KL 変換を実現する 3 層線形パーセプトロンの学習則 / Baldi・ Hornik の定理の拡張 *

筑波大学大学院博士課程工学研究科 高橋隆史

ume@viplab.is.tsukuba.ac.jp

1996 年 4 月 17 日

1　まえがき

Rumelhart[1] 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、関数近似への適用に関する近似万能性の証明 [2, 3, 4]、砂時計型多層パーセプトロンによる恒等写像近似に関する理論的解析 [5, 6, 7] 等が報告されている。本研究では、砂時計型多層パーセプトロンに関する諸結果

1. 3 層線形パーセプトロンは学習によって Karhunen-Loève(KL) 変換と等価な変換を実現できる [5, 6]
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は線形パーセプトロンの能力を超えない [7]
3. 5 層以上の非線形パーセプトロンは KL 変換の能力を上回る可能性がある [8, 9, 10]

の中で、 Baldi と Hornik の結果 [6] に注意を払う。

Baldi・ Hornik の定理 [6] によると、通常の二乗誤差関数に基づく線形パーセプトロンの学習によって、重みベクトルの張る部分空間は KL 基底のそれと一致し、恒等写像の最適線形近似を実現することが理解されている。しかし、 KL 変換ではその情報量に従って成分の順が定まり、近似に対する各々の寄与を計算できるのに対して、パーセプトロンの重みは KL 基底の線形結合に収束し、成分分析は困難である。そこで本研究では、 3 層線形パーセプトロンにおいて KL 変換を実現する学習則を検討し、その能力に対して理論的保証を与えることを目的とする。

以降、 2において 3 層砂時計型線形パーセプトロンの学習に関する Baldi・ Hornik[6] の定理とその意義を説明する。次に、エネルギー関数を変更することにより内部表現として KL 変換係数を実現する学習則を与え、その妥当性を理論的に明らかにする。 3において、与えられた学習則の有効性を数値計算によって検証する。最後に、 4において提示された学習則の 5 層砂時計型非線形パーセプトロンへの応用可能性について論議する。

2　Baldi・ Hornik の定理とその拡張

2.1　諸定義

定義 1 *3* 層線形パーセプトロン
　領域 $U(\subset \boldsymbol{R}^n)$ および領域 $V(\subset \boldsymbol{R}^p)$ 上の連続写像

$$\begin{array}{lcll} \psi & : & V \to W & (\text{ただし} W \subset \boldsymbol{R}^m) \\ \psi' & : & U \to V & \end{array} \tag{1}$$

による U 上の合成写像

$$\psi \circ \psi' \quad : \quad U \to W \tag{2}$$

を考える。二つの写像が共に線形、即ち

$$\begin{array}{lcll} \psi & : & \boldsymbol{u} \mapsto A_p \boldsymbol{u} & (u \in V) \\ \psi' & : & \boldsymbol{x} \mapsto B_p \boldsymbol{x} & (x \in U) \end{array} \tag{3}$$

ただし、

$$\begin{array}{lcl} A_p & = & [\boldsymbol{a}_1|\boldsymbol{a}_2|\cdots|\boldsymbol{a}_p] \\ B_p & = & {}^T[\boldsymbol{b}_1|\boldsymbol{b}_2|\cdots|\boldsymbol{b}_p] \\ \multicolumn{3}{l}{\boldsymbol{a}_i \text{は} m \text{ 次元列ベクトル}} \\ \multicolumn{3}{l}{\boldsymbol{b}_i \text{は} n \text{ 次元列ベクトル}} \end{array} \tag{4}$$

