主成分分析を行なうニューラルネットワークの学習と自己組織

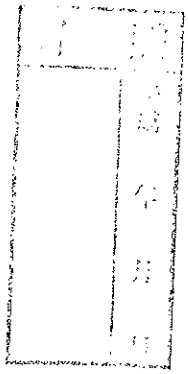
1993年3月

南藤　進光

主成分分析を行なうニューラルネットワークの学習と自己組織

1999年3月

高橋隆史



99012365

目 次

図 目 次

表 目 次

第 1 章

序論

1.1 情報圧縮とニューラルネットワーク

本論文では、ニューラルネットワークによる情報圧縮を研究対象とする。ここでいう情報圧縮とは、複数個の多次元データから成るデータ集合が与えられたときに、その本質的な情報・特徴をできる限り失うことなくデータの数やデータの記述に必要なパラメータ数を削減する操作をさす。パターン認識やデータ圧縮といった情報処理においては、与えられたデータをどれだけ効率の良い形で表現できるかによってその性能が大きく左右されるため、その処理過程の一部としてこのような情報圧縮が重要な役割を担っている。例えばパターン認識の場合、一般に観測データは非常にパラメータ数が多く、またノイズが含まれていることが多いため、そこからクラスの識別に必要な少数の特徴だけをうまく取り出すことが要求される。またデータ圧縮の場合、与えられたデータを、ビット数ができる限り少なく、かつ元のデータを復元したときの歪みもできる限り小さくなるような形式で表現することが必要となる。

データ集合として、T個の N次元実ベクトルの集合

$$\left\{\boldsymbol{x}_t \in \boldsymbol{R}^N\right\}_{t=1}^{T} \tag{1.1}$$

を考える。このとき、代表的な情報圧縮の手法は以下の二種類に分けられる (図 1.1 参照)。

- クラスタリングまたはベクトル量子化

  T個のデータの代表となるような$C(< T)$個のN次元ベクトルをデータ集合から選択する、あるいは新たに生成することによってデータ数を削減する方法

- 次元圧縮

  個々のN次元データをより次元数の少ない$H(< N)$次元空間へ写像することによってパラメータ数を削減する方法

クラスタリング・ベクトル量子化の手法としてはLBG アルゴリズム [1] が、次元圧縮の手法としては主成分分析 [2] を用いる方法が良く知られており、圧縮性能に関する理論、効率的な情報圧縮を実現するためのアルゴリズム、様々な実データに対する応用などが多様に研究されている。その中で本論文では、情報圧縮を行なうニューラルネットワークを学習によって構成する方法に注目する。

ニューラルネットワークを用いる方法の一般的な利点としては、並列計算あるいはハードウェア化に適した情報処理様式であるため、高速な処理を実現しやすいということが上げられる。また、環境などの変化によってデータの特性が徐々に変化するような場合でも、逐次的に学習を行なってネットワークを適応させることができるという特徴も有している。情報圧縮を行なうニューラルネットワークについてはこれまでに数多くの研究がなされており、上記二手法のどちらについても数多くの学習アルゴリズムが提案されている。例えば、クラスタリング・ベクトル量子化は、Kohonen のアルゴリズム [3] に代表される教師なしの競合学習 (文献 [4] 第 9 章) によって実現することができる。また次元圧縮は、Hebb[5] 型の教師なし自己組織学習 (文献 [4] 第 8 章, 文献 [2])、あるいは多層パーセプトロンの誤差逆伝搬学習 [6, 7] によって実現することができる。本論文では特に、次元圧縮を行なうニューラルネットワークを対象とし、その学習理論と応用について検討する。

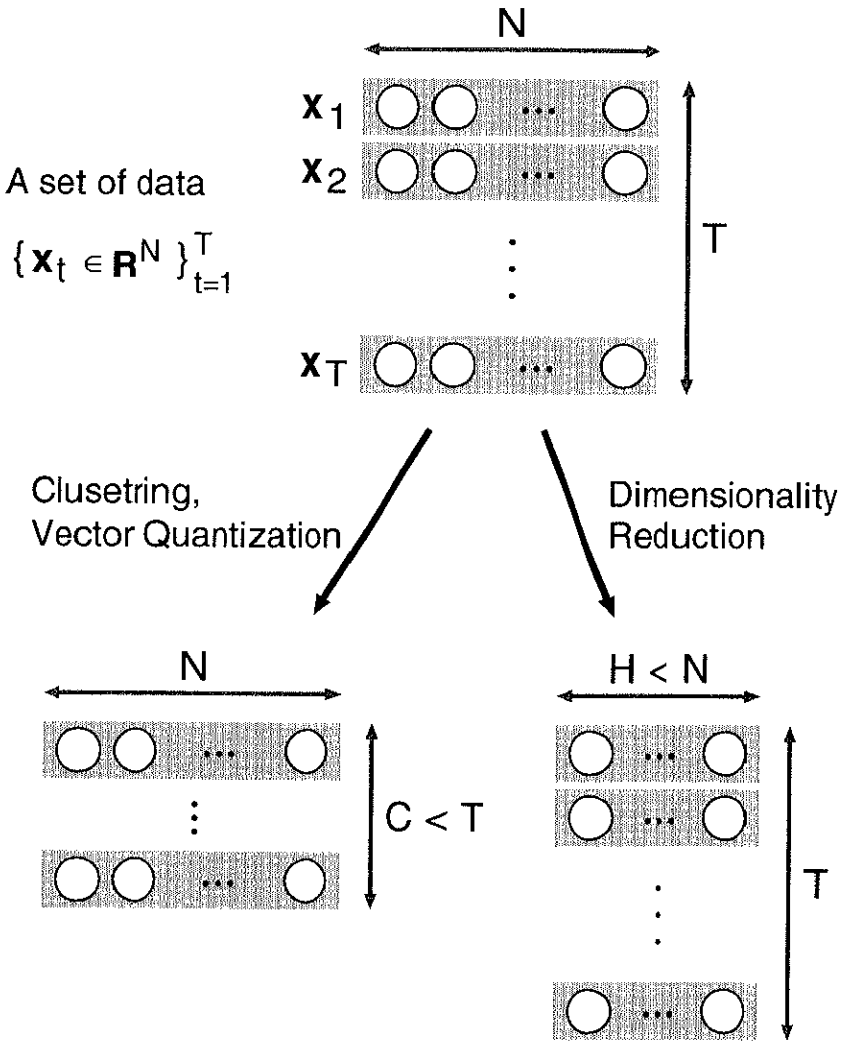


図 1.1: 情報圧縮の二手法

1.2 主成分分析

次章以降でニューラルネットワークによる次元圧縮について検討するための準備として、次元圧縮を実現する方法として一般的な主成分分析 (Karhunen-Loève 変換とも呼ばれる) について概説する。主成分分析の目的は、「与えられた N次元データの性質を良く表す $H(< N)$ 次元の特徴量 (特徴ベクトル) を求めること」である。

N次元の確率変数$\boldsymbol{x} \in \boldsymbol{R}^N$を考える。ただし、$\boldsymbol{x}$ はゼロ平均すなわち$E[\boldsymbol{x}] = \mathbf{o}$と仮定し、共分散行列

$$\begin{aligned} \Sigma &= E\left[(\boldsymbol{x} - E[\boldsymbol{x}])(\boldsymbol{x} - E[\boldsymbol{x}])^T\right] \\ &= E\left[\boldsymbol{x}\boldsymbol{x}^T\right] \quad \in \boldsymbol{R}^{N\times N} \end{aligned} \tag{1.2}$$

が与えられるものとする。通常の主成分分析では、H次元特徴ベクトル$\boldsymbol{y} \in \boldsymbol{R}^H$への写像として、線形写像

$$\boldsymbol{y} = W\boldsymbol{x} \tag{1.3}$$

を考える。ただし、Wの列ベクトルは$\boldsymbol{R}^N$の H次元部分空間の正規直交基底を成すものとする。したがって

$$WW^T = I \tag{1.4}$$

である。このとき、この部分空間への$\boldsymbol{x}$の射影は

$$\tilde{\boldsymbol{x}} = W^T\boldsymbol{y} = W^TW\boldsymbol{x} \tag{1.5}$$

で与えられる。$\tilde{\boldsymbol{x}}$は、特徴ベクトル$\boldsymbol{y}$から$\boldsymbol{x}$を近似的に再構成したものである。主成分分析における特徴量の「良さ」の基準としては、一般に

1. 特徴ベクトル$\boldsymbol{y}$が$\boldsymbol{x}$の分散をどの位良く表しているか

2. 再構成$\tilde{\boldsymbol{x}}$が$\boldsymbol{x}$のどの位良い近似となっているか

の二つが用いられる。1. の場合、その評価基準は

$$J_v = E\left[tr\left(\boldsymbol{y}\boldsymbol{y}^T\right)\right] = E\left[tr\left(\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^T\right)\right] \tag{1.6}$$

$$= tr\left(W\Sigma W^T\right) \tag{1.7}$$

で与えられ、これができる限り大きな値をとることが望まれる。一方、2. の場合は平均二乗誤差

$$J_e = E\left[||\boldsymbol{x} - \tilde{\boldsymbol{x}}||^2\right] \tag{1.8}$$

が基準となり、これができる限り小さな値をとることが望まれる。式 (1.4) の制約のもとでは J_vの最大化と J_eの最小化は等価であるため、主成分分析の目的は、「J_vを最大化する、あるいは J_eを最小化する行列 Wを求めること」と言い表すことができる。この目的を達成する最適な行列 Wは、以下の定理より与えられる。

定理 1.1 主成分分析 [2]

Σの固有値$\lambda_1, \lambda_2, \ldots, \lambda_N$が降順に並んでいるものとし、各固有値に対応する固有ベクトルを$\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_N$とおく。このとき、$WW^T = I$なる制約のもとで J_eを最小にする、あるいは J_vを最大にする行列 Wは

$$W = T[\boldsymbol{e}_1|\boldsymbol{e}_2|\ldots|\boldsymbol{e}_H]^T \tag{1.9}$$

で与えられる。ただし、Tは任意の $H \times H$(正規) 直交行列である。

式 (1.9) で特に $T = I$の場合

$$\boldsymbol{y} = W\boldsymbol{x} = \begin{bmatrix} \boldsymbol{e}_1^T\boldsymbol{x} \\ \vdots \\ \boldsymbol{e}_H^T\boldsymbol{x} \end{bmatrix} \tag{1.10}$$

となり、特徴ベクトル$\boldsymbol{y}$の h 番目の成分$\boldsymbol{e}_h^T\boldsymbol{x}$の分散は対応する固有値$\lambda_h$に一致する $(h = 1, 2, \ldots, H)$。また、J_vの最大値および J_eの最小値はそれぞれ

$$\max J_v = \sum_{h=1}^{H} \lambda_h \tag{1.11}$$

$$\min J_e = \sum_{h=H+1}^{N} \lambda_h \tag{1.12}$$

で与えられる。式 (1.11) および (1.12) からわかるように、特徴ベクトルをこのように定めると、任意の i $(1 \leq i \leq H)$ に対して、上位 i 個の成分 $\boldsymbol{e}_1^T\boldsymbol{x}, \boldsymbol{e}_2^T\boldsymbol{x}, \ldots, \boldsymbol{e}_i^T\boldsymbol{x}$ のみ取り出しても、i 次元で分散最大かつ誤差最小の特徴ベクトルとなる。このようにすると、上位の成分ほどデータの分散を良く表し、再構成時の近似誤差減少に大きく寄与することになる。主成分分析においては特に断りがない限り $T = I$ を仮定し、$\boldsymbol{e}_h^T\boldsymbol{x}$ のことを $\boldsymbol{x}$ の第 h 主成分と呼ぶ。このように、主成分分析では成分が重要度の順に並んだ特徴ベクトルを得ることができる。

1.3 本論文の構成

本論文では、ニューラルネットワークの学習アルゴリズムとして以下の二種類を考慮し、これらの学習理論とその応用について検討する。

- Hebb 型の教師なし (自己組織) 学習アルゴリズム
- 多層パーセプトロンの教師付き (誤差逆伝搬) 学習アルゴリズム

第 2 章では、学習によって主成分分析を実現する Hebb 型学習アルゴリズムの応用として、視覚系神経回路網の自己組織モデルを構築する。ニューロンの時間応答まで考慮したネットワークモデルによって、視覚系神経細胞の時空間応答特性を再現できることを示す。

第 3 章では、多層パーセプトロンの学習による次元圧縮を対象とする。従来パーセプトロンの学習に用いられてきたエネルギー関数には、主成分分析のように寄与順に並んだ特徴成分が得られないという問題があることを指摘し、寄与順の成分抽出を可能とする新たなエネルギー関数を検討する。線形素子から成る 3 層パーセプトロンを用いる場合には、中間層素子の出力が主成分に一致することを証明する。また、非線形素子を含む多層パーセプトロンによる非線形主成分分析についても検討する。

第 4 章では、第 3 章で検討したエネルギー関数を一般の学習課題に適用し、その効果を検討する。従来のエネルギー関数による学習と異なり、中間層素子の出力が寄与順に並んだ冗長性の低い内部表現が得られることを示す。

最後に第 5 章で本研究の論点をまとめ、結論を述べる。

第 2 章

視覚系時空間受容野の自己組織形成モデル

2.1 導入

生体視覚系の神経回路網は、外界からの光信号の符号化および特徴抽出を行なっている。その構成要素である個々の神経細胞は、視野内の光刺激の位置や形状などの様々な条件に選択的な応答を示すことが知られており、このような応答特性の分布は受容野と呼ばれている。これまでに、生理学実験によって様々な視覚系神経細胞の受容野特性が明らかにされており [8, 9, 10]、その機能の計算論的説明や応答の再現を目的としたモデルも多様に検討されている [11, 12, 13, 14, 15]。特に、大脳皮質視覚野に存在する神経細胞の受容野特性が生後間もない頃の視覚環境に影響を受けて変化する [16] ことが明らかになってからは、教師なし学習を行なうニューラルネットワークによる受容野の自己組織モデルが数多く提案されている [17, 18, 19, 20, 21, 22, 23]。例えばMalsburg は、上記の知見を説明するため、線分刺激を入力として方位選択性の受容野を自己組織するネットワークモデルを提案している [17]。また Linsker は、視覚系には生まれる前にすでに構造をもった受容野が存在するという知見を説明するため、ランダムノイズを入力とする自己組織モデルを提案している [18, 19]。この

モデルは線形素子から成る階層型のニューラルネットワークであり、各層のニューロン間の重みに段階的に Hebb 型の教師なし学習を適用することによって、視覚系に存在するものと同様の中心–周辺型や方位選択型などの受容野構造が自己組織的に獲得される。Linsker は、このようなネットワークの教師なし学習と主成分分析との間に密接な関係があること、および適当な条件下では主成分分析によって線形ネットワークの入力と出力の相互情報量最大化が実現されることを指摘し、「知覚系神経回路網は与えられた入力信号に含まれる情報をできる限り失わないように変換・符号化することを目的としている」という情報量最大保持原理を提唱している [19]。一方 Sanger は、入力層と出力層の 2 層のみから成る単純な線形ネットワークの教師なし学習で主成分分析を実現する学習アルゴリズムとして、GHA(Generalized Hebbian Algorithm) を提案している [22]。このアルゴリズムは入力の主成分を分散の大きさの順に抽出し、Linsker のモデルと同様のランダムノイズを入力とした学習によって中心–周辺型や方位選択型の受容野を自己組織することができる [23]。これらの他にも数多くの受容野自己組織モデルが提案されているが、これまでのモデルはいずれも受容野の空間特性の再現を目的としており、視覚刺激の時間変化に対する神経細胞の応答特性についてはほとんど考慮されてこなかった。そのため、例えば網膜の出力細胞である X 型および Y 型の神経節細胞はいずれも中心–周辺型の空間特性を有するものの、X 型は刺激に対して持続的な応答を示し、Y 型は過渡的な応答を示すというような生理学的知見 (文献 [14] 1.4.2 節) を説明できない。そこで本章では、時間応答特性まで含めた時空間受容野の自己組織形成モデルを検討する。

はじめに 2.2節で時間特性を考慮にいれたニューロンモデルを検討し、それを構成要素とするネットワークモデルの構造について説明する。また、自己組織の計算機シミュレーションに用いる学習アルゴリズムについて述べる。2.3節では計算機シミュレーションの結果を示し、ネットワークが獲得した受容野の時空間特性を調べる。2.4節ではこのモデルが獲得する受容野の特性を理論的に解析する。最後に 2.5節で本研究のモデルの妥当性について議論する。

2.2 ネットワークモデルの構成

時空間受容野の自己組織を実現するためのネットワークモデルについて述べる。はじめにネットワークの構造について説明し、次に受容野の自己組織をシミュレートするための教師なし学習アルゴリズムについて述べる。

2.2.1 ニューロンモデルとその時間応答特性

応答の時間変化を考慮に入れて受容野のモデルを構成するためには、ニューロンの時間応答をモデル化する必要がある。最も単純なモデルとしては、単位時間遅れ素子を多段接続したフィルタを介してニューロンが入力を受け取る形式[24]を考えることができる。しかし、このようなモデルでは一つの入力信号が時間遅れ幅の異なる多数のシナプスを介して伝達されるという仮定をおくことになり、生理学的知見と照らし合わてもニューロンのモデルとして不自然である。そこで本研究では、

- 固有の時間インパルス応答を有する少数の信号伝達経路が存在する
- 各伝達経路からの入力信号に対する重みの値によってニューロンの時間応答特性が決定される
- ニューロンの入出力特性は線形関数で表せる

という仮定をおいてモデル化を行う。この仮定に基づくニューロンモデルを図 2.1に示す。このモデルでは、一つの入力信号$\xi(t)$ は、それぞれ$\phi_i(t)$ なるインパルス応答をもつ N_{ch}個の伝達経路を経てニューロンへ入力される。したがって、実際のニューロンへの入力信号$\xi_i^{\mathcal{T}}(t)$ は

$$\xi_i^{\mathcal{T}}(t) = (\phi_i * \xi)(t) = \int_{-\infty}^{t} \phi_i(t-t')\xi(t')dt' \quad (i = 1, 2, \ldots, N_{ch}) \tag{2.1}$$

と表される。ニューロンの入出力は線形とするため、図 2.1のように$\xi(t)$ のみを入力として受け取っている場合、その出力は

$$y(t) = \sum_{i=1}^{N_{ch}} \omega_i \xi_i^{\mathcal{T}}(t) \tag{2.2}$$

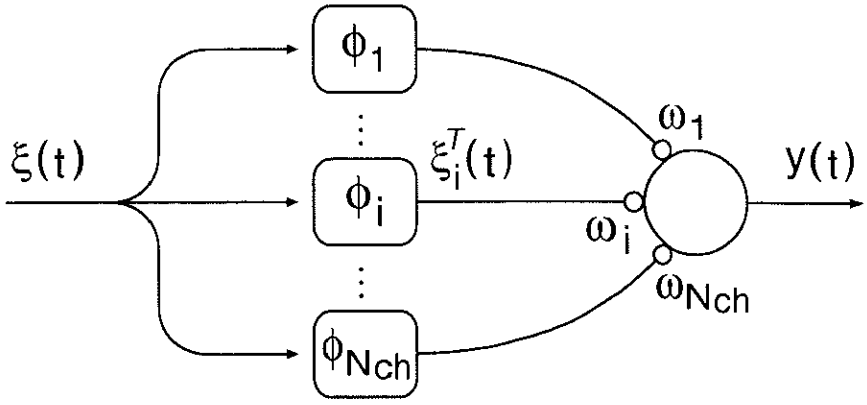


図 2.1: 複数の信号伝達経路を有するニューロンモデル

で与えられる。ただし、ω_i は i 番目の伝達経路からの入力に対する重みを表す。

本章のネットワークモデルでは、伝達経路のインパルス応答は線形再帰フィルタ [25, 26] により次式のように与えるものとした。

$$\phi_i(t) = \begin{cases} C_i \left(\frac{t}{\tau_i}\right)^{\alpha} e^{-\alpha\frac{t}{\tau_i}} & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{2.3}$$

ただし、αとτ_iは応答の持続時間と時間遅れの大きさを制御するパラメータである。C_iはインパルス応答を正規化して

$$\int_{-\infty}^{\infty} \phi_i(t)\, dt = 1 \tag{2.4}$$

を満たすための定数であり、

$$C_i = \frac{\alpha^{\alpha}}{\tau_i(\alpha-1)!} \tag{2.5}$$

で与えられる。$\{\phi_i(t)\}_{i=1}^{3}$の例を図 2.2に表す。ただし、$\alpha = 7$、$\tau_i = i$とした。この例が示すように、αの値が同一の場合、τ_iの値が大きくなるほど時間遅れが大きくなる。

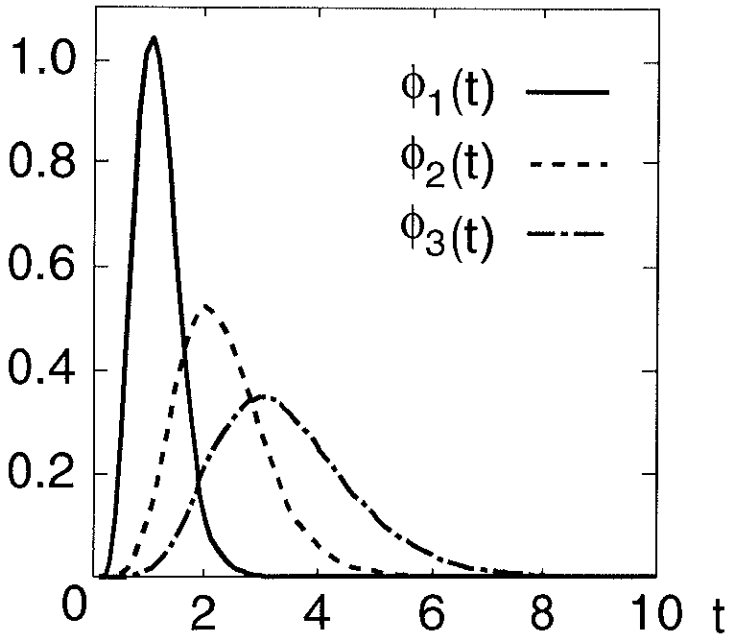


図 2.2: 伝達経路のインパルス応答関数

2.2.2 ネットワークの構造

上述のニューロンモデルを要素とするネットワークモデル全体の構造を図 2.3に示す。二次元平面上に配列された光受容器からの信号が N_{ch}個の経路を伝達されてネットワークへの入力を構成する。p 個の光受容器の出力信号を

$$\boldsymbol{\xi}^{\mathcal{S}}(t) = \left[\xi_1^{\mathcal{S}}(t), \xi_2^{\mathcal{S}}(t), \ldots, \xi_p^{\mathcal{S}}(t)\right]^T \in \boldsymbol{R}^p \tag{2.6}$$

