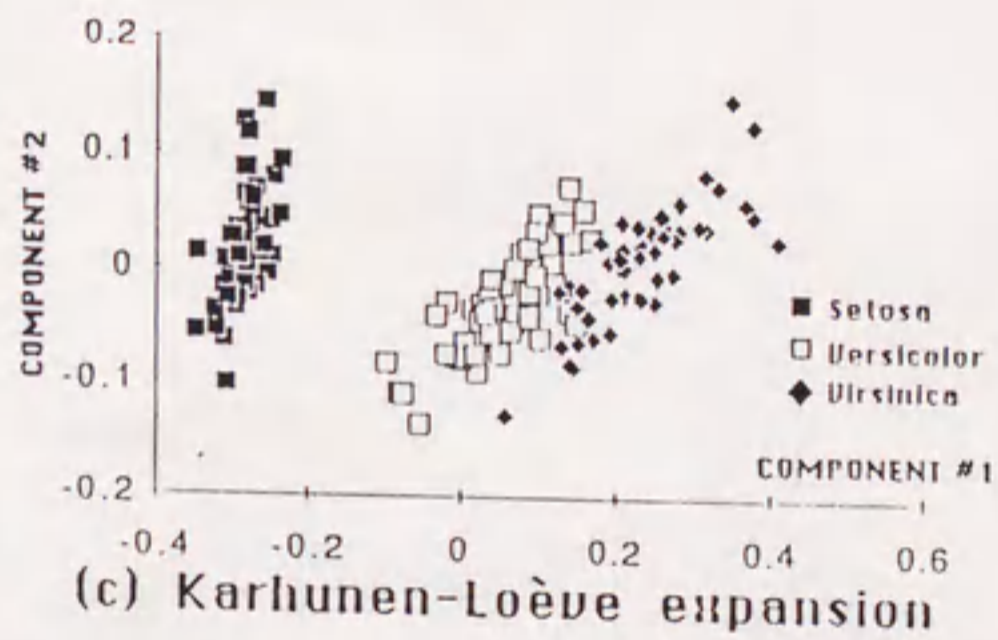
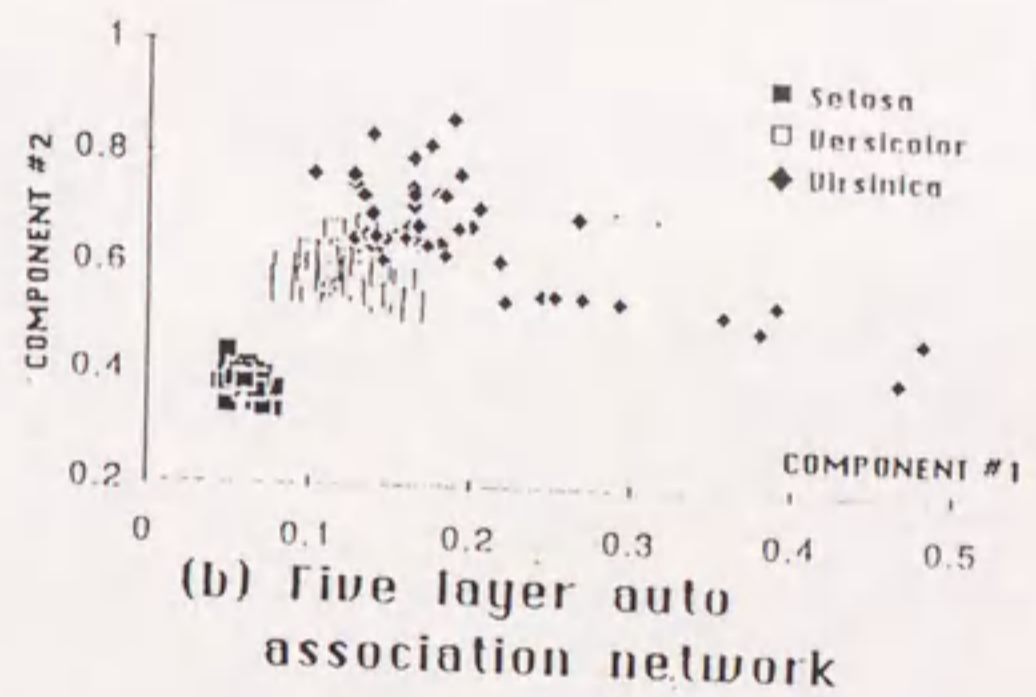
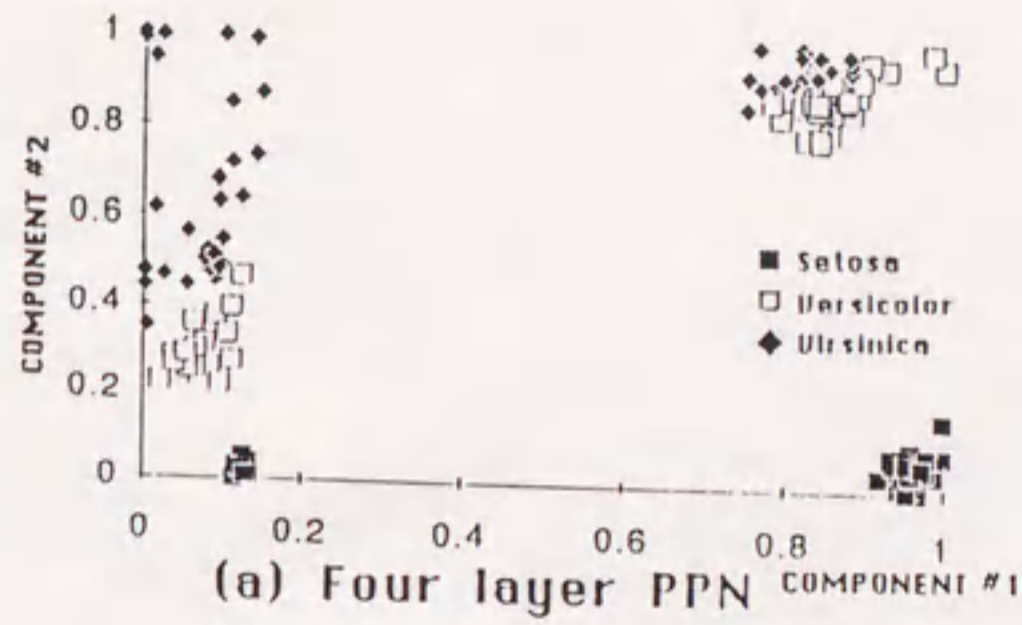
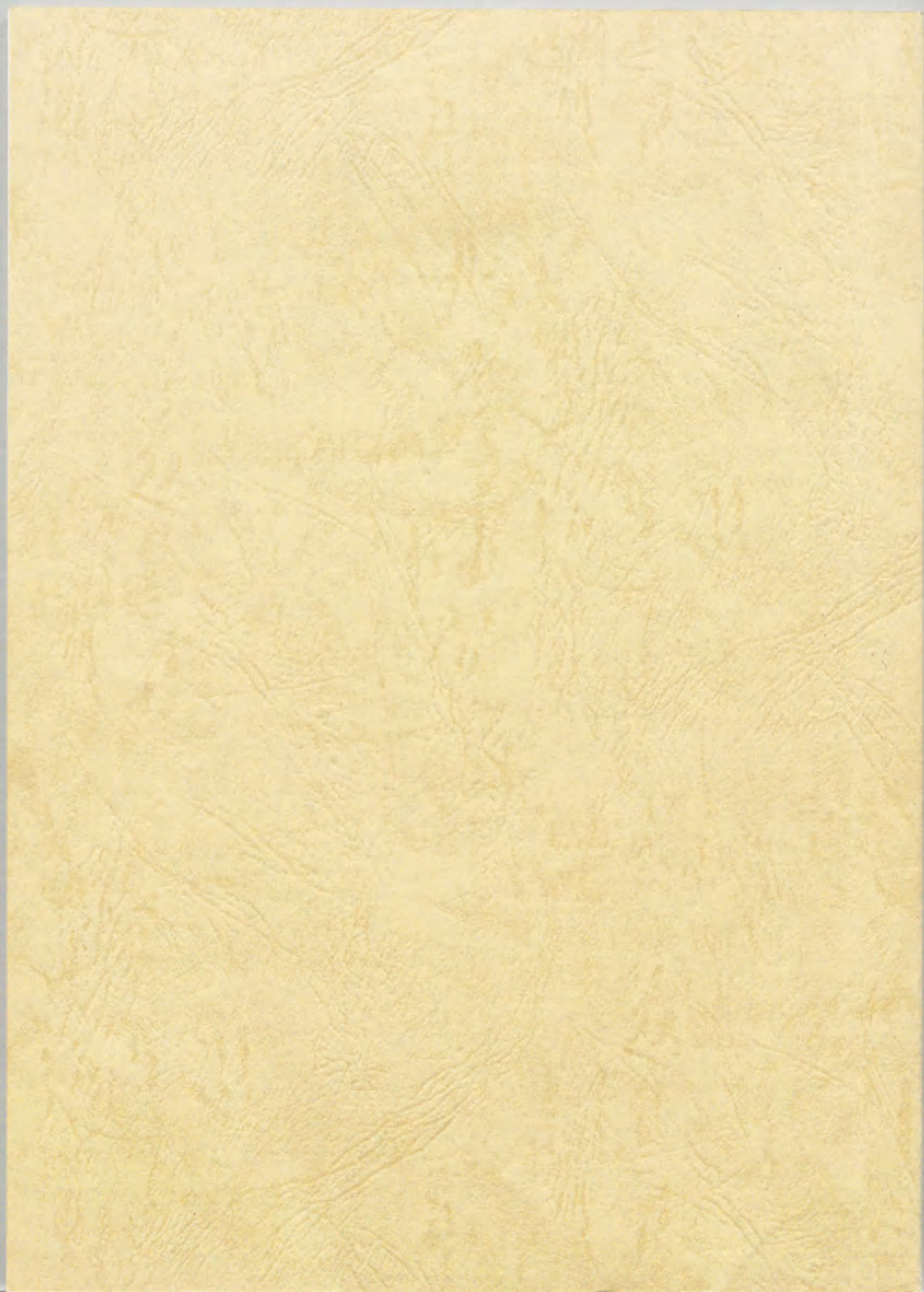


Fig.E Projection of the Iris data onto two dimensional space by using the four layer PPN, five layer auto-assoc.network, KLT, and discriminant analysis.



Inches 1 2 3 4 5 6 7 8
cm 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Kodak Color Control Patches

© Kodak, 2007 TM: Kodak



Kodak Gray Scale



© Kodak, 2007 TM: Kodak

A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19