である場合、合成写像

$$\psi \circ \psi' : \boldsymbol{x} \mapsto A_p B_p \boldsymbol{x} \tag{5}$$

を線形パーセプトロンと呼ぶ。

神経回路網の述語において、 $\boldsymbol{a}_i$ は出力層と中間層第 i 素子間の重みベクトル、 $\boldsymbol{b}_i$ は入力層と中間層第 i 素子間の重みベクトルと呼ばれる[1] *(*図 *1*参照*)*。また、ある入力値 $\boldsymbol{x} \in U$ に対する中間層素子の値 $\boldsymbol{u} = B_p \boldsymbol{x} \in V$ は入力 $\boldsymbol{x}$ に対する *3* 層パーセプトロンの内部表現と呼ばれる。

[1] 神経回路網の一般的な述語では、中間層 – 出力層間の重みベクトルは各中間層素子からある出力素子への伝達経路の重み値により定義するが、本論文では Baldi と Hornik[6] に従い、上述のように定義する。

* 現在電子情報通信学会和文論文誌に投稿中

定義 2 砂時計型線形パーセプトロン
式 *(5)* に対して、条件

$$n = m \geq p > 0 \tag{6}$$

が成立する場合、線形パーセプトロンの構造は砂時計型であると呼ぶ。

定義 3 恒等写像の学習
U 上の零平均 n 次元ベクトル値の集合 *(*データ集合*)*

$$\{\boldsymbol{x}_\mu \in U\}_{\mu=1,2,\ldots,N} \tag{7}$$

に対して、エネルギー関数 *(*平均二乗誤差*)*

$$E_p(A_p, B_p) = \frac{1}{N}\sum_{\mu=1}^{N} ||\boldsymbol{x}_\mu - A_p B_p \boldsymbol{x}_\mu||^2 \tag{8}$$

を定義する。式 *(8)* に対して勾配法を適用して E_p の最小値を与える重み行列の組 $\{\tilde{A}_p, \tilde{B}_p\}$ を求める過程を学習と呼ぶ。また得られたパーセプトロンは、$E_p(\tilde{A}_p, \tilde{B}_p)$ を誤差とする U 上の恒等写像

$$\phi : \boldsymbol{x}_\mu \mapsto \boldsymbol{x}_\mu \quad (\mu = 1, 2, \ldots, N) \tag{9}$$

の近似を与える。

定義 4 *Karhunen-Loève(KL)* 変換
データ集合 *(7)* に対する共分散行列

$$\Sigma = \frac{1}{N}\sum_{\mu=1}^{N} \boldsymbol{x}_\mu{}^T \boldsymbol{x}_\mu \tag{10}$$

は、

1. フルランクである

2. 相異なる固有値 $\lambda_1 > \cdots > \lambda_n > 0$ をもつ

と仮定する。このとき、λ_i に対応する正規化された固有 *(*列*)* ベクトル $\boldsymbol{u}_i$ によって構成される $n \times p$ 行列

$$U_p = [\boldsymbol{u}_1 | \boldsymbol{u}_2 | \ldots | \boldsymbol{u}_p] \tag{11}$$

は、p 次元 *KL* 逆変換を与える。また、$p \times n$ 行列

$${}^T U_p = \begin{bmatrix} {}^T\boldsymbol{u}_1 \\ \hline {}^T\boldsymbol{u}_2 \\ \hline \vdots \\ \hline {}^T\boldsymbol{u}_p \end{bmatrix} \tag{12}$$

は、p 次元 *KL* 変換を与える。

合成変換 $U_p{}^T U_p$ は、n 次元入力をより少ない p 次元線形空間へ写像し、逆変換で元の n 次元線形空間に戻す恒等写像の近似を実現する。*KL* 変換・*KL* 逆変換は p 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、平均二乗誤差 *(8)* を最小化するという意味で最適な変換であることが知られている。また、固有ベクトル $\boldsymbol{u}_i$ による *KL* 変換係数

$$\boldsymbol{u}_i \cdot \boldsymbol{x} \tag{13}$$

の分散は対応する固有値 λ_i に等しく、大きな固有値に対応する変換係数ほど誤差減少への寄与が大きくなる。

2.2 Baldi・Hornik の定理とその意義

[Baldi・Hornik の定理]