とおき、$i(=1,2,\ldots,N_{ch})$ 番目の伝達経路のインパルス応答をϕ_iとおくと、i 番目の経路を伝達されたニューロンへの入力$\boldsymbol{\xi}_i^{\mathcal{ST}}(t)$ は次式で表される。

$$\boldsymbol{\xi}_i^{\mathcal{ST}}(t) = \begin{bmatrix} (\phi_i * \xi_1^{\mathcal{S}})(t) \\ (\phi_i * \xi_2^{\mathcal{S}})(t) \\ \vdots \\ (\phi_i * \xi_p^{\mathcal{S}})(t) \end{bmatrix} \in \boldsymbol{R}^p \tag{2.7}$$

また、ネットワークへの入力信号全体が成すベクトルは

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{\xi}_1^{\mathcal{ST}} \\ \boldsymbol{\xi}_2^{\mathcal{ST}} \\ \vdots \\ \boldsymbol{\xi}_{N_{ch}}^{\mathcal{ST}} \end{bmatrix} \in \boldsymbol{R}^N \tag{2.8}$$

と表される。ただし、$N = p \times N_{ch}$である。表記を簡単にするため、以降では時刻を表す変数 t は適宜省略する。ネットワークは $M(< N)$ 個のニューロンから成り、各ニューロンの入出力関係は線形であるから、式 (2.8) の入力に対するこのネットワークの出力は次式で表される。

$$\boldsymbol{y} = W^T \boldsymbol{x} \tag{2.9}$$

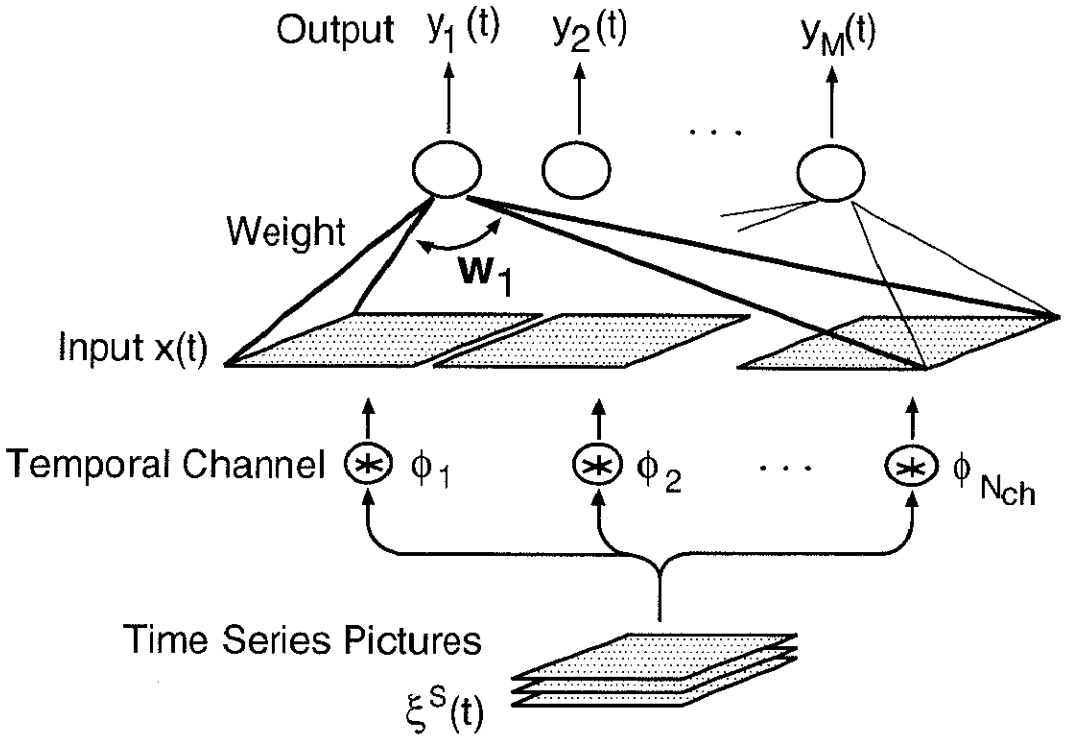


図 2.3: 時空間受容野自己組織のネットワークモデル

ただし、

$$
\begin{aligned}
W &= [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_M] && : \text{重み行列} \\
\boldsymbol{w}_j &= [w_{1j}, w_{2j}, \ldots, w_{Nj}]^T && : j\text{番目のニューロンの重みベクトル} \\
\boldsymbol{x} &= [x_1, x_2, \ldots, x_N]^T && : \text{入力ベクトル} \\
\boldsymbol{y} &= [y_1, y_2, \ldots, y_M]^T && : \text{出力ベクトル}
\end{aligned}
$$

である。このモデルでは、時空間受容野の特性は伝達経路のインパルス応答 $\{\phi_i\}$ とネットワークの重み Wによって決定される。インパルス応答については固定とし、重みに教師なし学習を適用することで時空間受容野の自己組織をシミュレートする。

2.2.3 受容野自己組織のための教師なし学習アルゴリズム

上述のネットワークモデルの重みに教師なし学習アルゴリズムを適用し、受容野の自己組織過程をシミュレートする。重みの学習においては、Linsker[19] の情報量最大保持原理を考慮し、線形ネットワークの教師なし学習によって主成分分析を実現する重みを獲得することのできる Sanger の学習アルゴリズム GHA(Generalized Hebbian Algorithm)[22] を用いる。式 (2.9) のネットワークにおいて、GHA は次式で与えられる。

$$\Delta W = \gamma \left(\boldsymbol{x}\boldsymbol{y}^T - W\mathcal{LT}\left[\boldsymbol{y}\boldsymbol{y}^T\right] \right) \tag{2.10}$$

成分毎に書き下すと

$$\Delta w_{ij} = \gamma y_j \left(x_i - \sum_{k=1}^{j} w_{ik} y_k \right) \tag{2.11}$$

と表される。ただし、γは学習速度を制御するパラメータであり、$\mathcal{LT}[A]$ は行列 A の下三角成分のみを取り出す、すなわち上三角および対角成分を 0 にする演算を表す。このアルゴリズムを用いると、j番目のニューロンの重みベクトル$\boldsymbol{w}_j$は入力の相関行列の j番目に大きな固有値に対応した固有ベクトルに収束する。特に入力がゼロ平均の場合には相関行列は共分散行列に一致し、j番目のニューロンは学習の進行につれて入力の第 j主成分を出力するようになる。したがって、学習収束時には、各ニューロンは入力信号の重要な特徴を順に出力することになる。

2.3 計算機シミュレーション

2.2節に述べたネットワークモデルを用いて受容野自己組織の計算機シミュレーションを行う。

2.3.1 実験条件

入力の画素数すなわち光受容器数、出力ニューロン数、および伝達経路数はそれぞれ $p = 9 \times 9 = 81$, $M = 16$, $N_{ch} = 3$ とする。したがって、ネットワークへの入力 $\boldsymbol{x}$ は $N = 243$ 次元、出力 $\boldsymbol{y}$ は $M = 16$ 次元となる。また、式 (2.3) のインパルス応答関数の性質を決めるパラメータは、それぞれ $\alpha = 7$, $\tau_i = i$ とする。ネットワークの動作は単位時間毎に離散化してシミュレートし、インパルス応答 $\phi_i(t)$ も単位時間毎に標本化したものを用いる。

伝達経路への入力 $\boldsymbol{\xi}^S(t)$ は二次元画像の時系列である。その画素値は、ガウシアンノイズを二次元ガウシアンでフィルタリングし、窓関数をかけることで生成する。ただし、ほとんどすべての値が正となるように、ガウシアンノイズには正のバイアスを加えている。ガウシアンフィルタの標準偏差は 1.5 画素とし、窓関数は画像平面の中心にピークをもつ標準偏差 2 画素のガウシアンとする。以上の操作によって空間相関を与えられたノイズ信号 $\boldsymbol{\xi}^S(t)$ は各伝達経路に入力され、それぞれのインパルス応答 $\phi_i(t)$ で畳み込まれる。したがって、各経路の出力 $\boldsymbol{\xi}_i^{ST}(t)$ およびこれより構成されるネットワークへの入力 $\boldsymbol{x}(t)$ は、空間的・時間的に相関をもった信号となる。

以上の条件でネットワークの重み $\boldsymbol{W}$ に式 (2.10) の学習則を適用し、計算機シミュレーションを実行する。

2.3.2 自己組織された受容野の特性

計算機シミュレーションで得られた時空間受容野を図 2.4に示す。ネットワークは様々なタイプの受容野を獲得しているが、その特性は空間的な重み分布を基に以下

の三種類に分類することができる。以降、No.jは j番目のニューロンを表し、図中の番号に対応する。

1. 空間的に一様な重み分布を有する受容野 (No.1 および 2)

   No.1 は全ての伝達経路からの入力に対して同符号の重みをもち、時空間のローパスフィルタの役割を果たす。一方 No.2 はϕ_1とそれ以外で符号を反転させており、空間平均の時間変化に対して応答を示すニューロンとなっている。

2. 方位選択性の受容野 (No.3–6,9 および 10, 12–16)

   No.3 の受容野は重みの符号の異なる 2 つの領域に分かれており、エッジ検出の働きをすると考えられる。No.4 も同様であるが、その方位選択性は No.3 のものと直交している。No.9 および No.10 はバー検出型細胞に似た空間形状を有するが、文献 [27] に述べられているように受容野は 4 つの部分に分割されている (2.4節も参照)。No.16 はチェッカーボードのような細かな画像に選択的に反応する。これら 5 つのニューロンは全ての伝達経路に対して同符号の重みをもち、それぞれの選択性に適合した刺激図形が静止している時に強い応答を示す。

   一方、 No.5, 6 および 12 から 15 までのニューロンは空間的には上記のものと同様の重み分布を成しているが、重みの符号は No.2 の様に反転している。これらのニューロンは、適当な方位に運動する刺激図形に対して選択的に応答すると考えられる。

3. 中心–周辺型の受容野 (No.7,8 および 11)

   No.7 および 8 は共に同心円形の重み分布を成しており、入力刺激の空間コントラストを検出する機能をもつ。一方、No.11 は空間コントラストの時間変化を検出すると考えられる。

ここで、No.7 および 8 の受容野はそれ以外のものと比べて時空間特性が大きく異なっていることに注意する。他のニューロンの受容野では、3 つの伝達経路に対す

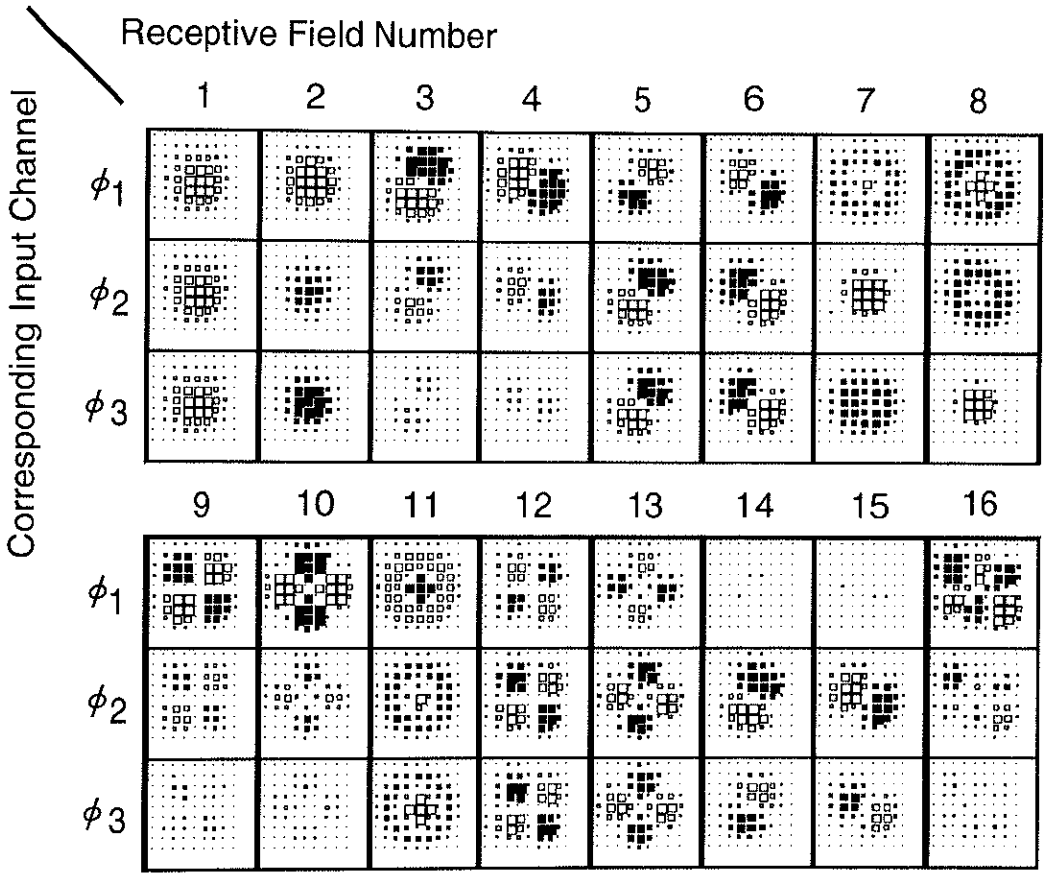


図 2.4: 計算機シミュレーションによって得られた時空間受容野：縦に並んだ三つの正方形パネルの組が一つのニューロンの重み分布を表し、各正方形パネル内の小さな正方形が重みの値を表す。正方形の大きさは重み値の絶対値の大きさを表し、白が正の値、黒が負の値に対応する。

る重みがいずれも同一形状の空間分布を成しており、その受容野内のどの位置にインパルス刺激を提示しても、得られる出力は定数倍の自由度を許して同じ時間経過をたどることになる。しかし、No.7 および 8 の場合、重みの空間分布は伝達経路毎に異なっており、インパルス刺激に対する応答は提示位置によって変化すると予想される。実験の結果、このような不均一な時間応答特性を示す受容野は、入力の平均をゼロとした場合には得られず、バイアスされた入力を与えた時にのみ出現することがわかった。この現象について、2.4節で詳しく検討する。

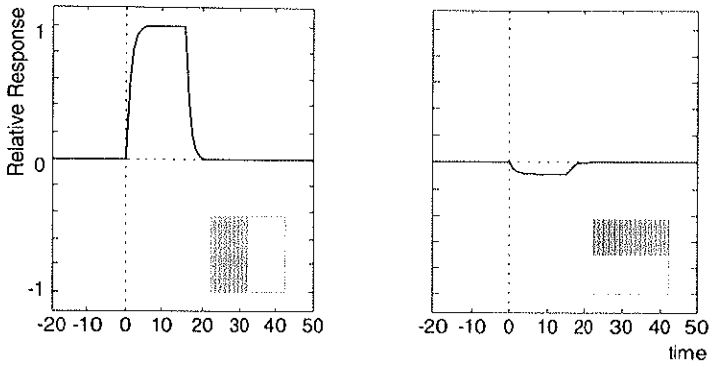
2.3.3 ニューロンの応答特性

次に、得られたニューロンのいくつかに対して静止あるいは運動刺激を入力として与えたときの応答を求め、これらの特性について述べる。図 2.5に No.3 および 5 の出力応答を、図 2.6に No.7 および 11 の出力応答を示す。

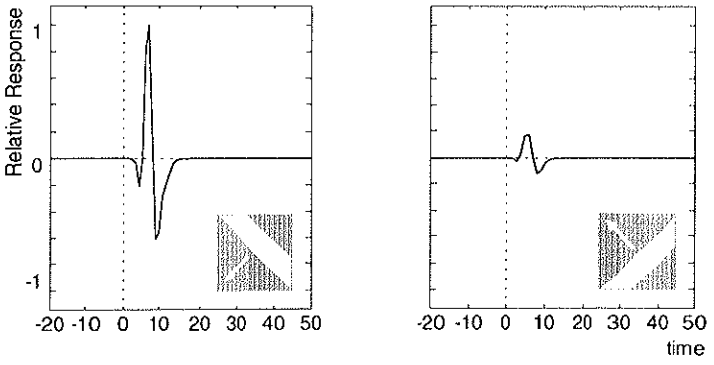
図 2.5(a) は、水平および垂直の静止エッジ刺激を提示したときの No.3 ニューロンの応答を示す。刺激は $t = 0$ から $t = 14$ まで提示した。このニューロンは水平エッジに対して最大の応答を示すが、垂直エッジに対しては弱い応答しか示さない。したがって、No.3 は水平エッジに選択的なエッジ検出器となっている。ただし、本章で検討しているモデルではニューロンの入出力は線形としているため、このニューロンは明暗が反転したエッジに対して正負反転した出力を示すことに注意する。

No.5 ニューロンは、前節に述べたように刺激の時間変化や運動を検出すると考えられる。これを確認するため、運動バー刺激に対する応答を計算した。結果を図 2.5(b) に示す。バーの幅は 3 画素とし、速さは単位時間あたり 2 画素とした。このニューロンは右上から左下へバー刺激が横切るときに最大の応答を示すが、これと直交する右下から左上への運動に対しては弱い応答しか示さない。また、ユニットの線形性から、運動方向が逆転すると応答は正負反転する。したがって、No.5 は方位選択性の運動検出器となっている。

No.7 および 11 のニューロンは、中心–周辺型の受容野を成している。No.7 ニュー

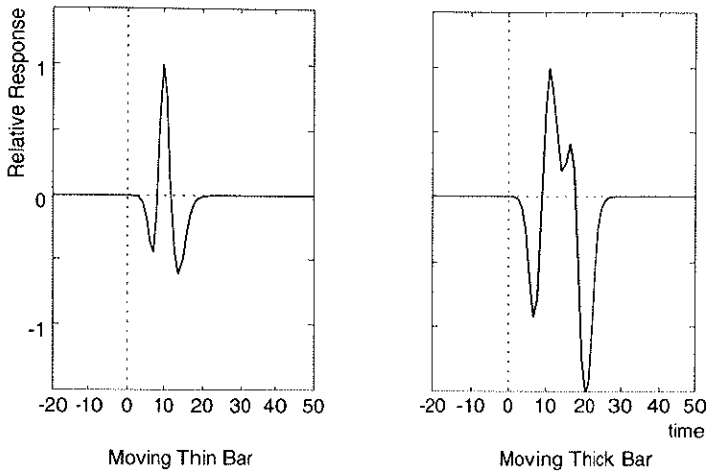


(a) 静止エッジ刺激に対する No.3 ニューロンの応答

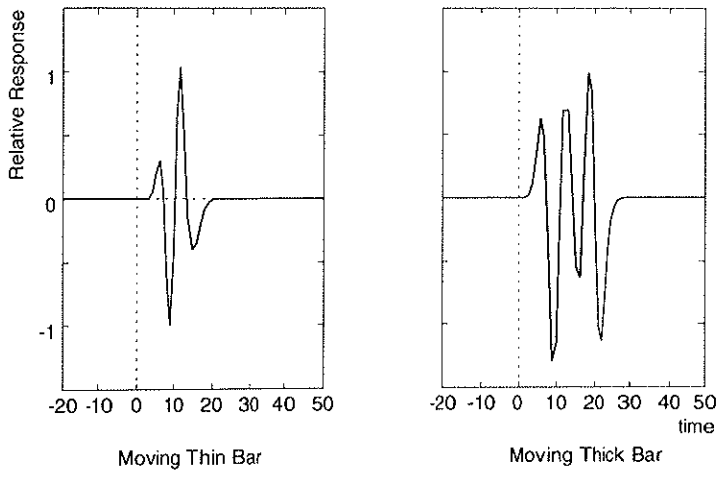


(b) 運動バー刺激に対する No.5 ニューロンの応答

図 2.5: 計算機シミュレーションによって得られたニューロンの出力応答 (その一): 横軸はシミュレーションの時間ステップを、縦軸は出力の大きさを表す。出力は最大値が 1 となるように正規化されている。また、静止刺激の提示開始時刻あるいは運動刺激が受容野内に入った時刻を $t = 0$ としている。



(a) 二種類の運動バー刺激に対する No.7 ニューロンの応答



(b) 二種類の運動バー刺激に対する No.11 ニューロンの応答

図 2.6: 計算機シミュレーションによって得られたニューロンの出力応答 (その二): 横軸はシミュレーションの時間ステップを、縦軸は出力の大きさを表す。出力は最大値が 1 となるように正規化されている。また、運動刺激が受容野内に入った時刻を $t = 0$ としている。

ロンはその重み分布からわかるように、受容野中心部の方が周辺部よりも持続時間の短い応答を示す (No.8 は逆に周辺部の方が短い応答を示す)。この特性は、ほにゅう類の網膜に存在しやはり中心-周辺型の受容野を有する X 型神経節細胞と類似している。一方、No.11 ニューロンは中心-周辺型で時間的に符号の反転した重みをもつため過渡的な応答特性を示すとみられ、これは同じく Y 型神経節細胞と対応づけることができる。そこで、これらニューロンの応答を生理実験データと比較するため、細いバー (Moving Thin Bar : バーの幅は 2 画素) および太いバー (Moving Thick Bar: 9 画素) の二種類の運動刺激に対する出力応答を求めた。結果を図 2.6(a) および (b) に示す。それぞれの応答は、ネコの生理実験より計算された神経節細胞の "response curve"[28] と類似していることがわかる (X 型 : 文献 [29] の図 1 および 2 ; Y 型 (細いバー) : 文献 [30] の図 1D および 2C ; Y 型 (太いバー) : 文献 [30] の図 2B および文献 [29] の図 5)。したがって、No.7 は X 型、No.11 は Y 型の網膜神経節細胞に対応したニューロンといえる。

2.4 受容野特性の理論的考察

2.3節の計算機シミュレーションの結果を理論的に解析する。ネットワークの重みの収束値を近似的に与える固有値問題を定式化し、このネットワークが獲得する時空間受容野の特性を解析する。また、入力信号のバイアスに対する受容野特性の変化についても考察する。

2.4.1 固有値問題の定式化

解析を容易にするため、ネットワークの入出力を時間的・空間的に連続な系として定式化し直す。ただし以下に示すように、本章で検討しているネットワークモデルにおいては、その時間特性はすべて伝達経路の番号に対する関数の形で表され、時刻を表す変数 t は陽には現れない。

ネットワークの入出力は次式のように書き表すことができる。

$$y_j(t) = \sum_{i=1}^{N_{ch}} \left\{ \int g_w(\boldsymbol{r}) w_j(\boldsymbol{r}, i) x(\boldsymbol{r}, i, t) d^2\boldsymbol{r} \right\} \tag{2.12}$$

ただし、

$w_j(\boldsymbol{r}, i)$: j番目のニューロンの重み関数
$x(\boldsymbol{r}, i, t)$: i 番目の伝達経路を経て入力される位置$\boldsymbol{r}$の信号
$y_j(t)$: j番目のニューロンの出力
$g_w(\boldsymbol{r})$: ガウシアン窓関数

である。窓関数は次式で定義される。

$$g_w(\boldsymbol{r}) = e^{\frac{\|\boldsymbol{r}\|^2}{2\sigma_w^2}} \tag{2.13}$$