フルランク行列 $\tilde{A}_p$、$\tilde{B}_p$ が $E_p(A_p, B_p)$ の最小値を与えるならば、正則な $p \times p$ 行列 C_p が存在して、

$$\begin{aligned} \tilde{A}_p &= U_p C_p \\ \tilde{B}_p &= C_p^{-1T} U_p \end{aligned} \tag{14}$$

が成立する。また、その逆が成立する。

式 (8) で与えられるエネルギー関数は二次形式であるため、学習によって重み行列 $\{A_p, B_p\}$ は式 (14) を満足する行列 $\{\tilde{A}_p, \tilde{B}_p\}$ に収束する。このとき、

$$\tilde{A}_p \tilde{B}_p = U_p{}^T U_p \tag{15}$$

より、$\tilde{A}_p \tilde{B}_p$ は KL 変換と等価な合成変換を実現している。しかし、重み行列 $\{\tilde{A}_p, \tilde{B}_p\}$ と $\{U_p, {}^T U_p\}$ は正則行列 $\{C_p, C_p^{-1}\}$ によって各々対応付けされるために、必ずしも

$$\boldsymbol{a}_i \mathbin{/\!/} \boldsymbol{u}_i, \quad \boldsymbol{b}_i \mathbin{/\!/} \boldsymbol{u}_i \quad (i = 1, 2, \ldots, p) \tag{16}$$

が成立せず、一般に内部表現は KL 変換係数と異なる。

ここで、「恒等写像の近似に関する寄与の大きさに従って各基底の順が定まること」が KL 変換の有効性であることに注意する。この性質によって、KL 変換では抽出された特徴の重要性が評価可能となり、情報圧縮・特徴抽出において効力を発揮する。以上の意味において Baldi・Hornik の定理には拡張の必要性があると考えられる。

2.3 KL 変換を実現する学習則

本節では KL 変換係数と一致する内部表現を実現するための学習則を提案し、その妥当性を理論的に保証する。

2.3.1 学習則の概略

中間層に p 個の素子をもつ砂時計型線形パーセプトロンを考慮する。それぞれの中間層素子を任意に番号

付けし、中間層第 k 素子まで考慮したパーセプトロンの重み行列を

$$\begin{aligned} A_k &= [\boldsymbol{a}_1|\boldsymbol{a}_2|\cdots|\boldsymbol{a}_k] \\ B_k &= {}^T[\boldsymbol{b}_1|\boldsymbol{b}_2|\cdots|\boldsymbol{b}_k] \end{aligned} \tag{17}$$

で与える。また、式 (17) に対応する出力

$$\boldsymbol{y}_k = A_k B_k \boldsymbol{x} \tag{18}$$

を用いて誤差関数

$$E_k = \frac{1}{N}\sum_{\mu=1}^{N} ||\boldsymbol{x}_\mu - A_k B_k \boldsymbol{x}_\mu||^2 \tag{19}$$

を定義する。一般の学習では E_p のみを用いるが、誤差 E_k の和

$$\tilde{E}_p = \sum_{k=1}^{p} E_k \tag{20}$$

でエネルギー関数を与えた場合の学習の効果について考えてみる。

例えば、$p = 2$ の場合、中間層第 2 素子を用いたことに起因する重み修正量は E_2 のみで決定されるが、第 1 素子に関する重み修正量は $\tilde{E}_2$ で決まり、$\tilde{E}_2$ に基づいて学習を行うパーセプトロンは E_2 を減らすと同時に、E_1 を減らす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を、中間層第 2 素子はこれを補う特徴を出力すると考えられる。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と期待できる。

2.3.2 Baldi・Hornik の定理の拡張

[定理]

フルランク行列 $\tilde{A}_p$、$\tilde{B}_p$ が $\tilde{E}_p$ の最小値を与えるならば、正則な $p \times p$ 対角行列 D_p が存在して、

$$\begin{aligned} \tilde{A}_p &= U_p D_p \\ \tilde{B}_p &= D_p^{-1\,T} U_p \end{aligned} \tag{21}$$