計算機シミュレーションにおいて検討したネットワークモデルでは、学習の進行につれて、j番目のニューロンの出力は入力の第j主成分に近づく。したがって、式(2.12)のネットワークにおいては、重み関数 $w_j(\boldsymbol{r}, i)$ は入力相関関数

$$C(\boldsymbol{r}, i, \boldsymbol{r}', i') = \langle g_w(\boldsymbol{r}) x(\boldsymbol{r}, i) g_w(\boldsymbol{r}') x(\boldsymbol{r}', i') \rangle \tag{2.14}$$

の j番目に大きな固有値に対応した固有関数として求めることができる。ここで、

$$C^{ST}(\boldsymbol{r}, i, \boldsymbol{r}', i') = \langle x(\boldsymbol{r}, i) x(\boldsymbol{r}', i') \rangle \tag{2.15}$$

として

$$C(\boldsymbol{r}, i, \boldsymbol{r}', i') = g_w(\boldsymbol{r}) C^{ST}(\boldsymbol{r}, i, \boldsymbol{r}', i') g_w(\boldsymbol{r}') \tag{2.16}$$

とおくと、$C^{ST}(\boldsymbol{r}, i, \boldsymbol{r}', i')$ は空間位置のみの関数 $S(\boldsymbol{r}, \boldsymbol{r}')$ と伝達経路の番号のみの関数 $T(i, i')$ を用いて

$$C^{ST}(\boldsymbol{r}, i, \boldsymbol{r}', i') = S(\boldsymbol{r}, \boldsymbol{r}') T(i, i') + \mu^2 \tag{2.17}$$

と表される ($S(\boldsymbol{r}, \boldsymbol{r}')$ および $T(i, i')$ の定義については付録 A.1節参照)。ただし、μは入力となるガウシアンノイズ信号の平均値を表す。式 (2.14) の固有値をλ、固有関数を $z(\boldsymbol{r}, i)$ とおくと、このネットワークにおける固有値問題は次式で与えられる。

$$\sum_{i'=1}^{N_{ch}} \left\{ \int C(\boldsymbol{r}, i, \boldsymbol{r}', i') z(\boldsymbol{r}', i') d^2\boldsymbol{r}' \right\} = \lambda z(\boldsymbol{r}, i) \tag{2.18}$$

2.4.2 ゼロ平均の入力に対する固有関数

はじめに、入力のバイアスがない、すなわちゼロ平均 ($\mu = 0$) の場合について考察する。この条件のもとでは、$C(\boldsymbol{r}, i, \boldsymbol{r}', i')$ は$\boldsymbol{r}, \boldsymbol{r}'$のみの関数 (空間相関関数) と i, i' のみの関数 (時間相関関数) の積として次式のように表される。

$$C(\boldsymbol{r}, i, \boldsymbol{r}, i') = g_w(\boldsymbol{r}) S(\boldsymbol{r}, \boldsymbol{r}') g_w(\boldsymbol{r}') T(i, i') \tag{2.19}$$

したがって、固有値および固有関数も以下のように空間パートと時間パートの積として表すことができる。

$$\lambda = \lambda^S \lambda^T \tag{2.20}$$

$$z(\boldsymbol{r}, i) = z^S(\boldsymbol{r}) z^T(i) \tag{2.21}$$

ただし、λ^Sおよび $z^S(\boldsymbol{r})$ は空間相関関数 $g_w(\boldsymbol{r}) S(\boldsymbol{r}, \boldsymbol{r}') g_w(\boldsymbol{r}')$ の固有値および固有関数であり、λ^Tおよび $z^T(i)$ は時間相関関数 $T(i, i')$ の固有値および固有関数である。

これより、入力がゼロ平均の場合に自己組織される受容野は、空間特性と時間特性を独立に表すことのできる、いわゆる時空間分離型 [31] のものに限られることがわかる。それぞれの特性を以下に示す。

- 空間特性：$z^S(\boldsymbol{r})$ の特性については、本質的に Sanger[22, 23] および Linsker の結果 [27, 32] と同一である。$S(\boldsymbol{r}, \boldsymbol{r}')$ は系の回転に対して不変であるから、$z^S(\boldsymbol{r})$ はさらに動径方向の関数 $z^{S_r}(r)$ と角度方向の関数 $z^{S_a}(\theta)$ の積

$$z^S(\boldsymbol{r}) = z^{S_r}(r) z^{S_a}(\theta) \tag{2.22}$$

として表され、$z^{S_a}(\theta)$ は適当な位相を ψ とおいて

$$z^{S_a}(\theta) = \cos(l\theta + \psi) \quad (l = 0, 1, 2, \ldots) \tag{2.23}$$

と表される [32]。したがって、固有関数の空間パートは、図 2.7に示すような形状をとる。

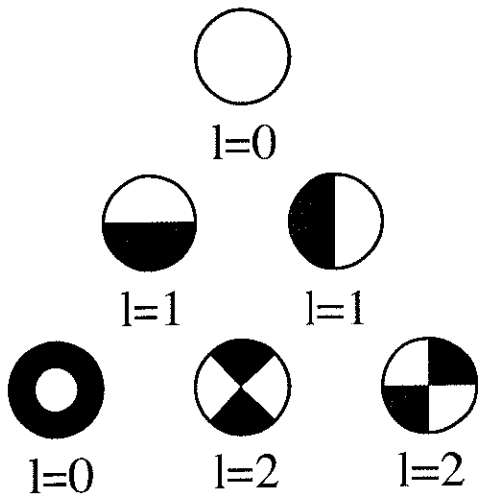


図 2.7: 固有関数の空間パートの形状：ゼロ平均の入力に対して得られる $z^S(\boldsymbol{r})$ の概形を上から対応する固有値の大きさの順に示す。

- 時間特性：$T(i,i')$ は $N_{ch} \times N_{ch}$の行列として$\boldsymbol{T} = [T(i,i')]$ と表せる。このとき、固有関数をならべてできるベクトル$\left[z^T(1), z^T(2), \ldots, z^T(N_{ch})\right]^T$は行列$\boldsymbol{T}$の固有ベクトルとなる。したがって、時間相関関数の固有関数は高々N_{ch}通りであり、系全体の固有関数 $z(\boldsymbol{r}, i)$ の時間特性も高々N_{ch}通りに分類される。数値計算によって求めた固有関数の値とインパルス応答関数から計算したニューロンの時間応答を図 2.8に示す。いずれも $N_{ch} = 3$ の条件で計算した時間応答 $\gamma_k(t)$ $(k = 1,2,3)$ を表し、(a) は$\alpha = 7$、(b) は$\alpha = 1$ の場合である。ただし $\gamma_k(t)$ は

$$\gamma_k(t) = \sum_{i}^{N_{ch}} z_k^T(i)\phi_i(t) \tag{2.24}$$

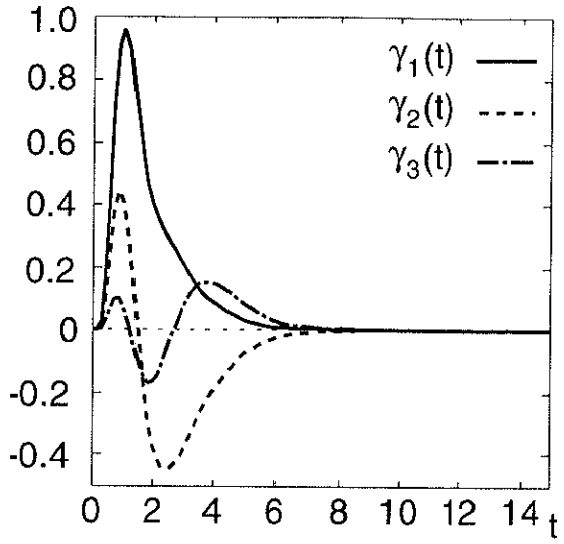
より求められ、固有関数の時間パートの k 番目に大きな固有値に対応する。ゼロ平均の場合に得られる受容野の時間特性は、入力のバイアスによらずこれらの応答のいずれかで与えられることになる。

2.4.3 バイアスされた入力の影響

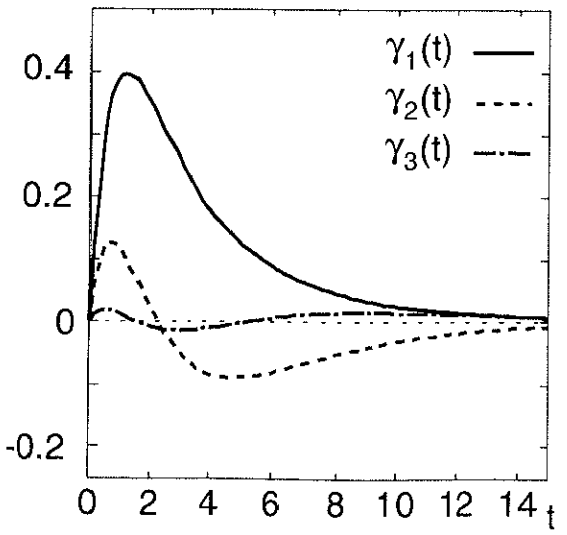
次に$\mu \neq 0$ の場合について考察する。式 (2.16) および式 (2.17) より、

$$\begin{aligned} & \int C(\boldsymbol{r}, i, \boldsymbol{r}', i') z(\boldsymbol{r}', i') d^2\boldsymbol{r}' \\ = & \int g_w(\boldsymbol{r}) \left(S(\boldsymbol{r}, \boldsymbol{r}') T(i,i') + \mu^2 \right) g_w(\boldsymbol{r}') z(\boldsymbol{r}', i') d^2\boldsymbol{r}' \qquad (2.25) \\ = & \; T(i,i') \int g_w(\boldsymbol{r})\; S(\boldsymbol{r}, \boldsymbol{r}') g_w(\boldsymbol{r}') z(\boldsymbol{r}', i') d^2\boldsymbol{r}' \\ & + \mu^2 \int g_w(\boldsymbol{r}) g_w(\boldsymbol{r}') z(\boldsymbol{r}', i') d^2\boldsymbol{r}' \qquad (2.26) \end{aligned}$$

と表せる。したがって、固有関数 $z(\boldsymbol{r}', i')$ が回転に対して不変でない、すなわち式 (2.23) において $l \neq 0$ である場合、式 (2.26) の第 2 項はゼロとなる。ゆえに、回転不変でない固有関数は入力のバイアスに影響を受けず、その空間・時間特性についてはゼロ平均の場合の結果がそのまま当てはまる。しかし、$l = 0$ の場合にはこの第 2 項はゼロになるとは限らず、回転不変な固有関数はバイアスの影響を受けて特



(a) $\alpha = 7$



(b) $\alpha = 1$

図 2.8: 固有関数の時間パートの値から求めたニューロンの時間応答

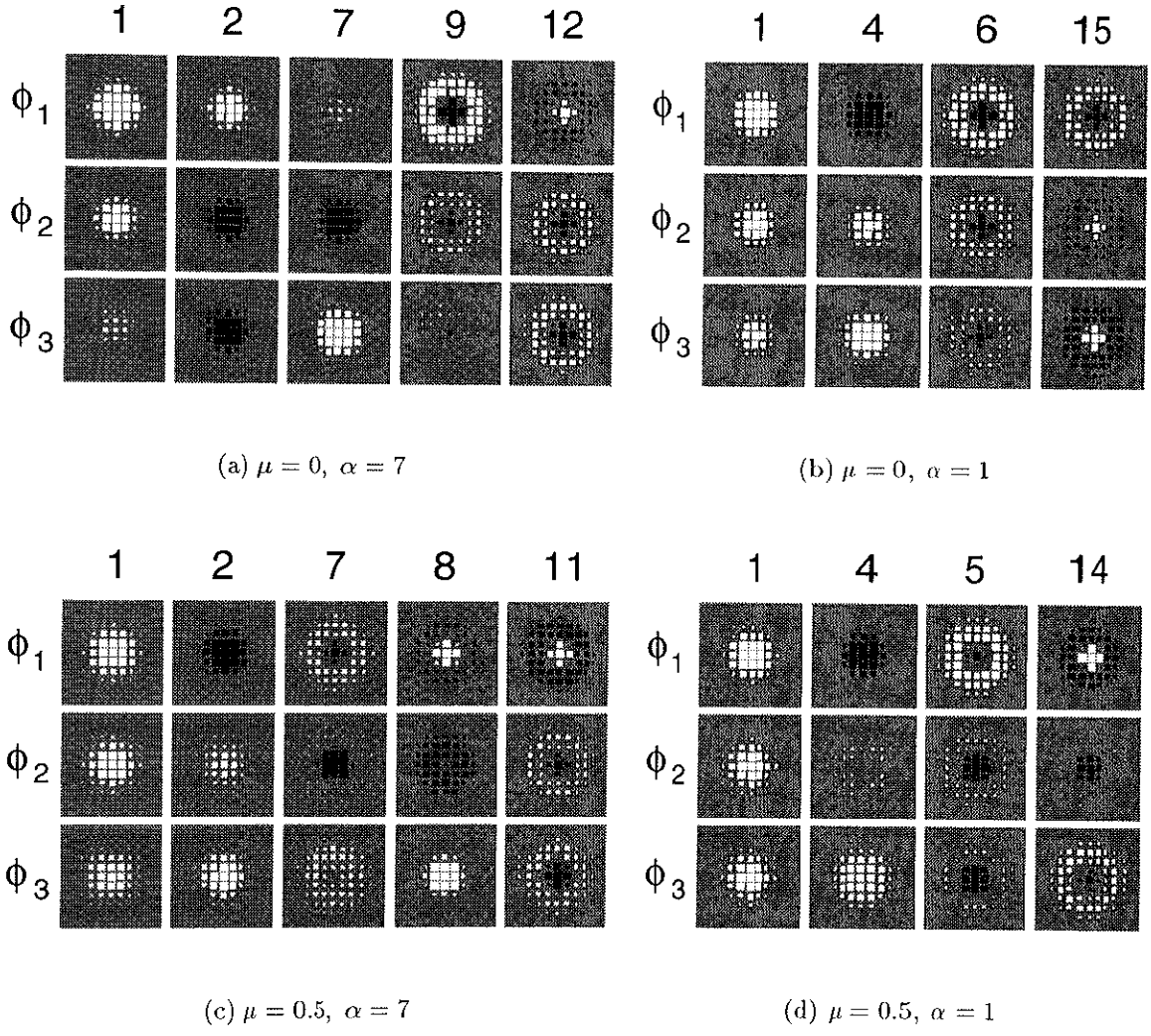
性を変化させ、空間特性と時間特性を分離して記述できない時空間非分離型となる可能性がある。

2.4.4 回転不変な受容野に関する数値計算

入力のバイアスと回転不変な受容野の特性の関係を詳しく調べるため、式 (2.18) の固有値問題を数値計算により解く。相関関数 $C(\boldsymbol{r}, i, \boldsymbol{r}', i')$ を計算機シミュレーションと同じ次元数に離散化し、その固有値および固有ベクトルを求めた。図 2.9は、図 2.4と同じ形式で固有ベクトルを表している。図 2.9(a) および (b) はゼロ平均で得られる固有ベクトルを示し、(c) および (d) はバイアスを加えて得られる固有ベクトルを表す。(a) および (c) は伝達経路のインパルス応答の持続時間を制御するパラメータを計算機シミュレーションと同じく $\alpha = 7$ とした場合であり、(b) および (d) は $\alpha = 1$ の場合である。これら以外のパラメータはすべて 2.3節の計算機シミュレーションと同じ値を用いている。したがって、(c) は計算機シミュレーションと同一の条件となっている。固有ベクトルの順位はαによって異なっているが、理論解析が示す通り、(a) および (b) では全ての固有ベクトルが時空間分離型となっている。一方、入力にバイアスの加わった (c) および (d) の場合には、一部の固有ベクトルが明らかに時空間で分離不可能な受容野となっていることがわかる。図 2.10は、図 2.9(c)No.7 ニューロンの受容野内の様々な位置に 1 画素のインパルス刺激を提示したときの応答を示す。刺激の提示位置によって応答曲線が変化しており、時空間非分離型の特性が確認できる。

2.4.5 考察

以上の解析結果は、図 2.4の計算機シミュレーション結果と良く一致している。例えば、No.3,9 および 16 のニューロンの受容野の空間パートはそれぞれ $l = 1, 2$ および 3 の場合に相当する。No.3,5 および 14 の空間パートは全て同一であり、時間パートのみ異なっている。また、16 種の受容野のうち No.1,2,7,8 および 11 の 5 つ



(a) $\mu = 0,\ \alpha = 7$

(b) $\mu = 0,\ \alpha = 1$

(c) $\mu = 0.5,\ \alpha = 7$

(d) $\mu = 0.5,\ \alpha = 1$

図 2.9: 数値計算によって得られた固有ベクトル：表示形式は図 2.4 に準ずる。上位 16 個の固有ベクトルのうち回転不変なもののみを示す。

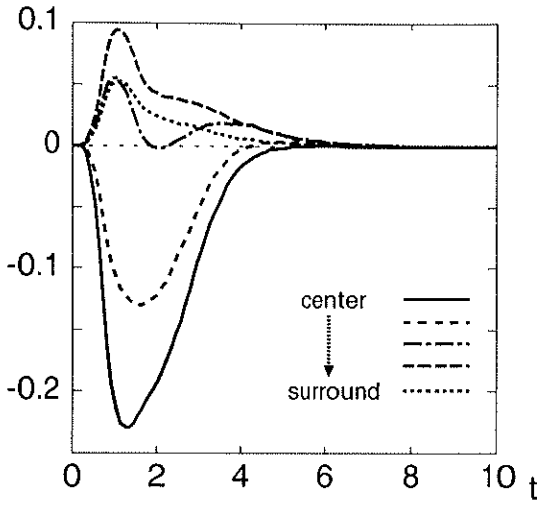


図 2.10: 回転不変な受容野の時間応答の例

が回転不変な空間特性を有しているが、いずれも重みの符号を除いて図 2.9(c) と順位まで含めて一致している。特に No.7 と 8 は明らかに時空間非分離型の受容野となっているが、No.7 は 2.3.3節の実験結果が示すように X 型の網膜神経節細胞と対応しており、この結果は時空間受容野の自己組織において入力のバイアスが必要である可能性を示唆している。

2.5 議論

本章では、線形ニューラルネットワークの教師なし学習による視覚系時空間受容野の自己組織形成モデルを提案した。ランダムノイズをローパスフィルタリングして空間的・時間的な相関を与えた信号を入力とし、その主成分を抽出することのできる教師なし学習アルゴリズムを用いて計算機シミュレーションを行なうことによって、生体の視覚系にみられるものと類似の特性を示す時空間受容野が自己組織されることを明らかにした。これは、従来のモデルでは検討されてこなかったニューロンの時間応答特性まで含めても、受容野の形成過程を情報量最大保持原理でうまく説明できることを示す結果といえる。特に、このモデルで獲得される中心-周辺型の受容野をもつ二つのニューロンが、それぞれ X 型および Y 型の網膜神経節細胞に類似した応答特性を示すという結果が得られた。さらに、理論解析の結果、このモデルでは入力のバイアスが中心-周辺型のような回転不変な受容野の特性に影響を与え、バイアスが存在する場合のみこれらの受容野の特性が時空間非分離型になることを明らかにした。実際に X 型や Y 型の神経節細胞では中心部と周辺部で時間応答が異なっているという生理学的知見 [33, 34, 35] があり、この点に関しては今後より詳細に検討する必要がある。

提案モデルでは、ニューロンへ信号を伝達する経路が複数存在し、それぞれ時間遅れ幅の異なるローパス型のインパルス応答特性を有することを仮定しており、この伝達経路がニューロンの時間特性を決定する重要な要素となっている。例えば、網膜においては光受容器からの信号が神経節細胞に伝達される過程には双極細胞、水平細胞やアマクリン細胞などの複数のニューロンが介在しており、複数の伝達経路が存在することは実際の視覚系においても自然なことである。また、式 (2.3) のローパス型のインパルス応答は、単純な線形再帰フィルタの多段接続としてモデル化される [26] ものである。以上のことから、上記の仮定はニューロンのモデルとして妥当なものと考える。

一方、このモデルの大きな問題点として、獲得されるニューロンの順位と実際の視覚系の階層構造との間に良い整合がみられないことがあげられる。実際の視覚系では大脳皮質に至ってはじめて方位選択性の受容野が出現するのに対して、モデルでは網膜神経節細胞に対応するニューロンよりも上位にエッジ検出型などの方位選択性のものが獲得されている。また、理論解析の結果が示すように、実際の視覚系にみられるようなバー検出型の受容野や方位選択性で時空間非分離型の受容野は、このモデルでは獲得することができない。したがって、ネットワークの構造を見直す、ニューロン出力間の相互作用を考慮に入れるなどの改良によってより整合性の良いモデルを構成することが今後の課題となる。

第 3 章

多層パーセプトロンの学習による主成分分析

3.1 導入

Rumelhart[6] による誤差逆伝搬学習則の提案以降、多層パーセプトロンの学習理論とその応用が多様に研究されている。その理論的側面においては、中間層にシグモイド型の非線形素子を含む 3 層パーセプトロンの近似万能性の証明 [36, 37, 38] がなされており、入力データ自身を出力教師とする恒等写像学習の近似能力に関する理論的解析 [39, 40, 42] も報告されている。本章では、多層パーセプトロンの恒等写像学習 [7] を対象とし、従来その学習に用いられてきたエネルギー関数の問題点を指摘する。その上でこの問題を解決するための新たなエネルギー関数を導入し、その学習特性を理論的・実験的に検討する。

多層パーセプトロンの恒等写像学習に関しては、これまでの研究により以下のことが知られている。

1. 3層線形パーセプトロンは学習によって主成分分析 (あるいはKarhunen-Loève(KL) 変換) と等価な変換を実現できる [39, 40, 41]。

2. 中間層素子にシグモイド関数を用いた 3 層非線形パーセプトロンの近似能力は線形パーセプトロンの能力を超えない [42]。

3. 5 層以上の非線形パーセプトロンは非線形な次元圧縮・主成分分析を実現することができ、その近似能力は線形の主成分分析を上回る [43, 44, 45]。

これらの中で、Baldi and Hornik[40] は、3 層線形パーセプトロンの恒等写像近似能力を理論的に保証する定理 (Baldi–Hornik の定理) を与えている。この定理によれば、データの平均二乗誤差をエネルギー関数とする従来の恒等写像学習においては、パーセプトロンの内部表現すなわち中間層の出力がデータの主成分と同一の部分空間を張ることがエネルギー関数最小化の必要十分条件となる。ゆえに、学習収束後のパーセプトロンは主成分分析と等価な恒等写像の最適線形近似を実現することが保証される。しかしこの場合、内部表現を構成する各中間層素子の出力が主成分と一対一に対応するわけではない。そのため、主成分分析では近似に対する各主成分の寄与の大きさを評価し、これらを重要度の順に並べることができるのに対して、パーセプトロンでは中間層素子の出力について寄与順が定まらず、このような成分分析は困難である。また、データの特徴量が中間層素子の出力全体で分散表現されるため、一部の成分のみを取り出して利用することも不可能である。そこで本章では、従来の平均二乗誤差をエネルギー関数とするパーセプトロンの恒等写像学習に関するこれらの問題を解決するため、寄与順に並んだ内部表現の自己組織を可能とする新しいエネルギー関数を検討する。

はじめに、3.2節で線形パーセプトロンの恒等写像学習に関する理論的検討を行なう。Baldi–Hornik の定理について述べた後、内部表現として主成分を獲得するためのエネルギー関数 (重畳エネルギー関数) を定義する。Baldi–Hornik の定理を拡張し、重畳エネルギー関数に基づく 3 層線形パーセプトロンの内部表現が主成分分析と一致することを証明する。次に、3.3節で重畳エネルギー関数を用いた恒等写像学

習の数値実験を行ない、学習の収束性を確認する。3.4節では重畳エネルギー関数の非線形主成分分析への適用可能性を検討する。5 層非線形パーセプトロンを用いた数値実験により、非線形な特徴抽出が必要とされる学習課題においても寄与順に並んだ効率的な内部表現を自己組織できることが示される。最後に、3.5節で本章の結果について議論する。特に、線形パーセプトロンで主成分を抽出する方法として渡辺ら [46, 47, 48] によって提案されている手法との比較を行なう。

3.2 重畳エネルギー関数の定義とその特性の理論解析

3.2.1 3 層線形パーセプトロン

線形素子のみから成り、入力層、中間層、出力層の素子数がそれぞれN, H, Mの3層線形パーセプトロンを以下のように定義する。

定義 3.1 3層線形パーセプトロン

領域$U(\subset \boldsymbol{R}^N)$および領域$V(\subset \boldsymbol{R}^H)$上の連続写像

$$\psi \ : \ V \to W \quad (\text{ただし } W \subset \boldsymbol{R}^M) \tag{3.1}$$

$$\phi \ : \ U \to V \tag{3.2}$$

によるUからWへの合成写像

$$\psi \circ \phi \quad : \quad U \to W \tag{3.3}$$

を考える。二つの写像が共に線形、すなわち

$$\begin{aligned} \psi \ &: \ \boldsymbol{u} \mapsto A_H \boldsymbol{u} \quad (\boldsymbol{u} \in V) \\ \phi \ &: \ \boldsymbol{x} \mapsto B_H \boldsymbol{x} \quad (\boldsymbol{x} \in U) \end{aligned} \tag{3.4}$$

ただし、

$$\begin{aligned} A_H &= [\boldsymbol{a}_1 | \boldsymbol{a}_2 | \cdots | \boldsymbol{a}_H] \quad \boldsymbol{a}_i \text{は} M \text{次元列ベクトル} \\ B_H &= [\boldsymbol{b}_1 | \boldsymbol{b}_2 | \cdots | \boldsymbol{b}_H]^T \quad \boldsymbol{b}_i \text{は} N \text{次元列ベクトル} \end{aligned} \tag{3.5}$$

である場合、合成写像

$$\psi \circ \phi \quad : \quad \boldsymbol{x} \mapsto A_H B_H \boldsymbol{x} \tag{3.6}$$

を3層線形パーセプトロンと呼ぶ。

定義 3.2 重みベクトル

上記の3層線形パーセプトロンにおいて、$\boldsymbol{a}_h \in \boldsymbol{R}^M$および$\boldsymbol{b}_h \in \boldsymbol{R}^N$を中間層第$h$素子に結合する重みベクトルと呼ぶ。$\boldsymbol{a}_h$は第$h$素子と出力層の各素子とを結合する

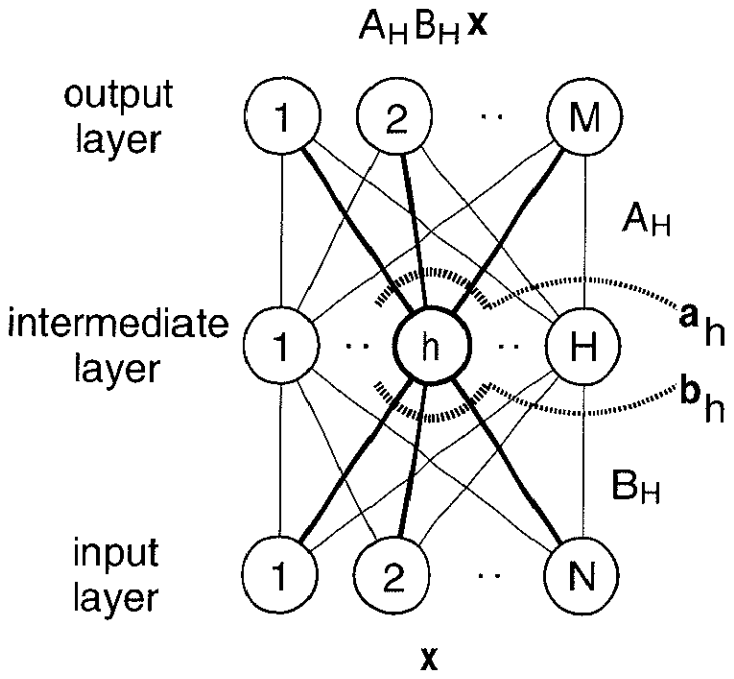


図 3.1: 3 層線形パーセプトロン

重み値から成るベクトルであり[1]、$\boldsymbol{b}_h$は第 h 素子と入力層との間の重み値から成るベクトルである（図 3.1参照)。

定義 3.3 内部表現

上記の 3 層線形パーセプトロンにおいて、ある入力値$\boldsymbol{x} \in U$が与えられたときの中間層素子の値

$$\boldsymbol{u} = B_H \boldsymbol{x} \in V \tag{3.7}$$

を入力$\boldsymbol{x}$に対するパーセプトロンの内部表現と呼ぶ。

定義 3.4 砂時計型 3 層線形パーセプトロン

[1] 一般的には、中間層-出力層間の重みベクトルはひとつの出力層素子と中間層の各素子とを結合する重み値から成るベクトルとして定義されるが、本論文では文献 [40] の表記に従い、このように定義する。

上記の 3 層線形パーセプトロンにおいて条件

$$N = M \geq H > 0 \tag{3.8}$$

が成立する場合、このパーセプトロンの構造は砂時計型であるという。

U上の N次元ベクトル値の集合（データ集合）

$$\{\boldsymbol{x}_t \in U\}_{t=1}^{T} \tag{3.9}$$

に対して、式 (3.6) の 3 層線形パーセプトロンが砂時計型であるならば、このパーセプトロンはU上の恒等写像の近似を与える。また、その平均二乗誤差は

$$E_H = \frac{1}{T}\sum_{t=1}^{T} ||\boldsymbol{x}_t - A_H B_H \boldsymbol{x}_t||^2 \tag{3.10}$$

で与えられる。

3.2.2 主成分分析

3.2.1節で述べた恒等写像近似の最適解を与える線形写像の構成法は、以下に示す主成分分析として知られている。以降、データ集合 (3.9) に対して以下の仮定をおく[2]。

1. ゼロ平均である

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t = \mathbf{o} \tag{3.11}$$

2. 共分散行列

$$\Sigma = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t^T \tag{3.12}$$

が N個の実固有値$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$ をもつ。ここで、λ_hに対応する正規固有ベクトルを$\boldsymbol{u}_h$と表す。

[2]厳密には、誤差の定義なども含めて標本平均ではなく期待値を用いるべきであるが、以降の議論を簡単にするため、このような定義を採用した。

このとき、以下の定理が成り立つ。

定理 3.1 主成分分析

$A_H = B_H^T = W^T, WW^T = I$なる制約のもとで式 (3.10) の平均二乗誤差を最小にする行列 Wは

$$W = TU_H^T \tag{3.13}$$

で与えられる。ただし、Tは任意の $H \times H$(正規) 直交行列であり、U_Hは

$$U_H = [\boldsymbol{u}_1|\boldsymbol{u}_2|\ldots|\boldsymbol{u}_H] \tag{3.14}$$

なる $N \times H$行列である。

合成写像 $W^TW = U_HU_H^T$は、N次元入力を H次元部分空間へ射影することによる恒等写像近似を実現し、H個の基底ベクトルを用いた恒等写像の線形近似において、式 (3.10) を最小化するという意味で最適な写像である。また、式 (3.13) において特に $T = I$とした場合、ベクトル$\boldsymbol{x}$に対して $W\boldsymbol{x}$の成分

$$\boldsymbol{u}_h^T\boldsymbol{x} \tag{3.15}$$

は$\boldsymbol{x}$の第 h 主成分と呼ばれる。第 h 主成分の分散は固有値λ_hに等しく、大きな固有値に対応する主成分ほど近似誤差減少に対する寄与が大きい。

通常主成分分析を行なう際には、データ集合の共分散行列Σに関する固有値問題を数値的に解いて固有ベクトルを求め、$W = U_H^T$を構成する。この場合、Wによって抽出される特徴量は主成分そのものであるから、各特徴量の寄与順が定まり、それらの重要度を評価することが可能となる。また、必要に応じて上位のいくつかの主成分のみを取り出して利用することも可能である。一般に次元数 Hをあらかじめ適切に定めておくことは困難であるから、抽出された成分の重要度を評価して一部の成分を取り出すことができるというこの性質は特徴抽出などにおいて重要なものである。

3.2.3 従来のエネルギー関数と Baldi–Hornik の定理

砂時計型 3 層線形パーセプトロンによる恒等写像近似においては、式 (3.10) の平均二乗誤差をエネルギー関数とする誤差逆伝搬学習が従来用いられてきた。この場合、パーセプトロンの重み行列に関して以下の定理が成り立つ。

定理 3.2 Baldi and Hornik[40]

フルランク行列$\tilde{A}_H$, $\tilde{B}_H$がE_Hの最小値を与えるならば、$H \times H$正則行列C_Hが存在し、

$$\begin{aligned} \tilde{A}_H &= U_H C_H \\ \tilde{B}_H &= C_H^{-1} U_H^T \end{aligned} \tag{3.16}$$

が成立する。また、その逆が成立する。

式 (3.10) で与えられるエネルギー関数は二次形式であるため、パーセプトロンの重み行列 A_H, B_Hは学習によって式 (3.16) の行列$\tilde{A}_H$, $\tilde{B}_H$に収束する。このとき

$$\tilde{A}_H \tilde{B}_H = U_H U_H^T \tag{3.17}$$

であるから、パーセプトロン$\tilde{A}_H \tilde{B}_H$は主成分分析と等価な最適線形近似写像を実現する。しかし、重み行列$\tilde{A}_H$および$\tilde{B}_H$と固有ベクトルから成る行列 U_HおよびU_H^Tはそれぞれ正則行列 C_HおよびC_H^{-1}によって対応付けられるため、$\tilde{A}_H$, $\tilde{B}_H$を構成する重みベクトル$\tilde{\boldsymbol{a}}_h$, $\tilde{\boldsymbol{b}}_h$について必ずしも

$$\tilde{\boldsymbol{a}}_h \mathbin{/\!/} \tilde{\boldsymbol{b}}_h \mathbin{/\!/} \boldsymbol{u}_h \quad (h = 1, 2, \ldots, H) \tag{3.18}$$

が成立しない。ゆえに、式 (3.10) で与えられる従来のエネルギー関数を用いた砂時計型 3 層パーセプトロンの恒等写像学習では、獲得される内部表現$\tilde{B}_H \boldsymbol{x}_t$は主成分分析の結果 $W\boldsymbol{x}_t = U_H^T \boldsymbol{x}_t$と一致せず、パーセプトロンが抽出する特徴量すなわち中間層素子の出力について、寄与順を定めることができない。この点が従来のパーセプトロンによる恒等写像学習の大きな欠点である。次節では、パーセプトロンの内部表現として主成分を獲得することのできる学習法を検討する。

3.2.4 重畳エネルギー関数と Baldi–Hornik の定理の拡張

3.2.1節の砂時計型3層線形パーセプトロンに対して、内部表現として主成分を獲得するためのエネルギー関数(重畳エネルギー関数)を定義する。また、Baldi–Hornikの定理(定理3.2)を拡張し、このエネルギー関数を用いた恒等写像学習に関する定理を証明する。

定義 3.5 部分パーセプトロン

式(3.6)および(3.8)で定義される砂時計型3層線形パーセプトロンにおいて、中間層素子を任意に番号付けし、中間層第i素子までを考慮した重み行列を

$$\begin{aligned} A_i &= [\boldsymbol{a}_1|\boldsymbol{a}_2|\cdots|\boldsymbol{a}_i] \\ B_i &= [\boldsymbol{b}_1|\boldsymbol{b}_2|\cdots|\boldsymbol{b}_i]^T \end{aligned} \tag{3.19}$$

とおく。このとき、合成写像

$$\boldsymbol{x} \mapsto A_i B_i \boldsymbol{x} \tag{3.20}$$

をこのパーセプトロンの第i部分パーセプトロンと呼ぶ。図3.2に示すように、第i部分パーセプトロンは$i+1$番目からH番目までの中間層素子を無視して得られる部分的なパーセプトロンである。

定義 3.6 部分エネルギー関数

上述のパーセプトロンにおいて、式(3.10)の平均二乗誤差と同様に、第i部分パーセプトロンの平均二乗誤差

$$E_i = \frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{x}_t - A_i B_i \boldsymbol{x}_t\|^2 \tag{3.21}$$

を求めることができる。これを第i部分エネルギー関数と呼ぶ。

ここで、第H部分パーセプトロンは与えられた素子を全て用いるパーセプトロンそのものであり、第H部分エネルギー関数は従来の恒等写像学習で用いられてきたエネルギー関数に一致することに注意する。以上の定義の下で、重畳エネルギー関数は以下のように定義される。

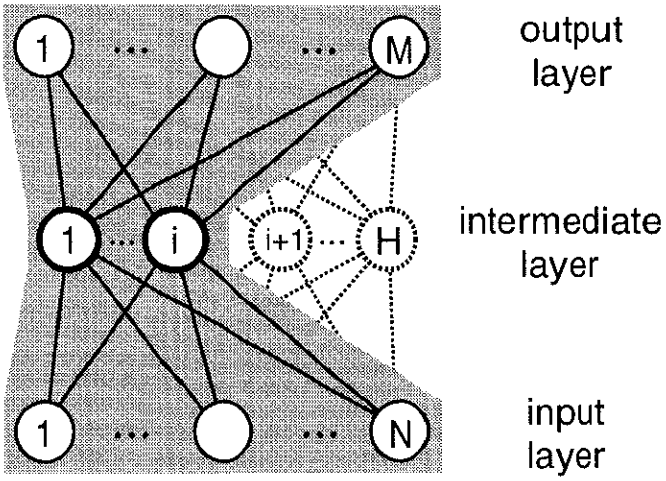


図 3.2: 3 層パーセプトロンとその部分パーセプトロン

定義 3.7 重畳エネルギー関数

全ての部分エネルギー関数の重みつき和

$$F_H = \sum_{i=1}^{H} \beta_i E_i \tag{3.22}$$

を重畳エネルギー関数と呼ぶ。ただし、β_iは任意の正定数とする。

例えば、$H = 2$としてF_2に基づく学習を行うパーセプトロンを考えると、中間層第 2 素子に結合する重みの修正量はE_2のみで決まるが、第 1 素子に結合する重みについては、E_2だけでなく、第 2 素子なしで計算される誤差E_1も同時に減少させるように学習が進行する。その結果、中間層第 1 素子は最も重要な特徴量を、第 2 素子はこれを補う特徴量を出力するようになると考えられる。よって一般に、中間層素子の番号順に寄与の大きな特徴を抽出するように学習が進行することが期待できる。重畳エネルギー関数を用いる恒等写像学習においては、以下の定理が成り立つ。

定理 3.3 フルランク行列 $\tilde{A}_H, \tilde{B}_H$ が F_H の最小値を与えるならば、正則な $H \times H$ 対角行列 D_Hが存在し、

$$\begin{aligned} \tilde{A}_H &= U_H D_H \\ \tilde{B}_H &= D_H^{-1} U_H^T \end{aligned} \tag{3.23}$$

が成立する。また、その逆が成立する。

学習収束時のパーセプトロン $\tilde{A}_H \tilde{B}_H$ が最適線形近似を与えるのは従来と同様である。しかし、$D_H = diag(d_1, \ldots, d_H)$ とおくと、式 (3.14), (3.23) より、$\tilde{A}_H$ および $\tilde{B}_H$ を構成する重みベクトル $\tilde{\boldsymbol{a}}_h, \tilde{\boldsymbol{b}}_h$ は

$$\begin{aligned} \tilde{\boldsymbol{a}}_h &= d_h \boldsymbol{u}_h \\ \tilde{\boldsymbol{b}}_h &= \tfrac{1}{d_h} \boldsymbol{u}_h \end{aligned} \tag{3.24}$$

と表される。したがって、従来のエネルギー関数の場合と異なり、

$$\tilde{\boldsymbol{a}}_h \,/\!/\, \tilde{\boldsymbol{b}}_h \,/\!/\, \boldsymbol{u}_h \quad (h = 1, 2, \ldots, H) \tag{3.25}$$

が重畳エネルギー関数 F_H の最小化の必要十分条件となる。ゆえに、中間層素子は寄与順に主成分に対応した値を出力するように自己組織される。

3.2.5 定理の証明

定理 3.3の証明のために以下の補題を準備する。

補題 3.4 N 個の正規直交ベクトル $\{\boldsymbol{u}_h \in \boldsymbol{R}^N\}_{h=1}^N$ から i 個を取り出してつくった行列

$$U_i = [\boldsymbol{u}_1 | \boldsymbol{u}_2 | \cdots | \boldsymbol{u}_i] \tag{3.26}$$

および非零の定数 $\{d_h\}_{h=1}^i$ から成る i 次対角行列

$$D_i = diag(d_1, \ldots, d_i) \tag{3.27}$$

を考える。これらを用いて $M \times i$ 行列 A_i および $i \times N$ 行列 B_i を

$$\begin{aligned} A_i &= U_i D_i \\ B_i &= D_i^{-1} U_i^T \end{aligned} \tag{3.28}$$

と定義し、これらに任意の M 次元ベクトル $\boldsymbol{a}_{i+1}$ および N 次元ベクトル $\boldsymbol{b}_{i+1}$ を付加して作った $M \times (i+1)$ 行列 A_{i+1} および $(i+1) \times N$ 行列 B_{i+1} を

$$\begin{aligned} A_{i+1} &= [A_i | \boldsymbol{a}_{i+1}] \\ B_{i+1} &= \left[\frac{B_i}{\boldsymbol{b}_{i+1}^T} \right] \end{aligned} \tag{3.29}$$

とする。このとき、

$$\begin{aligned} A_{i+1} &= U_{i+1} D_{i+1} \\ B_{i+1} &= D_{i+1}^{-1} U_{i+1}^T \end{aligned} \tag{3.30}$$

を満足する $(i+1)$ 次正則行列 D_{i+1} が存在するならば、 D_{i+1} は対角行列である。

補題 3.4の証明については、付録 A.2参照。

証明 以下の命題を順に証明する。

(i) $\tilde{A}_H, \tilde{B}_H$ が式 (3.23) で与えられるならば、すべての E_i $(i = 1, 2, \ldots, H)$ が最小となる。

(ii) $\tilde{A}_H, \tilde{B}_H$ がすべての E_i を最小化するならば、 $\tilde{A}_H, \tilde{B}_H$ は式 (3.23) の形式で与えられる。

(iii) すべての E_i が最小となるならば、 F_H も最小である。また、 F_H が最小であるならばすべての E_i が最小となる。

(i) D_H が対角行列であるため、 $\tilde{A}_H$,$\tilde{B}_H$ が式 (3.23) を満足するならば、すべての $i(= 1, 2, \ldots, H)$ において

$$\begin{aligned} \tilde{A}_i &= U_i D_i \\ \tilde{B}_i &= D_i^{-1} U_i^T \end{aligned} \tag{3.31}$$

が成立する。よって定理3.2より直ちに、すべての i において E_i は最小となることが示される。

(ii) 重み行列 $\tilde{A}_H, \tilde{B}_H$ がすべての E_i $(i = 1, 2, \ldots, H)$ を最小化するならば、定理3.2より、 $\tilde{A}_i, \tilde{B}_i$ は正則な i 次正方行列 C_i を適当に定めて

$$\begin{aligned} \tilde{A}_i &= U_i C_i \\ \tilde{B}_i &= C_i^{-1} U_i^T \end{aligned} \tag{3.32}$$

と表される。ただし、C_i は C_H の i 次主座小行列である。ここで補題より、$\tilde{A}_1, \tilde{B}_1$ が式 (3.32) を満たすとき、$\tilde{A}_2, \tilde{B}_2$ が式 (3.32) を満たすには C_2 は対角行列でなければならない。同様にして、帰納法により $i = 3, \ldots, H$ についても C_i は対角行列となる。したがって、対角行列 C_H を D_H と書き直せば、

$$\begin{aligned} \tilde{A}_H &= U_H D_H \\ \tilde{B}_H &= D_H^{-1} U_H^T \end{aligned} \tag{3.33}$$

が得られる。

(iii) (i) よりすべての i $(= 1, 2, \ldots, H)$ において E_i を最小化する重み行列の存在が示された。すべての E_i が最小であるならば明らかに F_H は最小となる。逆に、少なくとも一つの i において E_i が最小でないときには F_H は最小でないため、F_H が最小となるのはすべての E_i が最小の場合に限られる。 (証明終り)

3.3 数値実験による学習特性の検証

重畳エネルギー関数を用いた恒等写像学習の収束性を確認するため、数値実験を行なう。

3.3.1 実験条件

実験には

$$N = H = M = 3 \tag{3.34}$$

なる砂時計型 3 層線形パーセプトロンを用い、ゼロ平均のガウス分布に従うベクトル値からデータ集合

$$\{\boldsymbol{x}_t \in \boldsymbol{R}^3\}_{t=1}^{1000} \tag{3.35}$$

を生成した。学習には従来のエネルギー関数 E_H および重畳エネルギー関数 F_H の両方を適用し、学習回数と学習定数はそれぞれ 1000 回, 1.0 とした。また、F_H における定数 β_i はすべて同一の値すなわち

$$\beta_i = \frac{1}{H} \tag{3.36}$$

とした。

3.3.2 実験結果と考察

図 3.3 は、重畳エネルギー関数を用いた場合の学習回数に対する部分エネルギー関数 $\{E_i\}_{i=1}^3$ の値の変化を示す。各部分パーセプトロンの誤差が対応する次元数の主成分分析の誤差に収束していることがわかる。一方図 3.4 は、重畳エネルギー関数による学習後の重みベクトル $\{\boldsymbol{b}_h\}_{h=1}^3$ をデータ集合に重ねて表示している。中間層第 1 素子の重みベクトル $\boldsymbol{b}_1$ が第 1 主成分方向すなわちデータ集合の最大分散方向を向き、$\boldsymbol{b}_2$ がそれと直交した平面内で次に分散の大きな方向を、$\boldsymbol{b}_3$ が両者に直交した方向を向いていることが理解される。結果をより詳しく比較するため、パー

セプトロンの重みベクトルと主成分分析の固有ベクトル $\{\boldsymbol{u}_h\}_{h=1}^3$ との方向余弦を計算した。表 3.1 左に示すように、重畳エネルギー関数の場合には各重みベクトルが対応する固有ベクトルに収束している。しかし、同一条件下で従来のエネルギー関数による学習を行った場合 (表 3.1 右)、このような対応は見られない。