が成立する。また、その逆が成立する。

定理の証明のために補題を準備する。

[補題]

直交する正規列ベクトル $\{\boldsymbol{u}_i \in \boldsymbol{R}^n\}_{i=1,2,\ldots,k}$ $(1 \leq k \leq n)$ から成る行列

$$U_k = [\boldsymbol{u}_1|\ldots|\boldsymbol{u}_k] \tag{22}$$

及び、非零の定数 $\{d_i\}_{i=1,2,\ldots,k}$ を対角要素とする k 次対角行列

$$D_k = diag(d_1,\ldots,d_k) \tag{23}$$

を考える。これらを用いて $n \times k$ 行列 A_k および $k \times n$ 行列 B_k を

$$\begin{aligned} A_k &= U_k D_k \\ B_k &= D_k^{-1\,T} U_k \end{aligned} \tag{24}$$

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

$$\begin{aligned} A_{k+1} &= [A_k|\boldsymbol{a}_{k+1}] \\ B_{k+1} &= \left[\frac{B_k}{{}^T\boldsymbol{b}_{k+1}}\right] \end{aligned} \tag{25}$$

とする。このとき、

$$\begin{aligned} A_{k+1} &= U_{k+1} D_{k+1} \\ B_{k+1} &= D_{k+1}^{-1\,T} U_{k+1} \end{aligned} \tag{26}$$

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

[補題の証明] 付録参照

[定理の証明] $\tilde{A}_p$ の第 1 列から第 k 列 $(k = 1,\ldots,p)$ までの重みベクトルで構成された $n \times k$ 行列を $\tilde{A}_k$、$\tilde{B}_p$ の第 1 行から第 k 行までの重みベクトルで構成された $k \times n$ 行列を $\tilde{B}_k$ とおく。$\{\tilde{A}_p, \tilde{B}_p\}$ が式 (21) を満足するならば、Baldi・Hornik の定理より、すべての k について $\tilde{A}_k$、$\tilde{B}_k$ が E_k の最小値を与える。したがって、式 (21) を満足する $\{\tilde{A}_p, \tilde{B}_p\}$ は $\tilde{E}_p$ の最小値を与える。逆に、すべての E_k が同時に最小値をとるような $\{\tilde{A}_p, \tilde{B}_p\}$ が存在すれば、その $\{\tilde{A}_p, \tilde{B}_p\}$ に対して $\tilde{E}_p$ も最小値をとる。$k = 1,\ldots,p$ について Baldi・Hornik の定理を考慮すると、正則な p 次正方行列 $\tilde{D}_p$ が存在しかつその k 次の主座小行列 $\tilde{D}_k$ すべてが正則で、

$$\begin{aligned} \tilde{A}_k &= U_k \tilde{D}_k \\ \tilde{B}_k &= \tilde{D}_k^{-1\,T} U_k \end{aligned} \tag{27}$$

が成り立つ場合に、すべての E_k が最小値をとることが示される。$k = 1$ で式 (27) が成り立てば、補題より $k = 2$ で式 (27) が成り立つためには $\tilde{D}_2$ 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 $k = 3,\ldots,p$ についても $\tilde{D}_k$ は対角行列となる。故に、すべての E_k が最小値をとる $\{\tilde{A}_p, \tilde{B}_p\}$ が存在し、$\tilde{D}_p$ は対角行列となる。 (証明終り)

証明過程から明らかなように、任意の正定数 $\{\beta_i\}_{i=1,2,\ldots,p}$ によって重み付けされたエネルギー関数

$$\tilde{E}'_p = \sum_{k=1}^{p} \beta_k E_k \tag{28}$$

についても同様の定理が成り立つ。

[系]