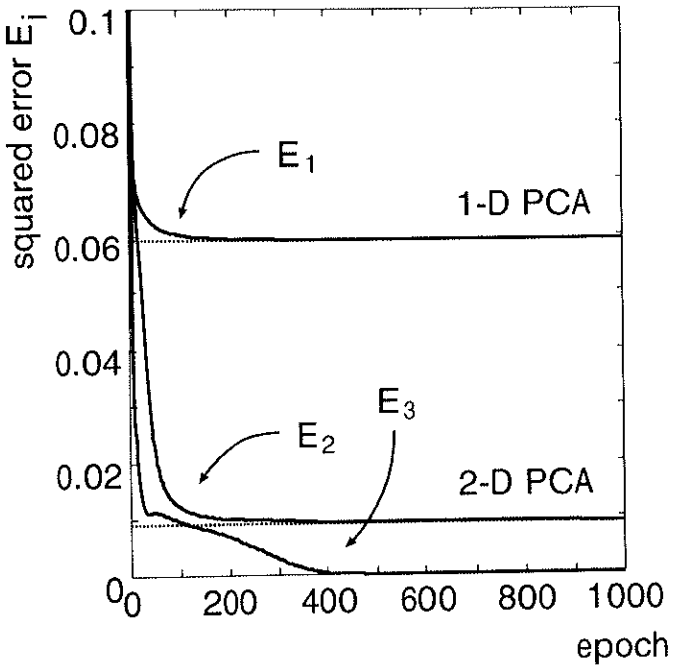


図 3.3: 重畳エネルギー関数を用いた砂時計型 3 層線形パーセプトロンの学習曲線：横軸はパーセプトロンの学習回数を、縦軸は二乗誤差 E_i の値を示す。図中の破線はデータ集合に対して第 1 主成分のみおよび 第 2 主成分までを用いて恒等写像近似を行なった場合の二乗誤差を表す。

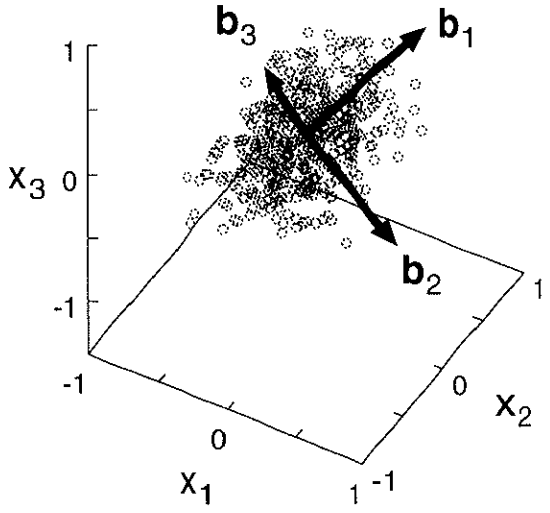


図 3.4: データ集合に重ねて表示した学習後の重みベクトル：破線の丸は入力データを表し、矢線は重みベクトルを表す。

表 3.1: 中間層素子の重みベクトル $\{\boldsymbol{b}_h\}_{h=1}^3$ と固有ベクトル $\{\boldsymbol{u}_h\}_{h=1}^3$ との方向余弦：左は重畳エネルギー関数を用いた場合、右は従来のエネルギー関数の場合を示す

	$\boldsymbol{b}_1$	$\boldsymbol{b}_2$	$\boldsymbol{b}_3$
$\boldsymbol{u}_1$	1.0000	0.0005	-0.0001
$\boldsymbol{u}_2$	-0.0006	0.9996	-0.0060
$\boldsymbol{u}_3$	0.0009	0.0271	1.0000

	$\boldsymbol{b}_1$	$\boldsymbol{b}_2$	$\boldsymbol{b}_3$
$\boldsymbol{u}_1$	0.2455	-0.5134	-0.7807
$\boldsymbol{u}_2$	-0.5730	0.5292	-0.5681
$\boldsymbol{u}_3$	-0.7819	-0.6755	0.2603

3.4 非線形主成分分析への適用

3.4.1 多層パーセプトロンによる非線形主成分分析

前節まででは、線形パーセプトロンの恒等写像学習で主成分分析を実現することを目標として重畳エネルギー関数を検討してきた。主成分分析は線形写像によって特徴量すなわち主成分を抽出し次元圧縮を行なう手法であるから、データの変数間の関係が線形であり、データの特徴が線形関数でうまくとらえられることを仮定している。そのため、変数間に非線形な関数関係が内在するような場合には、線形写像による主成分分析ではデータの本質的な特徴量を抽出することができず、次元圧縮の効果が得られない。このような場合、式 (3.1),(3.2) の ψ および ϕ に対して適当な非線形写像のクラスを設定し、合成写像 $\psi \circ \phi$ による恒等写像近似の平均二乗誤差

$$E_H = \frac{1}{T}\sum_{t=1}^{T} ||\boldsymbol{x}_t - \psi \circ \phi(\boldsymbol{x}_t)||^2 \tag{3.37}$$

を最小にする写像を構成することによって、効果的な次元圧縮を実現できると考えられる。このような手法は一般に非線形主成分分析と呼ばれる。

多層パーセプトロンにおいては、図 3.5に示すような非線形素子を含む 5 層の砂時計型パーセプトロン[3]、あるいはより多層の砂時計型非線形パーセプトロンを用いることで、非線形主成分分析を実現することができる [43, 44, 45]。この場合、ボトルネックとなる中間層の素子出力がデータの特徴量となる。本章では、3 層パーセプトロン同様にこのボトルネック層の出力を内部表現と呼ぶ。多層パーセプトロンによる非線形主成分分析では、線形の主成分分析と比較して、同次元数で近似したときの誤差のより小さい内部表現が獲得されることが期待できる。しかし、式 (3.37) をエネルギー関数とする従来の恒等写像学習では、3 層パーセプトロン同様に内部表現の寄与順が定まらないため、効率的な非線形次元圧縮を実現するには学習前に

[3]任意の多層パーセプトロンにおいても、入力層と出力層の素子数を N および M とし、それ以外の中間層の素子数の最小値を H としたとき、式 (3.8) が成立するパーセプトロンを砂時計型と定義する。

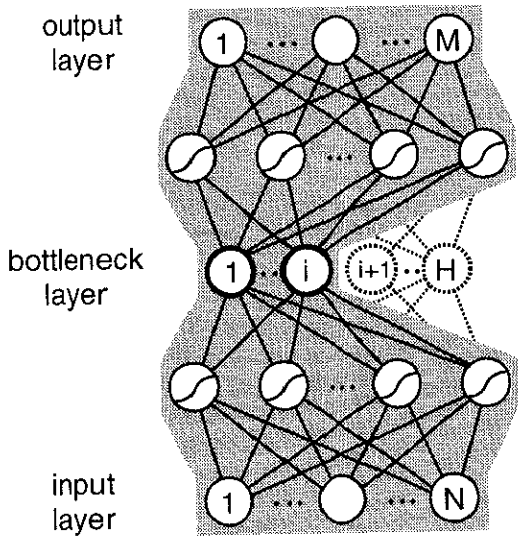


図 3.5: 5 層非線形パーセプトロンとその部分パーセプトロン

中間層の素子数を最適に定めてやらねばならないという問題がある。これに対して、重畳エネルギー関数は以下に示すように 3 層砂時計型パーセプトロン以外の任意の多層パーセプトロンに対して定義することができるため、非線形パーセプトロンの場合にも寄与順の内部表現が自己組織される可能性がある。そこで以下では、砂時計型非線形パーセプトロンに対する重畳エネルギー関数の定義を与え、5 層パーセプトロンを用いた数値実験によりその効果を検証する。

図 3.5の砂時計型 5 層パーセプトロンに対して、部分パーセプトロンおよび部分エネルギー関数を以下の様に定義する。

定義 3.8 部分パーセプトロン

図 3.5のパーセプトロンにおいて、H 個の中間層素子のうち第 i 素子まで考慮したパーセプトロンを第 i 部分パーセプトロンと呼ぶ。入力からこれら i 個の中間層素子出力への写像を ϕ_i と表記し、中間層素子出力からパーセプトロン出力への写像を ψ_i と表記すれば、第 i 部分パーセプトロンの合成写像は

$$f_i = \psi_i \circ \phi_i \tag{3.38}$$

と表される。

定義 3.9 部分エネルギー関数

上述のパーセプトロンに対して、第 i 部分パーセプトロンの平均二乗誤差

$$E_i = \frac{1}{T}\sum_{t=1}^{T} ||\boldsymbol{x}_t - f_i(\boldsymbol{x}_t)||^2 \tag{3.39}$$

を第 i 部分エネルギー関数と呼ぶ。

このとき、重畳エネルギー関数は定義 3.7 と全く同様に定義される。

定義 3.10 重畳エネルギー関数

全ての部分エネルギー関数の重みつき和

$$F_H = \sum_{i=1}^{H} \beta_i E_i \tag{3.40}$$

を重畳エネルギー関数と呼ぶ。ただし、β_i は任意の正定数とする。

3.4.2 半球面の近似

数値実験により、重畳エネルギー関数を用いたパーセプトロンによる非線形主成分分析の有効性を検証する。実験には入力層から順に 3–10–3–10–3 個の素子を配置した砂時計型 5 層パーセプトロンを用いる。したがって、入力層、ボトルネックとなる中間層および出力層の素子数は

$$N = H = M = 3 \tag{3.41}$$

である。また、第 2 層および第 4 層の素子の出力関数はシグモイド関数とし、その他の層の素子は線形とした。データ集合

$$\left\{\boldsymbol{x}_t \in \boldsymbol{R}^3\right\}_{t=1}^{100} \tag{3.42}$$

は $\boldsymbol{R}^3$ の半球面上に分布する点の座標値として生成した。学習回数および学習定数はそれぞれ 15,000 回, 0.05 とし、定数 0.9 の慣性項を付加して重畳エネルギー関数

および従来のエネルギー関数による学習を行う。重畳エネルギー関数における定数 β_i は、3.3節の実験同様に

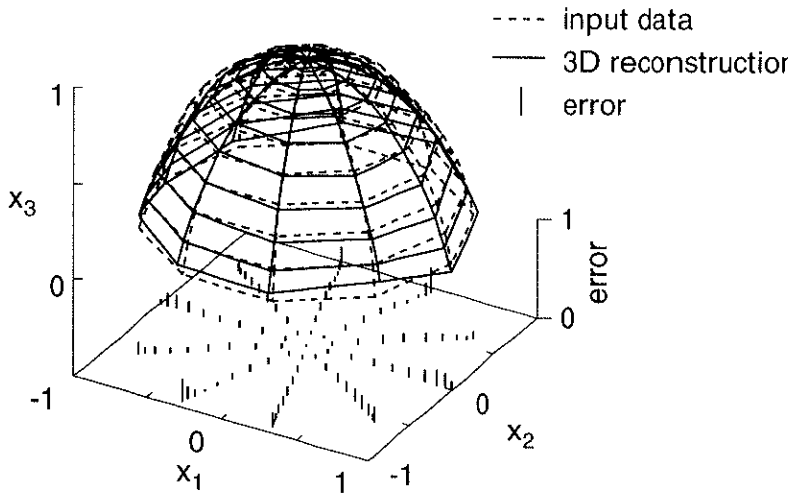
$$\beta_i = \frac{1}{H} \tag{3.43}$$

とした。

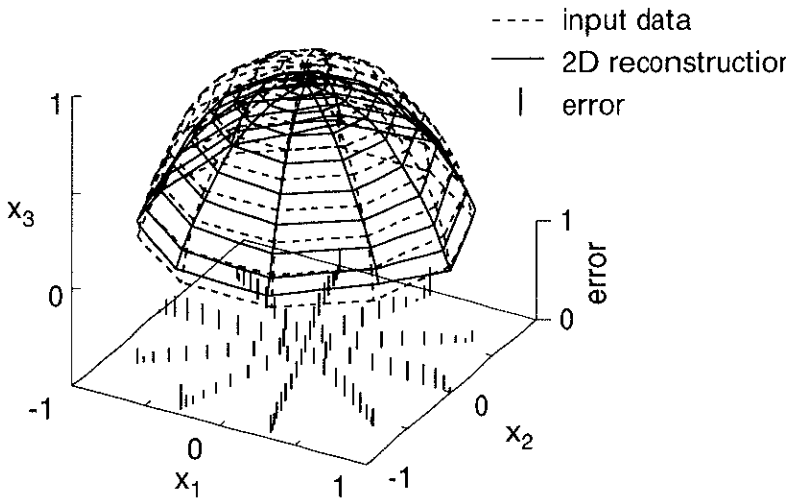
図 3.6は、重畳エネルギー関数を用いて学習したパーセプトロンの出力を示す。下部に描かれた線分は、$\boldsymbol{x}_t$ とそれに対するパーセプトロンの出力 $\tilde{\boldsymbol{x}}_t$ との誤差の大きさ

$$||\boldsymbol{x}_t - \tilde{\boldsymbol{x}}_t|| \tag{3.44}$$

を表す。中間層第 3 素子まで用いる (a) と比較すると、(b) に示す第 2 部分パーセプトロンの出力では半球面の天頂部分で若干誤差が大きくなるが、第 2 素子までの 2 次元の内部表現で半球面の形状がほぼ近似できていることがわかる。次に、従来のエネルギー関数を用いた場合と内部表現の違いを比較するため、3 つの中間層素子のうちそれぞれ 1 つだけを用いて出力を求めた。結果を図 3.7に示す。重畳エネルギー関数を用いた (a) の場合、中間層第 1 および第 2 素子で非線形な曲線座標が構成され、半球面が復元されている。また、第 3 素子だけを用いた出力は変異が小さく、この素子の半球面近似に対する寄与は小さいと考えられる。一方、従来のエネルギー関数を用いた (b) の場合には内部表現は単なる線形直交座標を構成しており、このパーセプトロンは内部表現として何ら有効な特徴量を獲得することができていない。無論このデータ集合の場合、あらかじめ中間層素子数を 2 に設定しておけば従来のエネルギー関数でも半球面の二次元内部表現を獲得することは可能である [49]。しかし、一般に学習を行なう前に最適な内部表現の次元数すなわち中間層素子数を決定することは困難であるから、冗長性の低い内部表現を自己組織的に獲得して最適な次元数を学習後に選択可能な重畳エネルギー関数は有効であると考えられる。

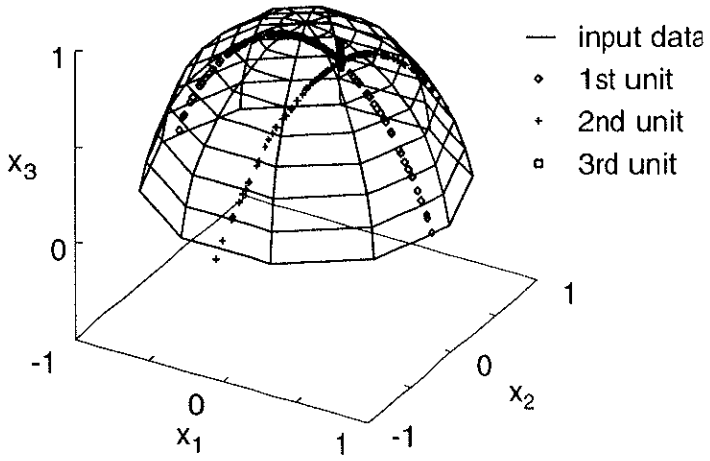


(a) 全ての素子を用いた場合

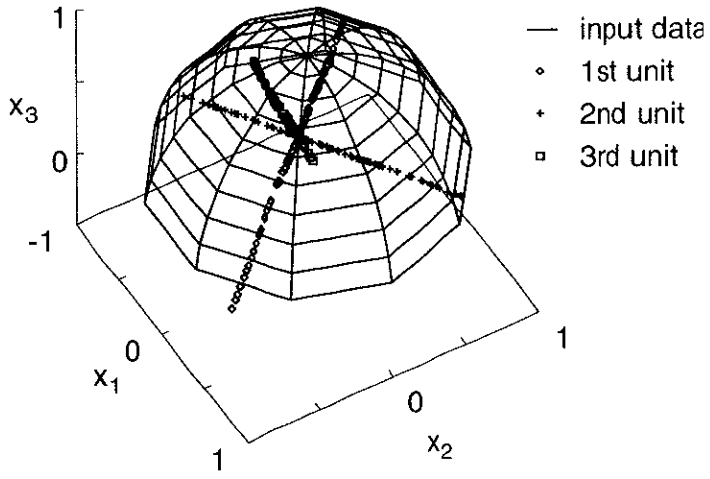


(b) 第 2 部分パーセプトロン

図 3.6: 砂時計型 5 層非線形パーセプトロンによる非線形主成分分析 (その一) : 重畳エネルギー関数を用いて学習したパーセプトロンの出力を示す。破線の格子がデータ集合を、実線の格子がパーセプトロンの出力を表す。下部の線分は誤差の大きさを表す。



(a) 重畳エネルギー関数の場合



(b) 従来のエネルギー関数の場合

図 3.7: 砂時計型 5 層非線形パーセプトロンによる非線形主成分分析 (その二) : 一つの中間層素子出力のみによる出力を示す。実線の格子はデータ集合を表す。

3.4.3 画像情報圧縮

次に、実データを用いた例題として、図3.8のように構成されたシステムによる画像情報圧縮の実験を行なう。対象画像には、SIDBA 標準画像 "Girl" (256 × 256 画素、8bit/画素) を選択した。これを 8 × 8 画素のブロックに分割して 1024 個の 64 次元ベクトルを生成し、主成分分析による次元圧縮を適用して 16 次元のデータ集合

$$\left\{\boldsymbol{x} \in \boldsymbol{R}^{16}\right\}_{t=1}^{1024} \tag{3.45}$$

を構成する。パーセプトロンの素子数は入力層から順に 16–32–H–32–16 (ただし $H = 4$ または 8) とし、第 2 層および第 4 層にはシグモイド型の非線形素子を、その他の層には線形素子を用いる。これらのパーセプトロンに対して重畳エネルギー関数および従来のエネルギー関数を適用して恒等写像学習を行なう。ただし、重畳エネルギー関数の定数β_iについては、

$$\beta_i \propto i^2 \tag{3.46}$$

となるように設定した (4.2.3節参照)。両者ともに同一の条件で学習を行なったのち、得られたパーセプトロンの出力から画像を再構成してその画質を評価する。画質の評価基準は、次式に示す SNR(Signal–to–Noise Ratio) とした。

$$\mathrm{SNR} = 20\log_{10}\frac{255}{\sqrt{\varepsilon^2}} \quad [dB] \tag{3.47}$$

ただしε^2は画素当たりの平均二乗誤差であり、本実験の条件では、部分エネルギー関数の値 E_iから

$$\varepsilon^2 = \frac{E_i}{64} \tag{3.48}$$

として求められる。

図 3.9は、学習後の部分パーセプトロンを用いて再構成した画像の SNR を、線形の主成分分析の結果と共に示している。与えられた中間層素子を全て用いて再構成した場合には、いずれのパーセプトロンも主成分分析を上回る SNR を得ており、従

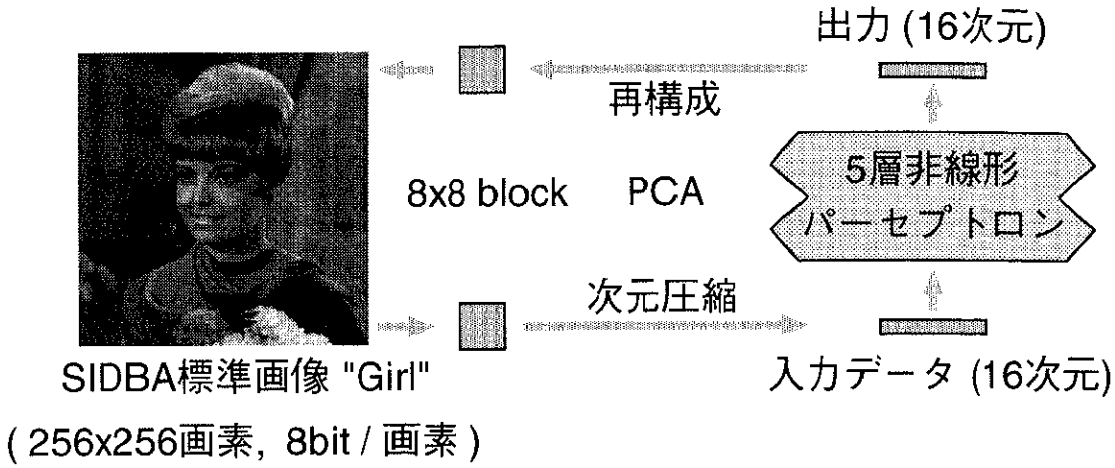
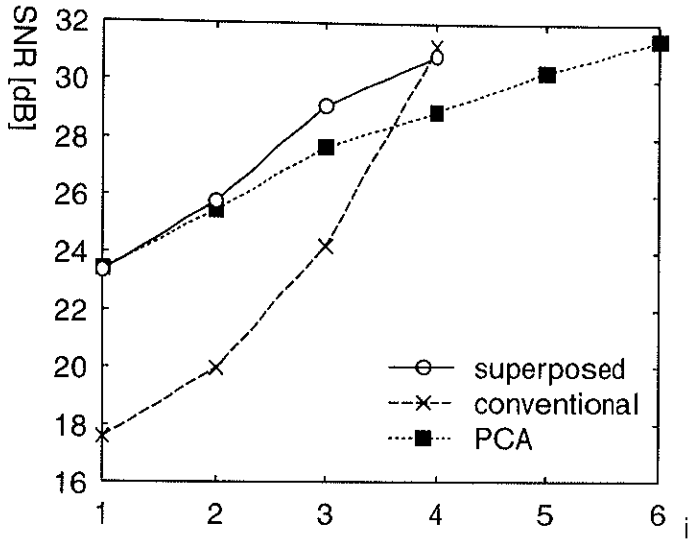
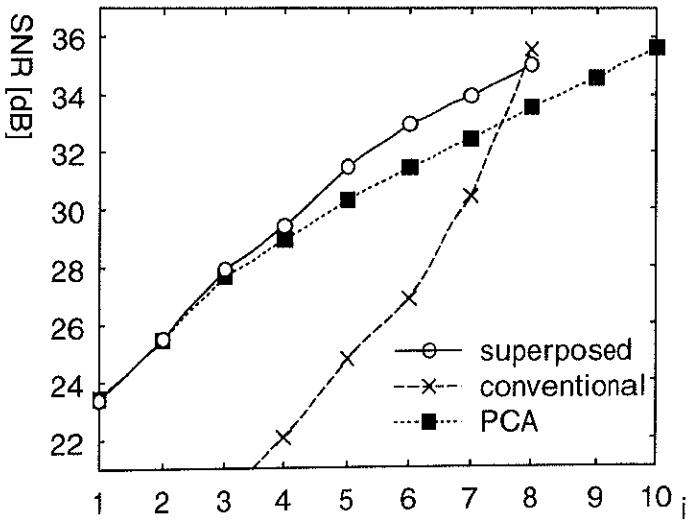