フルランク行列 $\tilde{A}_p$、$\tilde{B}_p$ が任意の正定数集合 $\{\beta_k\}_{k=1,2,\ldots,p}$ に対して、$\tilde{E}'_p$ の最小値を与えるならば、正則な $p \times p$ 対角行列 D_p が存在して、

$$\begin{aligned} \tilde{A}_p &= U_p D_p \\ \tilde{B}_p &= D_p^{-1T} U_p \end{aligned} \tag{29}$$

が成立する。また、その逆が成立する。

エネルギー関数 $\tilde{E}_p$ もしくは $\tilde{E}'_p$ の最小値を与える出力層 – 第 k 中間層素子間の重みベクトルを $\tilde{a}_k$、第 k 中間層素子 – 入力層間の重みベクトルを $\tilde{b}_k$ とおき、$D_p = diag(d_1, \ldots, d_p)$ と表すと、式 (11) と式 (21) より、

$$\begin{aligned} \tilde{a}_i &= d_i \boldsymbol{u}_i \\ \tilde{b}_i &= \frac{1}{d_i} \boldsymbol{u}_i \end{aligned} \tag{30}$$

となる。故に、

$$\boldsymbol{a}_i \,/\!/\, \boldsymbol{u}_i, \quad \boldsymbol{b}_i \,/\!/\, \boldsymbol{u}_i \quad (i = 1, 2, \ldots, p) \tag{31}$$

がエネルギー関数最小化の必要十分条件となることが定理によって保証された。

3 数値実験による検証

2において提案した学習則の効果を数値実験により確認する。実験には

$$n = m = p = 3 \tag{32}$$

なる線形パーセプトロンを用い、データ集合

$$\{\boldsymbol{x}_\mu \in \boldsymbol{R}^3\}_{\mu=1,2,\ldots,1000} \tag{33}$$

に対して恒等写像学習を行った。データ集合は 3 次元零平均ガウス分布に従うベクトル値であり、学習回数及び学習定数はそれぞれ 1000 回、1.0 と選択して $\tilde{E}_p$ に対する最急降下法を適用した。

図 2は、学習回数に対する二乗誤差 $\{E_i\}_{i=1,2,3}$ を提示している。図中の横線はデータ集合に対して 1 次元および 2 次元の KL 変換・KL 逆変換を適用した場合の二乗誤差を示している。一方、図 3はデータ集合に重ねて重みベクトル $\{\tilde{\boldsymbol{b}}_i\}_{i=1,2,3}$ を表示している。KL 基底同様に、$\tilde{\boldsymbol{b}}_1$ がデータ集合の最も分散の大きな方向を向き、$\tilde{\boldsymbol{b}}_2$ がそれと直交した平面内で分散の大きな方向を、$\tilde{\boldsymbol{b}}_3$ が両者に直交した方向を向いていることが理解される。より詳しく KL 基底と比較するため、これら重みベクトルとデータ集合の KL 基底 $\{\boldsymbol{u}_i\}_{i=1,2,3}$ との方向余弦を求める。表 1より各重みベクトルが KL 基底に平行になっていることがわかる。同一条件下で通常の誤差逆伝搬学習を行った場合、表 2が示すように各重みベクトルは KL 基底と直接対応しないことに注意を促す。

4 議論

一般に、砂時計型パーセプトロンの恒等写像学習を情報圧縮や特徴抽出に適用する場合、KL 変換を上回る効果を期待して、5 層以上の非線形パーセプトロンが利用される [8, 9, 10]。2において提案した学習則は、3 層砂時計型パーセプトロンに限らず、より多層の砂時計型パーセプトロン[2]、さらに任意の多層パーセプトロンの学習に適用可能である。これを提示するため、情報圧縮のために非線形性が必要となる簡単な問題に対して、5 層砂時計型非線形パーセプトロン (図 4参照) における提案手法を適用する。

実験には入力層から順に 3–10–3–10–3 の素子を配置した 5 層パーセプトロンを用い、第 2 層と第 4 層の素子の出力関数はシグモイド関数とした。データ集合 $\{\boldsymbol{x}_\mu \in \boldsymbol{R}^3\}_{\mu=1,2,\ldots,100}$ は半球面上の点とし、学習回数、学習定数はそれぞれ 15000 回、0.05 として定数 0.9 の慣性項を付加して学習を行った。

図 5(a) は提案した学習則による学習後のパーセプトロンの出力をデータ集合に重ねてプロットしたものである。出力は実線で、データ集合は破線で示され、それぞれの値は格子点上に存在する。また、図 5(b) は中間層第 2 素子までを用いて計算される出力を示している。第 3 素子まで用いた (a) と比較すると半球面の天頂部分で若干誤差が大きくなるが、学習則の効果により第 2 素子までの二次元の内部表現でほぼその形状を近似している事が理解される。

次に、獲得された内部表現を解析し、通常の誤差逆伝搬学習との比較を行う。図 5(c) は提案学習則による学習後のパーセプトロンで 3 つの中間層素子のうち一つだけを用いて出力を計算した結果を示す。中間層第 1 および第 2 素子が非線形な曲線座標を成し、これら二次元の内部表現でほぼ半球面が復元されている。また、第 3 素子の出力は変異が小さく、半球面再構成への寄与も小さい。一方、通常の誤差逆伝搬学習を適用する場合、あらかじめ中間層素子数を 2 として学習を行えば、半球面の二次元内部表現を学習することができる [10] が、素子数を 3 として上述の実験と同一条件で学習を行うと、その内部表現は図 5(d) に示すように単なる線形直交座標に対応したものとなる。また、この場合中間層で何らの特徴抽出もされておらず、内部表現に KL 変換を適用してもこのデータ集合が本質的に二次元であることを推定できないことに注意を促す。

一般に、情報圧縮・特徴抽出問題においては恒等写像近似に必要なパラメータ数は不明であり、多層パーセプトロンの利用に際して学習前に最適な中間層素子数を推定するのは困難である。しかし本研究で提案した学習則は非線形パーセプトロンによる情報圧縮にも

[2] 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 p としたとき、式 (6) が成立するパーセプトロンを砂時計型と定義する。

有効で、学習によって冗長性の低い内部表現を自動的に獲得できると考えられる。

参考文献

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

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

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

[4] G.Cybenko, Approximation by superpositions of a sigmoidal function, Math.Cont.Sig.Sys. vol.2, no.4, pp.303–314, 1989.

[5] H.Bourlard and Y.Kamp, Auto-association by multilayer perceptrons and singular value decomposition, Biol.Cybern., 1988.

[6] 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.

[7] 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, IEICE Trans, vol.J73-A, no.1, 1990.

[8] S.Usui, S.Nakauchi and M.Nakano, Internal color representation acquired by a five-layer neural network, ICANN, pp.867–872, 1991.

[9] D.DeMers and G.Cottrell, Non-linear dimensionality reduction, NIPS 5, pp.580–587, 1993.

[10] 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, IEICE Trans., vol.J73-D-II, no.8, pp.1173–1178, 1990.

[補題の証明]

$D_{k+1} = {}^T U_{k+1} A_{k+1}$ より、

$$
\begin{aligned}
D_{k+1} &= \begin{bmatrix} {}^T\boldsymbol{u}_1 \\ \vdots \\ {}^T\boldsymbol{u}_{k+1} \end{bmatrix} \left[\begin{array}{ccc|c} d_1\boldsymbol{u}_1 & \dots & d_k\boldsymbol{u}_k & \boldsymbol{a}_{k+1} \end{array}\right] \\
&= \left[\begin{array}{c|c} & \alpha_1 \\ D_k & \vdots \\ & \alpha_k \\ \hline \mathbf{o} & \alpha_{k+1} \end{array}\right]
\end{aligned} \tag{34}
$$