図 3.8: 画像情報圧縮の実験系

来のエネルギー関数を用いた方が重畳エネルギー関数よりも良い値を示している。しかし、従来のエネルギー関数の場合には素子を一つでも削除すると SNR が大幅に低下するのに対して、重畳エネルギー関数の場合には全ての部分パーセプトロンで主成分分析を上回る SNR が得られている。重畳エネルギー関数によって自己組織された内部表現では、必要に応じて任意の次元数の特徴ベクトルを取り出すことが可能となっていることがわかる。



(a) $H = 4$



(b) $H = 8$

図 3.9: 画像情報圧縮の実験結果：学習後の部分パーセプトロンを用いて再構成された画像の SNR を示す。実線は重畳エネルギー関数、破線は従来のエネルギー関数、点線は主成分分析の結果を表す。

3.5 議論

本章では、多層パーセプトロンによる次元圧縮において寄与順に並んだ内部表現の自己組織を可能とする方法として、重畳エネルギー関数に基づく恒等写像学習を提案し、その学習特性を理論的・実験的に検討した。その結果、重畳エネルギー関数を用いた 3 層線形パーセプトロンの恒等写像学習では、学習後のパーセプトロンの中間層素子がそれぞれ寄与の大きなものから順にデータの主成分を出力するようになることが証明された。また、5 層非線形パーセプトロンの恒等写像学習に適用した場合には、内部表現にデータの非線形な特徴を寄与順に抽出し、非線形主成分分析を実現できることを示した。

パーセプトロンの恒等写像学習で寄与順の主成分抽出を可能にする方法としては、本研究で検討した方法の他に、中間層素子数 1 個の 3 層パーセプトロンを多段に接続する形式のもの [46, 47]、入力層に N個、中間層に H個、出力層に $N \times H$個の素子を配置した 3 層パーセプトロンを用いるもの [48] が存在する。しかし、本研究で検討した 3 層砂時計型パーセプトロンと比較すると、これらはネットワークの構成により多くの素子や重みを必要とするため、近似出力を求める際に必要な計算量が多くなるという欠点を有している。また、これらはいずれも線形素子を用いることを前提としており、ネットワークの構造や重みに制約を加えることによって寄与順の主成分抽出を実現している。そのため、非線形主成分分析に適用できないという点も大きな問題である。これに対して本研究では、ネットワーク構造を制約するかわりにエネルギー関数に変更を加える方法を採用したため、任意の構造の多層パーセプトロンに適用可能な学習アルゴリズムを構成することができた。これによって、内部表現が寄与順に並んだ非線形主成分分析を実現している。しかし、主成分分析との一致が証明されている 3 層線形パーセプトロンの場合と違い、非線形パーセプトロンの恒等写像学習における内部表現の特性に関しては明確になっていない。したがって、重畳エネルギー関数を用いる非線形主成分分析に関しては、今後理論的な解析を行なう必要がある。

表 3.2: 重畳エネルギー関数と従来のエネルギー関数における計算量の比較

	従来型	重畳型
前進計算	1600 (回)	5152 (回)
後退計算	2624	11808
計	4224	16960
比率	1	4.02

一方、重畳エネルギー関数の応用上の問題点として、従来のエネルギー関数を用いるよりも学習に多くの計算量を要することが上げられる。表3.2は、3.4.3節の実験に用いたパーセプトロン ($H = 8$) において、一つのデータに対する一回の学習当たり必要な乗除算回数をパーセプトロン出力の計算 (前進計算) と誤差の逆伝搬および重みの修正 (後退計算) の二つに分けて算出したものである。従来のエネルギー関数では一つの誤差 E_Hのみを評価して学習を行なえば良いところを、重畳エネルギー関数では E_1から E_Hまでの全ての部分エネルギー関数を評価して学習を行なう必要があるため、計算量がおよそ $H/2$倍となっている。この計算量の増加は、中間層素子数の多い大規模なパーセプトロンを用いる場合には無視できなくなると考えられる。したがって、より少ない計算量で寄与順の成分抽出を実現する改良方式の検討も今後の課題である。

第 4 章

多層パーセプトロンの学習における内部表現の冗長性削減

4.1 導入

本章では、第3章で多層パーセプトロンによる恒等写像学習を対象に検討した重畳エネルギー関数を、一般の多層パーセプトロンの学習に応用する。パーセプトロンの学習理論においては、十分な数のシグモイド型非線形素子を含む4層パーセプトロンが任意の関数を近似可能であること [50]、および3層パーセプトロンが任意の連続関数を近似可能 [36, 37, 38] であることが知られている。これらの結果は、学習に用いるデータの近似に関しては、中間層の素子数を増やせばいくらでもパーセプトロンの近似能力を向上できることを示している。しかし実際には、計算コストの点から中間層素子数を任意に増やすことは不可能である。さらに、必要以上の素子を与えた場合には、中間層素子の冗長性によって

- 未学習データに対する近似精度 (汎化能力) が低下する
- 獲得した内部表現やネットワーク構造の解析が困難になる

といった問題が生じることが指摘されている [51, 52]。したがって応用の際には、中間層素子の数や重みの数などを調節して、与えられた学習課題に適した構造のパーセプトロンを用いることが重要となる。そのため、冗長性の低いパーセプトロンを構成する方法がこれまで数多く検討されてきた [51]。その方法は、以下のように大きく三つに分類することができる。

1. 素子数や重み数などが異なる複数のパーセプトロンで学習を行ない、それらの中から一定の基準にしたがって最適な構造のものを選択する方法 [53, 54, 55]

2. 素子数や重み数の少ないパーセプトロンで学習を始め、必要に応じて素子・重みを追加していく方法 [56, 57, 58, 59]

3. 十分な数の素子・重みを有するパーセプトロンの構造を変化させて冗長性を削減する方法

   (a) 学習中あるいは学習後に一定の基準にしたがって不要な素子と重みを削除する方式 [60, 61]

   (b) 自由度に関する制約項を付加したエネルギー関数によって冗長な結合を自律的に減弱させる方式 [52, 62, 63]

一方、重畳エネルギー関数は、恒等写像学習において主成分分析同様に寄与順に並んだ内部表現を自己組織するためのエネルギー関数として検討されているが、特にパーセプトロンの構造に関する制約はなく、任意の多層パーセプトロンの学習に適用可能である。この場合、重畳エネルギー関数を用いる学習は上述の 3.(b) の方式の一つと考えられ、寄与順に並んだ内部表現を自己組織することによってパーセプトロンの冗長性を自律的に削減する効果が期待できる。そこで本章では、重畳エネルギー関数を用いた学習による内部表現の冗長性削減効果を検討する。はじめに 4.2 節において重畳エネルギー関数に基づく学習則の導出過程を説明する。次に 4.3節で一変数の関数近似および二次元のパターン分類の例題を対象として数値実験を行

ない、重畳エネルギー関数に基づく学習によって冗長性の低い内部表現を持つパーセプトロンが構成できることを示す。最後に 4.4節でその有効性を議論する。

4.2 重畳エネルギー関数に基づく学習則の導出

非線形素子を含む3層パーセプトロンに対して重畳エネルギー関数を定義し、これに基づく誤差逆伝搬学習則を導出する。以下の議論では3層パーセプトロンを仮定するが、より多層のパーセプトロンに関しても容易に拡張可能である。

4.2.1 重畳エネルギー関数の定義

定義 4.1 部分パーセプトロン

入力層素子 N個、中間層素子 H個、出力層素子 M個から成る 3 層パーセプトロンにおいて、入力

$$\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^T \tag{4.1}$$

が与えられたときの中間層素子の出力を

$$\boldsymbol{u} = [u_1, u_2, \ldots, u_H]^T \tag{4.2}$$

とおく。ただし、 $u_h (h = 1, 2, \ldots, H)$ は

$$u_h = \sigma\left(\sum_{n=1}^{N} v_{h,n} x_n + v_{h,0}\right) \tag{4.3}$$

で与えられる。 $v_{h,n}$ は中間層第 h 素子と入力層第 n 素子を結合する重みの値、 $v_{h,0}$ はしきい値を表し、 $\sigma(\cdot)$ は中間層素子の出力関数である。通常のパーセプトロンでは u_1 から u_H まで全ての中間層素子出力を用いるが、 u_1 から u_i までのみを考慮し、残りの素子の出力 $u_{i+1}, \ldots, u_H$ を 0 に置き換えて得られるパーセプトロンを第 i 部分パーセプトロンと呼ぶ (図 4.1参照)。

入力 $\boldsymbol{x}$ に対する第 i 部分パーセプトロン $(i = 1, 2, \ldots, H)$ の出力を

$$f_i(\boldsymbol{x}) = {}^i\boldsymbol{y} = [{}^iy_1, {}^iy_2, \ldots, {}^iy_M]^T \tag{4.4}$$

と表すと、 ${}^iy_m$ $(m = 1, 2, \ldots, M)$ は

$${}^iy_m = \rho\left(\sum_{h=1}^{i} w_{m,h} u_h + w_{m,0}\right) \tag{4.5}$$

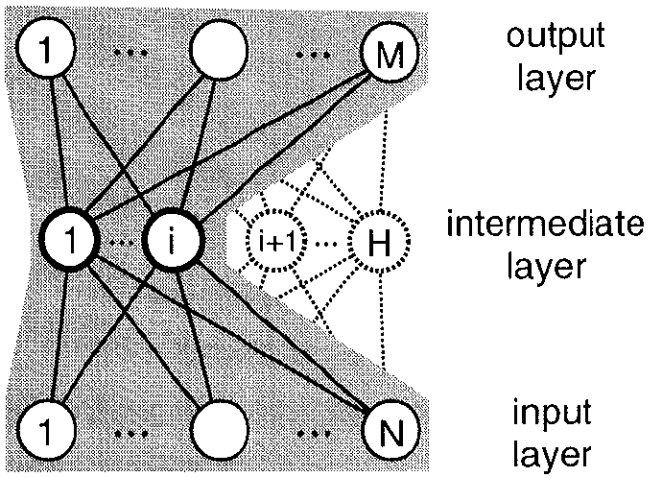


図 4.1: パーセプトロンと部分パーセプトロン

で与えられる。ただし、$w_{m,h}$ は出力層第 m 素子と中間層第 h 素子を結合する重みの値、$w_{m,0}$ はしきい値を表し、$\rho(\cdot)$ は出力層素子の出力関数である。通常のパーセプトロンにおける出力の計算と異なるのは、総和記号の上の数が H から i になっている点のみである。第 H 部分パーセプトロンの出力 $f_H(\boldsymbol{x})$ は全ての中間層素子を用いる通常のパーセプトロンの出力に一致する。

定義 4.2 部分エネルギー関数

上記のパーセプトロンに対して、教師データを T 個の入出力データ対

$$\{\boldsymbol{x}_t \in \boldsymbol{R}^N, \boldsymbol{y}_t \in \boldsymbol{R}^M\}_{t=1}^T \tag{4.6}$$

で与える。このとき、誤差関数

$$E_i = \frac{1}{2T}\sum_{t=1}^{T} ||\boldsymbol{y}_t - f_i(\boldsymbol{x}_t)||^2 \tag{4.7}$$

を第 i 部分エネルギー関数と呼ぶ。

第 i 部分エネルギー関数の値は、パーセプトロン全体が目標出力値を近似する上で第 i 部分パーセプトロンがどれだけ貢献しているかを示す指標と考えることがで

きる。例えば、E_H の値が十分収束した時点で、ある素子数 I ($1 \leq I \leq H$) に対して

$$E_{i \leq I} \approx E_H \tag{4.8}$$

が成り立つならば、$I+1$ 番目以降の中間層素子は冗長であり、その学習課題に対しては素子数は I 個で十分であると判断することができる。しかし、従来のパーセプトロン学習では E_H のみが最小化されるため、このような方法で中間層素子の冗長性を判定することは困難である。そこで、すべての部分エネルギー関数を同時に最小化するため、以下のようなエネルギー関数を検討する。

定義 4.3 重畳エネルギー関数

部分エネルギー関数 E_i の重み付き和

$$F = \sum_{i=1}^{H} \beta_i E_i \tag{4.9}$$

を重畳エネルギー関数と呼ぶ。β_i は正の定数である。

重畳エネルギー関数を用いた学習では、第 i 中間層素子に関与する重みの修正量は $E_i, \ldots, E_H$ より計算されるが、第 $i-1$ 素子の重みについては、これらに加えて i 番目以下の素子を用いないで得られる誤差 E_{i-1} も減少させるように修正される。したがって、番号の小さな素子に関する重みほど少ない素子数で目標出力を近似するように学習が進行する。そのため、上位の素子ほど全体の誤差 E_H の減少に対して大きく寄与し、内部表現には素子の番号順に重要な特徴が抽出されると期待できる。

4.2.2 学習則の導出

式 (4.9) の重畳エネルギー関数に基づく誤差逆伝搬学習則を導出する。第 h 中間層素子と出力層素子を結合する重み $w_{m,h}$ の修正量 $\Delta w_{m,h}$ の場合、F に関する最急降下法より直ちに

$$\Delta w_{m,h} = -\eta \frac{\partial F}{\partial w_{m,h}} = -\eta \sum_{i=1}^{H} \beta_i \frac{\partial E_i}{\partial w_{m,h}} \tag{4.10}$$

が導かれる。ただし η は学習定数である。ここで、$i < h$ のとき

$$\frac{\partial E_i}{\partial w_{m,h}} = 0 \tag{4.11}$$

であるから、式 (4.10) は

$$\Delta w_{m,h} = -\eta \sum_{i=h}^{H} \beta_i \frac{\partial E_i}{\partial w_{m,h}} \tag{4.12}$$

となる。したがって、$\Delta w_{m,h}$ は対応する h 番目から H 番目までの各部分パーセプトロンに通常の誤差逆伝搬学習則を適用したときの重み修正量の和で与えられる。入力層側の重み $v_{h,n}$およびしきい値 $v_{h,0}$ についても同様である。一方、しきい値 $w_{m,0}$ はすべての部分パーセプトロンに関与するため、その学習則は

$$\Delta w_{m,0} = -\eta \sum_{i=1}^{H} \beta_i \frac{\partial E_i}{\partial w_{m,0}} \tag{4.13}$$

で与えられる。

4.2.3 係数 β_i の選択

3章で検討したように、線形パーセプトロンの恒等写像学習に重畳エネルギー関数を適用する場合、理論上は正定数 β_i を任意に設定することができる。しかし、以下に示すように学習効率が係数 β_i に依存して変化するため、実際に一般の学習に適用する際には、適切な値を選択する必要がある。β_i は 一般性を失わずに $\sum_{i=1}^{H} \beta_i = 1$ とできることから、以降ではこれを仮定して議論する。

$w_{m,h}$を例に、すべての部分エネルギー関数に関するこの重みの修正量がほぼ等しい場合を考慮する。すなわち、

$$\frac{\partial E_h}{\partial w_{m,h}} \approx \frac{\partial E_{h+1}}{\partial w_{m,h}} \approx \cdots \approx \frac{\partial E_H}{\partial w_{m,h}} \tag{4.14}$$

を仮定する (ただし $h = 1, 2, \ldots, H$である)。このとき、式 (4.12) より

$$\Delta w_{m,h} \approx -\eta \left(\sum_{i=h}^{H} \beta_i \right) \frac{\partial E_H}{\partial w_{m,h}} \tag{4.15}$$

であるから、$\Delta w_{m,1}$ と $\Delta w_{m,h}$ を比較すると、$w_{m,h}$ の更新に関する学習定数は、実質的に $w_{m,1}$ のそれの$\sum_{i=h}^{H}\beta_i$ $(\leq \sum_{i=1}^{H}\beta_i = 1)$ 倍となる。他の重みについても同様の議論が成り立つ。ゆえに、Hが大きい場合に例えば $\beta_i = \frac{1}{H}$ とすると、下位の中間層素子の重み修正量が上位のものに比べて非常に小さくなる。その結果、学習が停滞して各部分エネルギー関数の間の差が大きくならず、各素子の寄与が偏りにくくなる可能性がある。これを回避するためには、$\{\beta_i\}_{i=1}^{H}$ が i に対する単調増加数列となるように設定し、実質的な学習定数の差を軽減する方法が考えられる。

以上の考察に基づき、以降の数値実験では経験的に β_i を

$$\beta_i = \frac{i^2}{\sum_{j=1}^{H} j^2} = \frac{6i^2}{H(H+1)(2H+1)} \tag{4.16}$$

と設定した。

4.3 数値実験による学習特性の検証

重畳エネルギー関数の冗長性削減効果を検証するためにいくつかの数値実験を行う。

4.3.1 一変数関数近似問題

次式の連続関数 [64] を学習対象とする関数近似問題を取り上げる。

$$g(x) = \frac{(x-2)(2x+1)}{x^2+1} \tag{4.17}$$

入力データ x_t は区間 $[-8, 8]$ において刻み幅 0.3 で等間隔に選択し、教師データ y_t は $g(x_t)$ に平均 0 、分散 0.1 のガウシアンノイズを加えたものとする (図 4.2参照)。実験に用いる 3 層パーセプトロンは中間層素子数 H を多くとり、

$$N = 1, \quad H = 12, \quad M = 1 \tag{4.18}$$

とした。また、中間層の出力関数は

$$\sigma(x) = \tanh(x) \tag{4.19}$$

とし、出力層の出力関数は線形すなわち $w_{m,0} \equiv 0$ $(m = 1)$ のもとで

$$\rho(x) = x \tag{4.20}$$

とした。その他の実験条件を表 4.1に示す。学習は重畳エネルギー関数および従来のエネルギー関数それぞれについて重みの初期値を変えながら 10 回ずつ行なった。

重畳エネルギー関数を用いて構成された 10 個のパーセプトロンの部分エネルギー関数 E_i の値を図 4.3(a) に示す。部分エネルギー関数の値は中間層素子数を増すにつれて減少しており、初期値に依存することなく $i = 7$ すなわち第 7 部分パーセプトロンで収束している。したがって、この例題においては中間層素子数は 7 個で十分であり、8 から 12 番目の中間層素子は冗長であると推定できる。

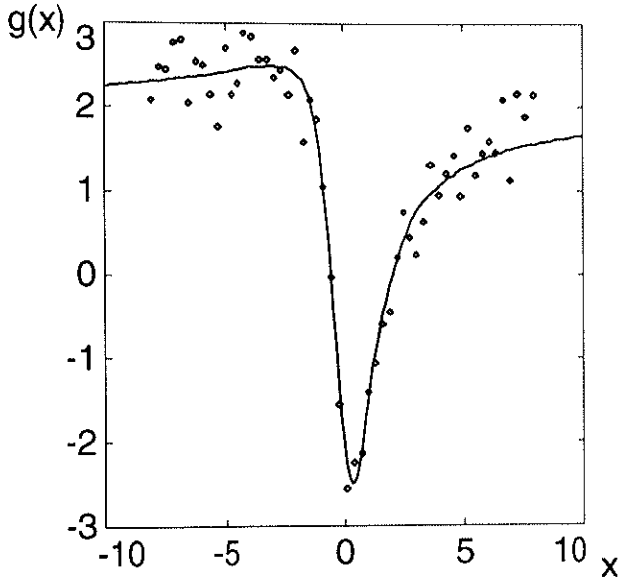


図 4.2: 関数近似問題の実験に用いた関数と教師データ

表 4.1: 関数近似問題の実験条件

重みの初期値	$[-0.1, 0.1]$ の一様乱数
学習定数	0.01
慣性定数	0
学習回数	50,000 epoch

次に、この結果が重畳エネルギー関数に固有のものであることを示すため、従来のエネルギー関数に基づく結果と比較する。ただし、従来のエネルギー関数では中間層素子は全く順位付けされないため、学習後のパーセプトロンに対して以下の操作を適用して寄与の順に素子の並べかえを行なった。

1. $i = H$ とする。

2. 1 番目から i 番目の中間層素子のうち、その素子を削除して誤差を計算したときに E_i に対する誤差の増加が最も少ないものを選択して、その素子を新たに i 番目とする。

3. $i = 2$ ならば終了、さもなくば i を一つ減らして 2. へ。

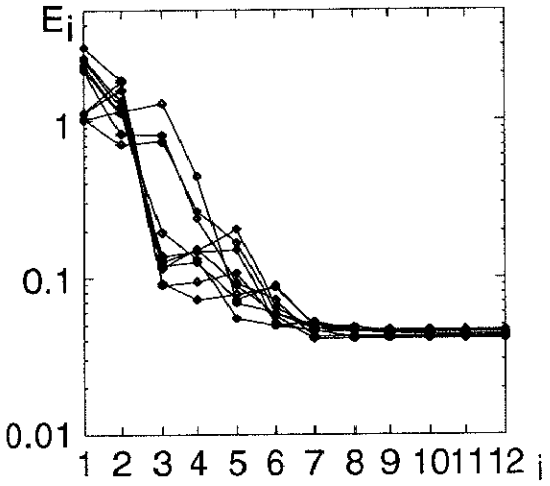
図 4.3(b) は、この操作を行なった後の部分エネルギー関数 E_i の値を示している。中間層素子数の増加にともなって E_i が減少する傾向は見られるが、どの部分パーセプトロンにおいても誤差の収束は確認できない。したがって、これらのパーセプトロンの内部表現は冗長な分散表現となっていると考えられる。

図 4.4は、二つのエネルギー関数で獲得される内部表現の違いを表す。破線は第 i 部分パーセプトロンの出力 $f_i(x)$ を表している。重畳エネルギー関数を用いたパーセプトロンでは上位の少数の中間層素子を利用して概形が表現され、素子の追加にともなって徐々に細部が近似されるようになっているが、従来のエネルギー関数を用いたパーセプトロンでは部分パーセプトロンの近似精度が悪く、それぞれの中間層素子の役割が不明確である。

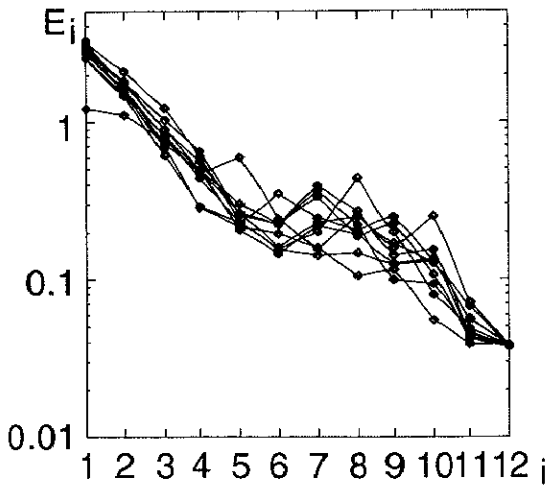
4.3.2 二次元パターン分類問題

2 次元のらせん上の点にノイズを加えて生成される教師データ (図 4.5参照) を用いた 2 クラスのパターン分類問題を対象として実験を行なう。パーセプトロンの構成は

$$N = 2, \quad H = 10, \quad M = 2 \tag{4.21}$$

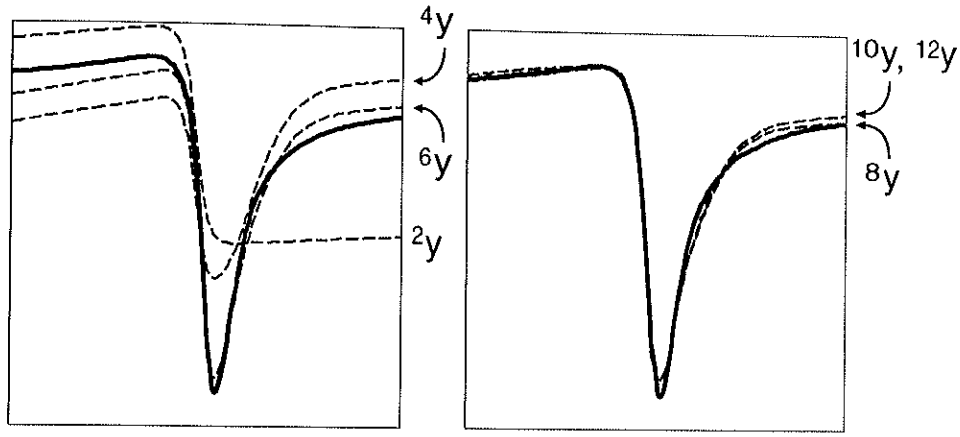


(a) 重畳エネルギー関数

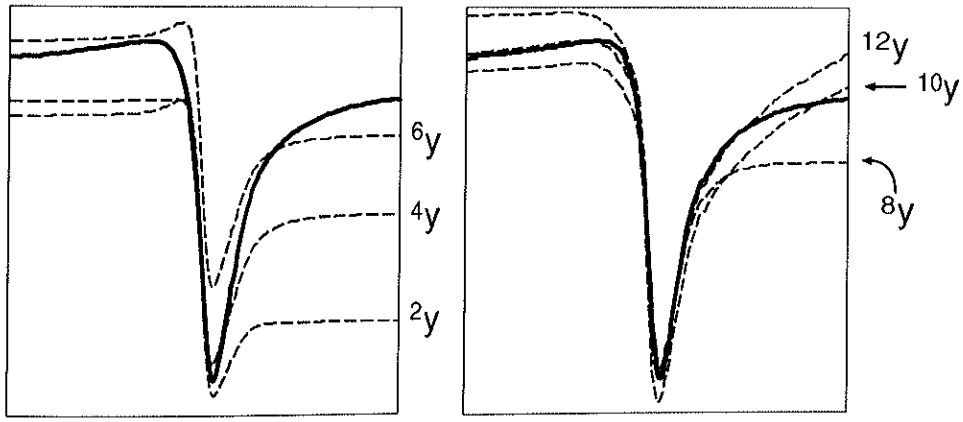


(b) 従来のエネルギー関数

図 4.3: 関数近似問題における部分エネルギー関数の値：横軸は中間層素子数 i を、縦軸は第 i 部分エネルギー関数 E_i の値を表す



(a) 重畳エネルギー関数



(b) 従来のエネルギー関数

図 4.4: 関数近似問題における部分パーセプトロンの出力 : 実線は真の関数曲線を、破線は第 i 部分パーセプトロンの出力 $f_i(x)$ $(i = 2, 4, 6, 8, 10, 12)$ を表す

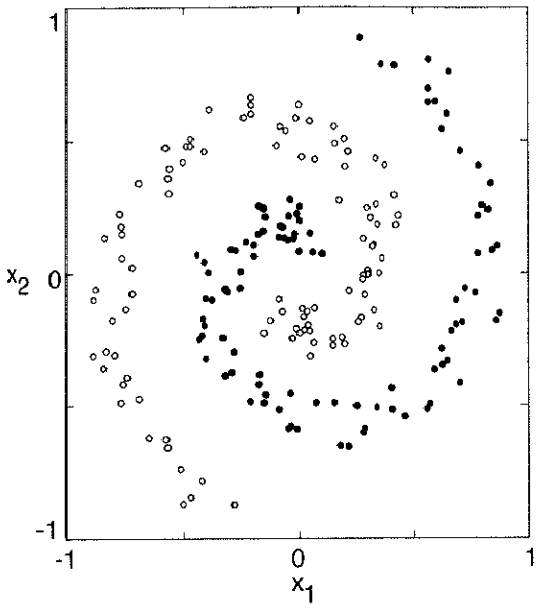


図 4.5: パターン分類問題の教師データ：白丸がクラス 1、黒丸がクラス 2 のデータを表す。

とし、中間層および出力層の出力関数は

$$\sigma(x) = \rho(x) = \frac{1}{1 + \exp(-x)} \tag{4.22}$$

とした。2 つの出力素子をクラス 1 およびクラス 2 のそれぞれに対応させ、出力の教師データは、そのデータのクラスに対応する出力素子に対して 0.9 を、残りの素子に対して 0.1 を割り当てた。その他の実験条件を表 4.2に示す。学習は重畳エネルギー関数および従来のエネルギー関数それぞれについて重みの初期値を変えながら 10 回ずつ行い、教師データとは別のテストデータに対する各部分パーセプトロンの認識率を求める。テストデータは各点の座標値に加えるノイズの値を変更して生成した。また、二つの出力素子のうちより大きな値を出力した方が示すクラスをパーセプトロンの選択したクラスとした。認識率は、パーセプトロンの選択したクラスが正しいクラスに一致した割合として算出した。

表 4.2: パターン分類問題の実験条件

	重畳型	従来型
重みの初期値	$[-0.1, 0.1]$ の一様乱数	
学習定数	0.5	
慣性定数	0.9	
学習回数	50,000 epoch	100,000 epoch

重畳エネルギー関数で学習したパーセプトロンの認識率を図4.6(a) に示す。認識率は 7 番目の中間層素子でほぼ収束しており、8 から 10 番目までの中間層素子は冗長なものと判断できる。一方、図4.6(b) は従来のエネルギー関数の結果を示す。学習後に 4.3.1節の実験と同様に中間層素子の並べかえを行なっている。ただし、50,000 回の反復では学習が停滞したままで十分な性能のパーセプトロンを構成できなかったため、学習回数を 100,000 回とした。しかし、10 個のうち 3 つのパーセプトロンでは 100,000 回の反復でも十分な認識率が得られなかった。それ以外のパーセプトロンについても認識率の収束は見られず、内部表現が冗長になっていると考えられる。

図 4.7は、重畳エネルギー関数および従来のエネルギー関数で構成されたパーセプトロンについて、部分パーセプトロンを用いて入力平面の領域分割を行った結果を示す。図中の実線は

$$ {}^{i}y_1 = {}^{i}y_2 \tag{4.23} $$

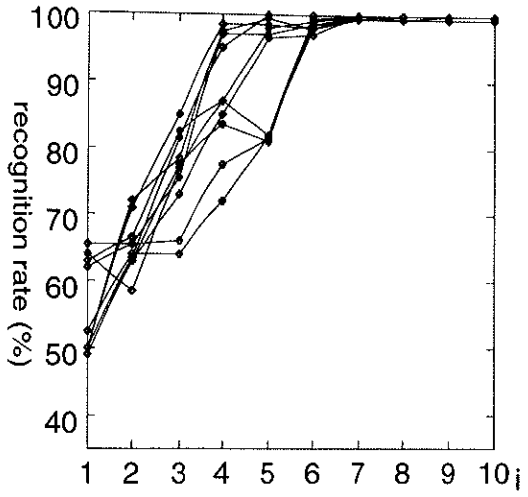
を満たす 2 つのクラスの境界を示す[1]。また、以下の条件にしたがって領域を分割し、塗分けを行った。

白色 : ${}^{i}y_1 > {}^{i}y_2$ かつ ${}^{i}y_1 \geq 0.75$

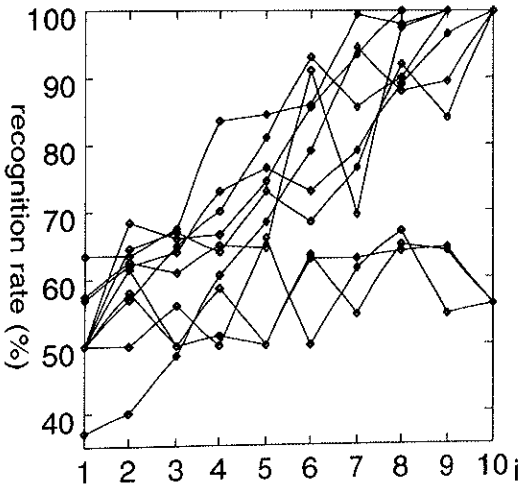
濃色 : ${}^{i}y_1 < {}^{i}y_2$ かつ ${}^{i}y_2 \geq 0.75$

淡色 : 上記以外

[1]厳密には、式 (4.23) を満たす入力 (x_1, x_2) の集合は領域となる。しかし本実験の場合、図に示すスケールではそのような領域を確認できないため、境界領域を曲線として扱う。



(a) 重畳エネルギー関数



(b) 従来のエネルギー関数

図 4.6: パターン分類問題における部分パーセプトロンの認識率：横軸は中間層素子数 i を、縦軸は第 i 部分パーセプトロンの認識率を表す

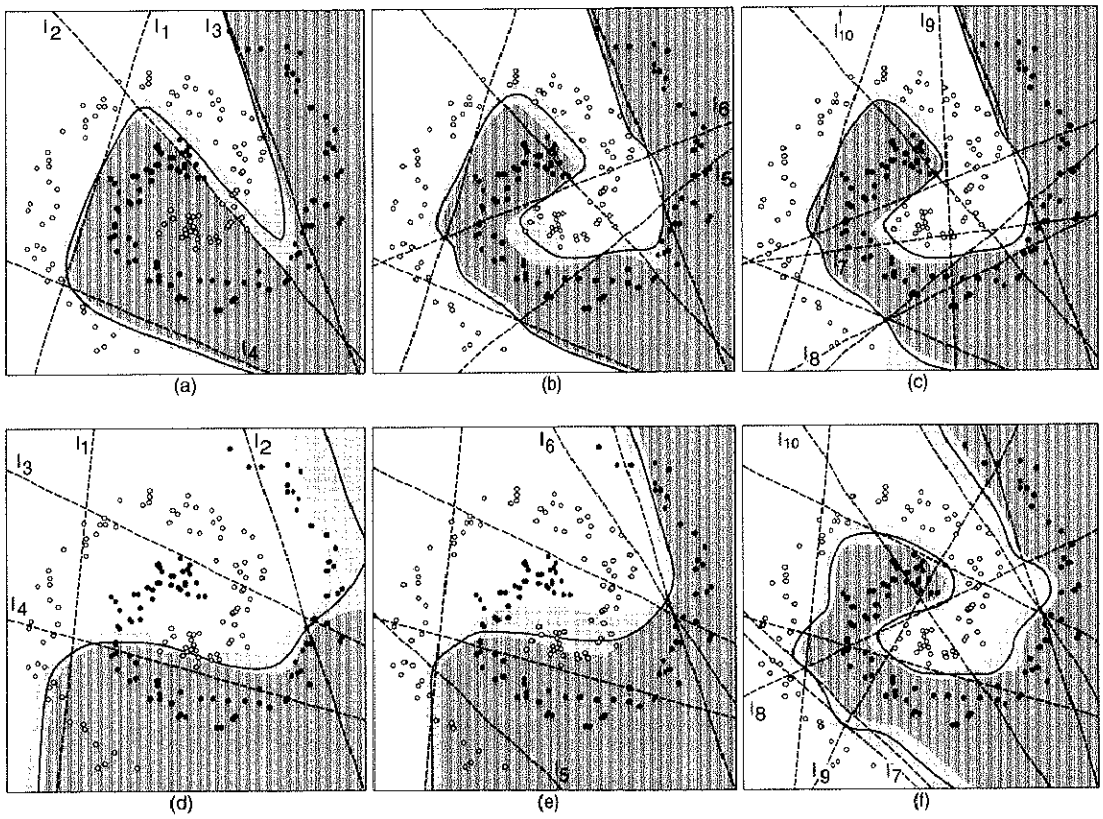


図 4.7: 部分パーセプトロンによる入力平面の領域分割 : ((a)–(c) : 重畳エネルギー関数、(d)–(f) : 従来のエネルギー関数、(a),(d) $i = 4$、(b),(e) $i = 6$、(c),(f) $i = 10$ の部分パーセプトロンに対応。実線 : 分類境界、破線 : 識別線 l_h)

図中の破線は以下の式で定義される各中間層素子の識別線を表している。

$$\begin{aligned} l_h &= \{(x_1, x_2)|u_h = 0.5\} \\ &= \left\{(x_1, x_2)\middle|\sigma\left(\sum_{n=1}^{2} v_{h,n}x_n + v_{h,0}\right) = 0.5\right\} \\ &= \{(x_1, x_2)|v_{h,1}x_1 + v_{h,2}x_2 + v_{h,0} = 0\} \end{aligned} \tag{4.24}$$

重畳エネルギー関数の場合、上位 4 つの中間層素子で大まかな分類が行なわれ (図 4.7(a))、さらに 2 素子の追加で主に中心領域の分類が修正されて概形が完成し (同 (b))、残りの素子で細部の修正が行なわれている (同 (c))。一方従来のエネルギー関数の場合、上位の中間層素子のみで構成された部分パーセプトロンでは概形は得られず (図 4.7(d),(e))、各中間層素子の役割が明確にならない。また、l_5 と l_7 の識別線 (同 (e),(f)) がほぼ平行であり、明らかに冗長な素子が存在することが理解される。

4.3.3 汎化能力の検討

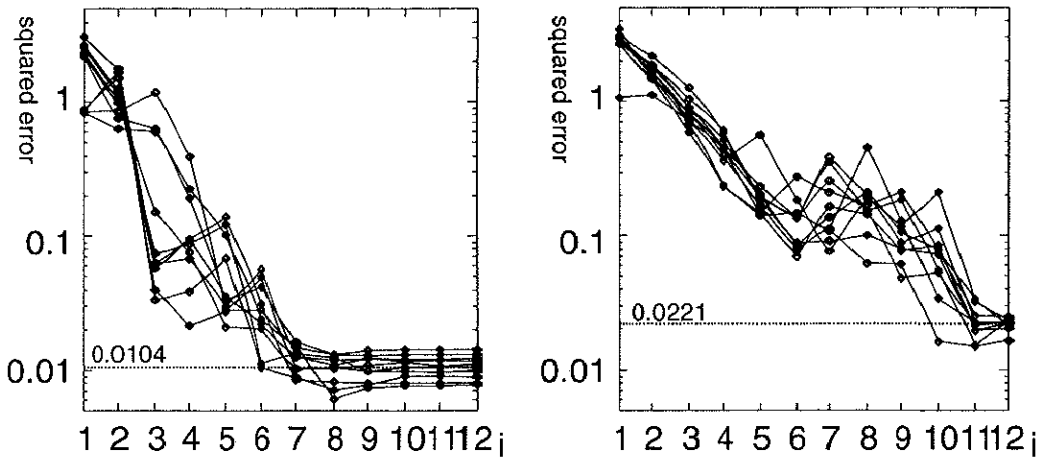
重畳エネルギー関数に基づく学習では、冗長性の低い内部表現が自己組織されるため、単純に平均二乗誤差をエネルギー関数とする場合と比較して汎化能力が改善される可能性がある。これを調べるため、4.3.1 節の実験で構成したパーセプトロンを用い、真の関数値に対する誤差を指標として汎化能力を評価する。なお、テストデータの入力値は区間 $[-10, 10]$ においてきざみ幅 0.1 で等間隔に選択した。

重畳エネルギー関数および従来のエネルギー関数により構成されたパーセプトロンの誤差を図 4.8(a) および (b) にそれぞれ示す。重畳エネルギー関数の場合、教師データに対する誤差と同様にテストデータに対する誤差も 7 個程度の中間層素子数で収束している。また、10 個の第 12 部分パーセプトロンの平均誤差 (図中破線) は 0.0104 となっており、通常のエネルギー関数の平均 0.0221 の半分程度となっている。したがって、重畳エネルギー関数を用いたことによって内部表現の冗長性削減と共に汎化能力改善の効果が得られている。

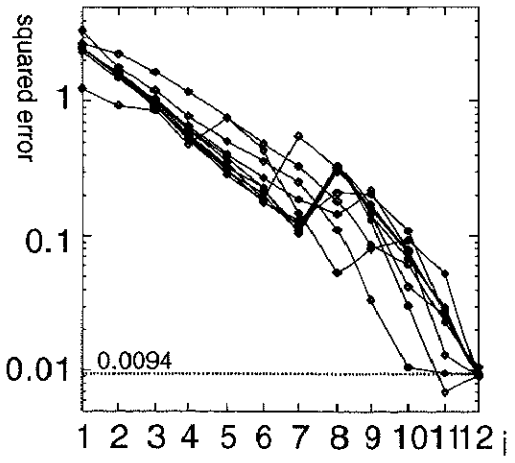
この結果を他の汎化能力改善手法と比較するため、重みの忘却項 (正則化項) を付加したエネルギー関数 [55, 65]

$$G = E_H + \frac{\alpha}{2}\left(\sum_{h=1}^{H}\sum_{n=0}^{N} v_{h,n}^2 + \sum_{m=1}^{M}\sum_{h=1}^{H} w_{m,h}^2\right) \tag{4.25}$$

を用いて追実験を行う。図 4.8(c) は、4.3.1節の実験と同様の条件で 10 個のパーセプトロンを構成し、中間層素子の置換後にテストデータに対して計算した誤差の値を示している。この場合、係数 α について様々な値を用いて試行錯誤を行い、テストデータに対する誤差が最小となる値 $\alpha = 0.32$ を選択した。第 12 部分パーセプトロンの平均誤差は重畳エネルギー関数よりも若干小さくなっている。しかし、図 4.8(b) と同 (c) を比較すると、忘却項を付加した場合には従来のエネルギー関数よりも中間層素子の利用に偏りが少なくなっており、分散表現が強まっているとみられる。したがって、重みの忘却項を併用する方法では汎化能力は改善されるものの、内部表現の冗長性削減効果は得られない。



(a) 重畳エネルギー関数 (b) 従来のエネルギー関数



(c) 従来のエネルギー関数+重みの忘却項

図 4.8: テストデータに対する誤差 : 横軸は中間層素子数 i を、縦軸は第 i 部分パーセプトロンの誤差を表す。破線は 10 個の第 12 部分パーセプトロンの平均値を示す。

4.4 議論

本章では、重畳エネルギー関数に基づく学習によって内部表現の冗長性の低いパーセプトロンが自己組織されることを示した。パーセプトロンの獲得する内部表現は、重要な特徴量が少数の中間層素子に集中して寄与順に並んだ明確なものとなる。そのため、各素子についてそれぞれの貢献度の評価が可能であり、パーセプトロンの解析が容易となる。さらに、内部表現の冗長性削減の効果によって汎化能力の優れたパーセプトロンが構成される可能性を指摘した。

重畳エネルギー関数を用いて構成されたパーセプトロンの場合、学習後に用いる中間層素子数を任意に選択することができる。本章の実験では、単純に部分エネルギー関数の値から必要素子数を推定したが、素子毎の寄与の大きさの評価基準も含めて、必要素子数の推定法に関しては今後より詳細に検討する必要がある。また、中間層素子数の多い条件では、β_iの選択が重畳エネルギー関数の学習性能や獲得される内部表現の性質に影響を与えると考えられる。この点に関する考察も今後の課題となる。他方、汎化能力の向上については、本章では一つの例題に対して実験的考察を行なったにすぎず、統計的手法 [53, 54, 55] との比較など、より厳密な考察が必要である。したがって、重畳エネルギー関数の学習性能に関する理論的検討も今後の重要な課題である。また、重畳エネルギー関数を用いた冗長性削減方式は、エネルギー関数の変更によるものであるから、従来提案されている冗長性削減手法と併用することが可能である。この観点から、より性能の高いパーセプトロンの構築法を検討することも必要であろう。

第 5 章

結論

本論文では、次元圧縮を行なうニューラルネットワークの学習の理論と応用について検討した。第 2 章では、Hebb 型の教師なし学習の応用として、視覚系時空間受容野の自己組織形成モデルを構成した。第 3 章および第 4 章では、多層パーセプトロンの誤差逆伝搬学習で寄与順に並んだ内部表現を自己組織することのできるエネルギー関数について検討した。本論文の論点は以下の通りである。

- 時空間受容野の自己組織モデル (第 2 章)

  時間特性の異なる複数の信号伝達経路をもつニューロンモデルを考え、これを構成要素とするネットワークの重みに Hebb 型の教師なし学習を適用することによって、視覚系の神経細胞と類似の時間・空間応答特性を示すニューロンが得られることを示した。重みの自己組織に用いた学習アルゴリズムは主成分分析と同様の次元圧縮を実現するものであり、上記の結果は、視覚系ニューロンの時空間受容野の形成過程が効率的な次元圧縮を目的とする教師なし学習によって説明できることを示している。

- 寄与順の成分抽出を可能にするエネルギー関数 (第 3 章および第 4 章)

  任意の多層パーセプトロンの学習において中間層素子の出力が寄与順に並んだ内部表現を自己組織することのできる重畳エネルギー関数を提案した。重畳

エネルギー関数を線形パーセプトロンの恒等写像学習に適用した場合、主成分分析と同一の内部表現が得られることが理論的に保証される。5 層非線形パーセプトロンの恒等写像に適用した場合、データの非線形な特徴成分を寄与順に抽出する非線形主成分分析を実現することができる。また一般のパーセプトロンの学習においても、中間層素子の出力を寄与順に自己組織するために冗長性が低く効率的な内部表現が得られることを示した。

両者は、特徴量が寄与順に並ぶように制約された次元圧縮を行なっている点が共通している。第 2 章のネットワークモデルにおいては、寄与順に主成分を抽出可能な学習アルゴリズムを使用することによって、計算機シミュレーション結果の解析やモデルの理論的考察が容易になっている。上記の制約のない学習アルゴリズム (例えば文献 [66, 67]) を用いた場合、ニューロン数の選択や学習の初期条件によって実験結果が変動し、得られる受容野の特性を重要度に関連づけて評価することは不可能である。また多層パーセプトロンにおいても、第 3 章および第 4 章の結果が示すように、寄与順に並んだ内部表現を自己組織することでこれまでになかった様々な利点が生まれている。従来、特に多層パーセプトロンに関してはこのような視点に立った研究はほとんど行なわれていないが、このようにニューラルネットワークの学習において寄与順の成分抽出を実現する方法を検討することは重要な問題と考えられる。特に、ニューラルネットワークの工学的応用を考えると、非線形パーセプトロンにおいて効率的な内部表現を構成することが大きな問題であり、この点に関して、本論文の結果を基に理論・実験の両面からより詳細な検討を行なうことが今後の課題である。

付録

A.1 相関関数 C^{ST} の導出

2.4節の解析において用いられる入力 $x(\boldsymbol{r},i)$ の相関関数 $C^{ST}(\boldsymbol{r},i,\boldsymbol{r}',i')$ の導出過程を述べる。2.3節の計算機シミュレーションにおいては、ネットワークへの入力は、ガウシアンノイズを二次元ガウシアンでフィルタリングしこれを各伝達経路のインパルス応答で畳み込むことで生成されていた。$\xi(\boldsymbol{r},t)$ を平均 μ、分散 σ^2 のガウシアンノイズとし、$g_f(\boldsymbol{r};\sigma_f)$ を分散 σ_f^2 の二次元ガウシアンフィルタ

$$g_f(\boldsymbol{r};\sigma_f) = \frac{1}{2\pi\sigma_f^2}\, e^{-\frac{\|\boldsymbol{r}\|^2}{2\sigma_f^2}} \tag{A.1}$$

とする。このとき、伝達経路への入力 $\xi^{\mathcal{S}}(\boldsymbol{r},t)$ は

$$\xi^{\mathcal{S}}(\boldsymbol{r},t) = \int \xi(r-\boldsymbol{r}',t) g_f(\boldsymbol{r}';\sigma_f) d^2\boldsymbol{r}' \tag{A.2}$$

と表される。したがって、$\xi^{\mathcal{S}}(\boldsymbol{r})$ の共分散関数 $Q^{\mathcal{S}}(\boldsymbol{r},\boldsymbol{r}')$ は

$$\begin{aligned} Q^{\mathcal{S}}(\boldsymbol{r},\boldsymbol{r}') &= \left\langle (\xi^{\mathcal{S}}(\boldsymbol{r})-\mu^{\mathcal{S}})(\xi^{\mathcal{S}}(\boldsymbol{r}')-\mu^{\mathcal{S}}) \right\rangle \\ &= \sigma^2 g_f(\boldsymbol{r}-\boldsymbol{r}';\sqrt{2}\sigma_f) \end{aligned} \tag{A.3}$$

で与えられる。ただし、

$$\mu^{\mathcal{S}} = \langle \xi^{\mathcal{S}}(\boldsymbol{r}) \rangle$$

$$
\begin{aligned}
&= \mu \int g_f(\boldsymbol{r};\sigma_f) d^2\boldsymbol{r} \\
&= \mu
\end{aligned}
\tag{A.4}
$$

である。また、i 番目の伝達経路を経てネットワークへ入力される信号 $x(\boldsymbol{r},i,t)$ は

$$
x(\boldsymbol{r},i,t) = \int_{-\infty}^{t} \xi^{\mathcal{S}}(\boldsymbol{r},t')\phi_i(t-t')dt' \tag{A.5}
$$

と表される。したがって、その共分散関数 $Q^{\mathcal{ST}}(\boldsymbol{r},i,\boldsymbol{r}',i')$ は

$$
\begin{aligned}
Q^{\mathcal{ST}}(\boldsymbol{r},i,\boldsymbol{r}',i') &= \left\langle (x(\boldsymbol{r},i)-\mu_i^{\mathcal{ST}})(x(\boldsymbol{r}',i')-\mu_{i'}^{\mathcal{ST}}) \right\rangle \\
&= Q^{\mathcal{S}}(\boldsymbol{r},\boldsymbol{r}') \int_{-\infty}^{t} \phi_i(t-t')\phi_{i'}(t-t')dt' \\
&= \sigma^2 g_f(\boldsymbol{r}-\boldsymbol{r}';\sqrt{2}\sigma_f) T(i,i')
\end{aligned}
\tag{A.6}
$$

で与えられる。ただし、

$$
\begin{aligned}
\mu_i^{\mathcal{ST}} &= \langle x(i,\boldsymbol{r}) \rangle \\
&= \mu^{\mathcal{S}} \int_{-\infty}^{t} \phi_i(t-t')dt' \\
&= \mu
\end{aligned}
\tag{A.7}
$$

である。また、

$$
\begin{aligned}
T(i,i') &= \int_{-\infty}^{t} \phi_i(t-t')\phi_{i'}(t-t')dt' \\
&= \int_{0}^{\infty} \phi_i(t)\phi_{i'}(t)dt
\end{aligned}
\tag{A.8}
$$

である。式 (2.3) のインパルス応答関数に対して式 (A.8) を計算すると、

$$
T(i,i') = \frac{\alpha(2\alpha)!}{(\alpha!)^2} \cdot \frac{(\tau_i\tau_{i'})^{\alpha}}{(\tau_i+\tau_{i'})^{2\alpha+1}} \tag{A.9}
$$

が得られる。以上の結果より、入力の相関関数 $C^{\mathcal{ST}}(\boldsymbol{r},i,\boldsymbol{r}',i')$ は次式で表される。

$$
\begin{aligned}
C^{\mathcal{ST}}(\boldsymbol{r},i,\boldsymbol{r}',i') &= Q^{\mathcal{ST}}(\boldsymbol{r},i,\boldsymbol{r}',i') + \mu_i^{\mathcal{ST}}\mu_{i'}^{\mathcal{ST}} \\
&= S(\boldsymbol{r},\boldsymbol{r}')T(i,i') + \mu^2
\end{aligned}
\tag{A.10}
$$

ただし、$S(\boldsymbol{r},\boldsymbol{r}') = Q^{\mathcal{S}}(\boldsymbol{r},\boldsymbol{r}')$ である。

A.2 補題 3.4の証明

$D_{k+1} = U_{k+1}A_{k+1}^T$より、

$$
\begin{aligned}
D_{k+1} &= \left[\begin{array}{c} \boldsymbol{u}_1^T \\ \vdots \\ \boldsymbol{u}_{k+1}^T \end{array}\right] \left[\begin{array}{ccc|c} d_1\boldsymbol{u}_1 & \dots & d_k\boldsymbol{u}_k & \boldsymbol{a}_{k+1} \end{array}\right] \\
&= \left[\begin{array}{ccc|c} & & & \alpha_1 \\ & D_k & & \vdots \\ & & & \alpha_k \\ \hline \multicolumn{3}{c|}{\mathbf{0}} & \alpha_{k+1} \end{array}\right]
\end{aligned} \tag{A.11}
$$

である。ただし、$\alpha_i = \boldsymbol{u}_i^T\boldsymbol{a}_{k+1}$ とした。また、$D_{k+1}^{-1} = B_{k+1}U_{k+1}$より、

$$
\begin{aligned}
D_{k+1}^{-1} &= \left[\begin{array}{c} \frac{1}{d_1}\boldsymbol{u}_1 \\ \vdots \\ \frac{1}{d_k}\boldsymbol{u}_k \\ \hline \boldsymbol{b}_{k+1}^T \end{array}\right] \left[\begin{array}{ccc} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_{k+1} \end{array}\right] \\
&= \left[\begin{array}{ccc|c} & & & \\ & D_k^{-1} & & \mathbf{o} \\ & & & \\ \hline \beta_1 & \dots & \beta_k & \beta_{k+1} \end{array}\right]
\end{aligned} \tag{A.12}
$$

である。ただし、$\beta_i = \boldsymbol{b}_{k+1}^T\boldsymbol{u}_i$ とした

これらより、

$$
\begin{aligned}
&D_{k+1}D_{k+1}^{-1} \\
&= \left[\begin{array}{ccc|c} 1+\alpha_1\beta_1 & \cdots & \alpha_1\beta_k & \alpha_1\beta_{k+1} \\ \vdots & \ddots & \vdots & \vdots \\ \alpha_k\beta_1 & \cdots & 1+\alpha_k\beta_k & \alpha_k\beta_{k+1} \\ \hline \alpha_{k+1}\beta_1 & \cdots & \alpha_{k+1}\beta_k & \alpha_{k+1}\beta_{k+1} \end{array}\right]
\end{aligned} \tag{A.13}
$$

となる。仮定から、D_{k+1}は正則であり、また$\alpha_{k+1}, \beta_{k+1} \neq 0$であることに注意すると、$\alpha_1, \cdots, \alpha_k, \beta_1, \cdots, \beta_k$はすべて0でなければならない。ゆえに$D_{k+1}$は対角行列となる。

(証明終り)

研究業績一覧

B.1 本論文に関する研究業績

B.1.1 原著論文

1. T. Takahashi and Y. Hirai,
   "Self-organization of spatio-temporal visual receptive fields,"
   IEICE Transactions on Information and Systems, vol.E79-D, no.7, pp.980–989, 1996.

2. 高橋隆史 徳永隆治 平井有三
   「KL変換を実現する3層線形パーセプトロンの教師付き学習則 — Baldi-Hornikの定理の拡張 — 」
   電子情報通信学会和文論文誌, vol.J80–D–II, no.5, pp.1267–1275, 1997

3. 高橋隆史 徳永隆治
   「重畳エネルギー関数による多層パーセプトロンの冗長性削減」
   電子情報通信学会和文論文誌, vol.J80–D–II, no.9, pp.2532–2540, 1997

B.1.2 国際会議 (査読つき)

1. T. Takahashi and Y. Hirai,
   "Self-organization of spatio-temporal receptive fields,"
   Proceedings of the Internatinal Conference on Neural Information Processing '94 (ICONIP'94), vol.2, pp.960–965, 1994.

2. T. Takahashi and R. Tokunaga,
   "Removing the Redundancy of Perceptrons in Terms of a Simple Energy Function,"
   Progress in Connectionist-Based Information Systems, Proceedings of the 1997 Internatinal Conference on Neural Information Processing and Intelligent Information Systems (ICONIP'97), vol.1, pp.271–274, 1997.

3. T. Takahashi and R. Tokunaga,
   "Energy Functions for Efficient Nonlinear Dimensionality Reduction by Multi Layer Perceptrons," Proceedings of the Internatinal Conference on Neural Information Processing (ICONIP'98), vol.1, pp.494–497, 1998.

B.1.3 研究報告

1. 高橋隆史 平井有三
   「視覚系時空間受容野の自己組織形成」
   電子情報通信学会技術研究報告, vol.NC–94, no.129, pp.33–40, 1994.

2. 高橋隆史 平井有三
   「視覚系時空間受容野の自己組織形成—理論解析—」
   日本神経回路学会第 6 回全国大会講演論文集, pp.255–256, 1995.

3. 高橋隆史 平井有三
「網膜神経節細胞の時空間受容野自己組織形成モデル」
電子情報通信学会技術研究報告, vol.NC-96, no.599, pp.223-230, 1996.

4. 高橋隆史 徳永隆治
「多層パーセプトロンによる情報圧縮に適した二つのエネルギー関数」
電子情報通信学会技術研究報告, vol.PRMU-98, no.126, pp.41-46, 1998.

B.2 その他の研究業績

B.2.1 原著論文

1. 高橋隆史 徳永隆治
「画像のブロック平均値から交流成分を予測する高速演算アルゴリズム」
電子情報通信学会和文論文誌, vol.J81-D-II, no.4, pp.778-780, 1998.

2. ユ ヒョンベ 高橋隆史 長谷川友紀 徳永隆治
「コンデンセーション変換の拡張による LIFS 画像符号化法の改良」
電子情報通信学会和文論文誌, vol.J81-D-II, no.7, pp.1576-1583, 1998.

3. ユ ヒョンベ 高橋隆史 河野裕之 徳永隆治
「LIFS 画像符号化法の画品質改善 —グラム-シュミットの直交化を用いた拡張コンデンセーション変換の複合—」
電子情報通信学会和文論文誌, vol.J81-D-II, no.12, pp.2731-2737, 1998.

B.2.2 国際会議 (査読つき)

B.2.3 研究報告

1. ユ ヒョンベ 高橋隆史 河野裕之 徳永隆治
   「グラム–シュミットの直交化を応用したLIFS画像符号化法の画品質改善」
   電子情報通信学会技術研究報告, vol.NLP-98, no.146, pp.37-42, 1998.

参考文献

[1] Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantizer Design," IEEE Transactions on Communications, vol.28, no.1, pp.84–95, 1980.

[2] K. I. Diamantaras and S. Y. Kung, "Principal Component Neural Networks," Wiley, 1996.

[3] T. Kohonen, "Self-Organization and Associative Memory," Springer Verlag, 1989.

[4] J. Hertz, A. Krogh, and R. G. Palmer, "Introduction to The Theory of Neural Computation," Addison–Wesley, 1991.

[5] "The Organization of Behavior," Wiley, 1949.

[6] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning internal representations by error propagation," in Parallel Distributed Processing vol.1, eds. J. L. McCleland, D. E. Rumelhart, and The PDP Research group, chap.8, MIT Press, Cambridge, 1986.

[7] G. W. Cottrell, P. Munro, and D. Zipser, "Leraning Internal Representations from Gray-Scale Images: An Example of Extensional Programming," Ninth Annual Conference of the Cognitive Science Society, pp.462–473, 1987.

[8] S. W. Kuffler, "Discharge patterns and functional organization of mammalian retina," Journal of Neurophysiology, vol.16, pp.37–68, 1953.

[9] D. H. Hubel and T. N. Wisel, "Receptive fields of single neurons in the cat's striate cortex," Journal of Physiology, vol.148, pp.574–591, 1959.

[10] D. H. Hubel and T. N. Wisel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," Journal of Physiology, vol.160, pp.106–154, 1962.

[11] D. Marr, and S. Ullman, "Directional selectivity and its use in early visual processing," Proceedings of Royal Society of London B, vol.211, pp.151–180, 1981.

[12] J. Richter, and S. Ullman, "A model for the temporal organization of X- and Y-type receptive fields in the ptimate retina," Biological Cybernetics, vol.43, pp.127–145, 1982.

[13] D. Marr, "VISION –A Computational Investigation into the Human Representation and Processing of Visual Information–," W. H. Freeman & Company, 1982.
乾敏郎, 安藤広志 訳, "ビジョン –視覚の計算理論と脳内表現–," 産業図書, 1987.

[14] 乾敏郎, "視覚情報処理の基礎," サイエンス社, 1990.

[15] 平井有三, "視覚と記憶の情報処理," 培風館, 1995.

[16] C. Blakemore and G. F. Cooper, "Development of the brain depends on the visual environment," Nature, vol.228, pp.477-478, 1970.

[17] C. von der Malsburg, "Self-organization of orientation sensitive cells in the striate cortex," Kybernetik, vol.14, pp.85–100, 1973.

[18] R. Linsker, "From basic network principles to neural architecture," Proceedings of National Academy of Science USA, vol.83, pp.7508–7512, 8390–8394, 8779–8783, 1986.

[19] R. Linsker, "Self-organization in a perceptual network," IEEE Computer, pp.105–117, 1988.

[20] D. M. Kammen, and A. L. Yuille, "Spontaneous symmetry-breaking energy functions and the emergence of orientation selective cortical cells," Biological Cybernetics, vol.59, pp.23–31, 1988.

[21] A. L. Yuille, D. M. Kammen, and D. S. Cohen, "Quadrature and the development of orientation selective cells by Hebb rules," Biological Cybernetics, vol.61, pp.183–194, 1989.

[22] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network," Neural Networks, vol.2, pp.459–473, 1989.

[23] T. D. Sanger, "Analysis of the two-dimensional receptive fields learned by the Generalized Hebbian Algorithm in response to random input," Biological Cybernetics, vol.63, pp.221–228, 1990.

[24] S. Wimbauer, W. Gerstner, and J. L. van Hemmen, "Emergence of spatiotemporal receptive fields and its application to motion detection," Biological Cybernetics, vol.72, pp.81–92, 1994.

[25] D. W. Tank, and J. J. Hopfield, "Neural computation by concentrating information in time," Proceedings of the National Academy of Science USA, vol.84, pp.1896–1900, 1987.

[26] 野村正英, "MT 野の細胞応答モデリング," 電子情報通信学会技術研究報告 NC94-107, pp.241-247, 1995.

[27] R. Linsker, "Designing a sensory processing system: What can be learned from principal components analysis?," Proceedings of IJCNN'90 (Washington D.C.), vol.2, pp.291–297, Jan. 1990.

[28] R. W. Rodieck, "Quantitative analysis of cat retinal ganglion cell response to visual stimuli," Vision Research, vol.5, pp.583–601, 1965.

[29] R. W. Rodieck, and J. Stone, "Response of cat retinal ganglion cells to moving visual patterns," Journal of Neurophysiology, vol.28, pp.819–832, 1965a.

[30] B. Dreher, and K. J. Sanderson, "Receptive field analysis: responses to moving visual contours by single lateral geniculate neurones in the cat," Journal of Physiology, vol.234, pp.95–118, 1973.

[31] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, "Receptive-field dynamics in the central visual pathways," Trends in Neuroscience, vol.18, no.10, 1995.

[32] D. J. C. MacKay, and K. D. Miller, "Analysis of Linsker's application of Hebbian rules to linear networks," Network, vol.1, pp.257-297, 1990.

[33] R. W. Rodieck, and J. Stone, "Analysis of receptive fields of cat retinal ganglion cells," Journal of Neurophysiology, vol.28, pp.833–849, 1965b.

[34] C. Enroth-Cugell, and L. H. Pinto, "Pure central responses from off-centre cells and pure surround responses from on-centre cells," Journal of Physiology, vol.220, pp.441–464, 1972.

[35] F. M. De Monasterio, "Center and surround mechanisms of opponent-color X and Y ganglion cells of retina of macaques," Journal of Neurophysiology, vol.41, pp.1418–1434, 1978.

[36] G. Cybenko, "Approximation by superpositions of a sigmoidal function," Mathematics of Control, Signals, and Systems, vol.2, no.4, pp.303–314, 1989.

[37] K. Funahashi, "On the approximate realization of continuous mappings by neural networks," Neural Networks, vol.2, pp.183–192, 1989.

[38] K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators," Neural Networks, vol.2, pp.359–366, 1989.

[39] H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition," Biological Cybernetics, 1988.

[40] P. Baldi and K. Hornik, "Neural networks and principal component analysis: Learning from examples without local minima," Neural Networks, vol.2, pp.53–58, 1989.

[41] P. F. Baldi and K. Hornik, "Learning in linear neural networks : A survey," IEEE Transactions on Neural Networks, vol.6, no.4, pp.837–858, 1995.

[42] 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実現についての理論的考察," 電子情報通信学会論文誌, vol.J73-A, no.1, 1990.

[43] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," AIChe Journal, vol.37, no.2, pp.233-243, 1991.

[44] S. Usui, S. Nakauchi, and M. Nakano, "Internal color representation acquired by a five-layer neural network," Proceedings of International Conference of Artificial Neural Networks, vol.1, pp.867–872, 1991.

[45] D. DeMers, and G. Cottrell, "Non-linear dimensionality reduction," in Advances in Neural Information Processing Systems 5, eds. S. J. Hanson, J. D. Cowan and C. L. Giles, pp.580–587, 1993.

[46] 渡辺一央, 伊東英彦, 増田一, 大堀隆文, "KL 変換用多段接続型パーセプトロン," 電子情報通信学会論文誌, vol.J75-D-II, no.11, pp.1925–1932, 1992.

[47] 渡辺一央, 大堀隆文, 下沢楯夫, "KL 変換用単位パーセプトロンの収束性に関する理論的考察," 電子情報通信学会論文誌, vol.J75-D-II, no.11, pp.1933–1939, 1992.

[48] 増田一, 大堀隆文, 渡辺一央, "KL 変換用 3 層構造ニューラルネットワーク," 電子情報通信学会論文誌, vol.J77-D-II, no.2, pp.397–404, 1994.

[49] 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," 電子情報通信学会論文誌, vol.J73-D-II, no.8, pp.1173–1178, 1990.

[50] G. Cybenko, "Continuous Valued Neural Networks with Two Hidden Layers Are Sufficient," Technical Report, Department of Computer Science, Tufts University, 1988.

[51] R. Reed, "Pruning algorithms — A survey," IEEE Transactions on Neural Networks, vol.4, no.5, pp.740–747, 1993.

[52] M. Ishikawa, "Structural learning with forgetting," Neural Networks, vol.9, no.3, pp.509–521, 1996.

[53] 栗田多喜夫, "情報量基準による 3 層ニューラルネットの隠れ層のユニット数の決定法," 電子情報通信学会論文誌, vol.J73-D-II, no.8, pp.1872–1878, 1990.

[54] N. Murata, S. Yoshizawa, and S. Amari, "Network information criterion — Determining the number of hidden units for an artificial neural network model, " IEEE Transactions on Neural Networks, vol.5, no.6, pp.865–872, 1994.

[55] D. J. C. MacKay, "Bayesian methods for adaptive models," Ph. D. Thesis, California Institute of Technology, 1992.

[56] M. Marchand, M. Golea, and P. Ruján, "A Convergence Theorem for Sequantial Learning in Two-Layer Perceptrons," Europhysics Letters, vol.11, pp.487–492.

[57] M. Frean, "The Upstar Algorithm: A Method for Constructing and Training Feedforward Neural Networks," Neural Computation, vol.2, pp.198–209.

[58] S. E. Fahlmann, and C. Lebiere, "The cascade-correlation learning architecture," Advances in Neural Information Processing System 2, pp.524–532, 1990.

[59] M. Mézard, and J.–P. Nadal, "Lerning in Feedforward Layered Networks: The Tiling Algorithm," Journal of Physics A, vol.22, pp.2191–2204.

[60] Y. Le Cun, J. S. Denker, and S. A. Solla, "Optimal brain damage," in Advances in neural information processing systems 2, pp.598–605, 1990.

[61] 萩原将文, "淘汰機能を有するバックプロパゲーション—学習回数の低減と中間層ユニットの削減法—," 電子情報通信学会論文誌, vol.J74-D-II, no.6, pp.812–818, 1991.

[62] S. J. Hanson, and L. Y. Pratt, "Comparing biases for minimal network construction with back-propagation," in Advances in neural information processing systems 1, pp.177–185, 1989.

[63] Y. Chauvin, "A back-propagation algorithm with optimal use of hidden units," in Advances in neural information processing systems 1, pp.519–526, 1989.

[64] M. H. Hassoun, "Fundamentals of artificial neural networks," MIT Press, pp.221–226, 1995.

[65] A. Krogh and J. A. Hertz, "A simple weight decay can improve generalization," in Advances in neural information processing systems 4, pp.951–957, 1992.

[66] E. Oja, "A Simplified Neuron Model As a Principal Component Analyzer," Journal of Mathematical Biology, vol.15, pp.267–273, 1982.

[67] E. Oja, "Neural networks, principal components, and subspaces," International Journal of Neural Systems, vol.1, no.1, pp.61–68, 1989.

謝辞

本研究は、筑波大学電子・情報工学系の平井有三教授ならびに徳永隆治助教授の御指導のもと、視覚情報処理研究室およびカオス研究室において行なわれました。懇切なる御指導を賜わった両先生に心より御礼申し上げます。また、様々な御協力や暖かい励ましを頂いた両研究室の皆様に感謝致します。さらに、本研究の遂行にあたって様々の助言や激励を頂いた理化学研究所の松本元氏、電子技術総合研究所の重松征史氏、栗田多喜夫氏、NEC 基礎研究所の岡島健治氏、野村正英氏、宮下真信氏、ならびに本論文の審査にあたって貴重な助言を頂いた筑波大学電子・情報工学系の名取亮教授、板橋秀一教授、寅市和男教授、安永守利助教授に深謝し御礼申し上げます。最後に、筆者を支え暖かく見守って下さった明子さん、そして父母に感謝します。

筑波大学附属図書館
1 00990 12365 6

本学関係