ただし、$\alpha_i = {}^T\boldsymbol{u}_i \boldsymbol{a}_{k+1}$

また、$D_{k+1}^{-1} = B_{k+1} U_{k+1}$ より、

$$
\begin{aligned}
D_{k+1}^{-1} &= \left[\begin{array}{c} \frac{1}{d_1}\boldsymbol{u}_1 \\ \vdots \\ \frac{1}{d_k}\boldsymbol{u}_k \\ \hline {}^T\boldsymbol{b}_{k+1} \end{array}\right] \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_{k+1} \end{bmatrix} \\
&= \left[\begin{array}{ccc|c} & D_k^{-1} & & \mathbf{o} \\ \hline \beta_1 & \dots & \beta_k & \beta_{k+1} \end{array}\right]
\end{aligned} \tag{35}
$$

ただし、$\beta_i = {}^T\boldsymbol{b}_{k+1} \boldsymbol{u}_i$

これらより、

$$
D_{k+1} D_{k+1}^{-1} = \left[\begin{array}{ccc|c} 1+\alpha_1\beta_1 & \cdots & \alpha_1\beta_k & \alpha_1\beta_{k+1} \\ \vdots & \ddots & \vdots & \vdots \\ \alpha_k\beta_1 & \cdots & 1+\alpha_k\beta_k & \alpha_k\beta_{k+1} \\ \hline \alpha_{k+1}\beta_1 & \cdots & \alpha_{k+1}\beta_k & \alpha_{k+1}\beta_{k+1} \end{array}\right] \tag{36}
$$

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。

(補題の証明終り)

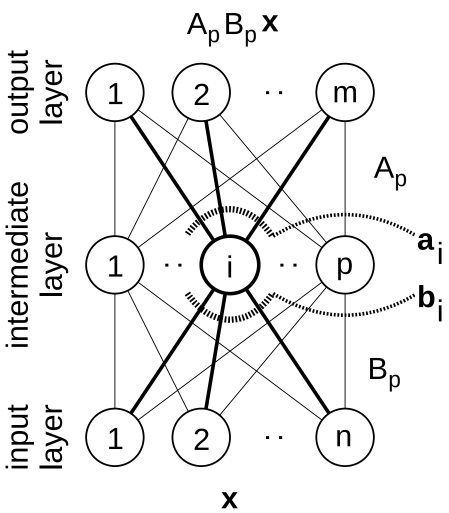


図 1: 3 層線形パーセプトロン

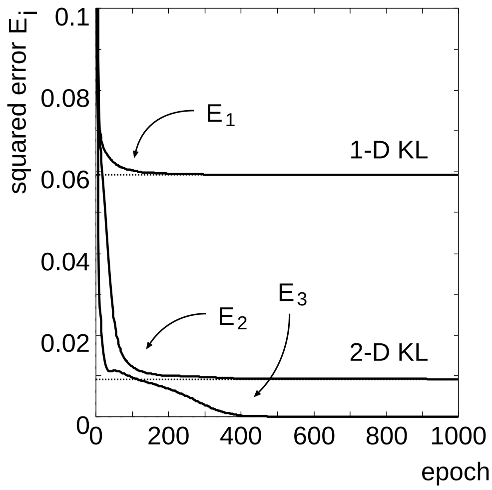


図 2: 提案学習則を用いた線形パーセプトロンと KL 変換・ KL 逆変換の二乗誤差の比較：横軸はパーセプトロンの学習回数を表し、図中の点線はデータ集合に対して 1 次元および 2 次元の KL 変換・ KL 逆変換を適用した場合の二乗誤差を表す。

表 1: 提案学習則における中間層素子の重みベクトル $\{\tilde{\boldsymbol{b}}_i\}_{i=1,2,3}$ とデータ集合の KL 基底 $\{\boldsymbol{u}\}_{i=1,2,3}$ の方向余弦。

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

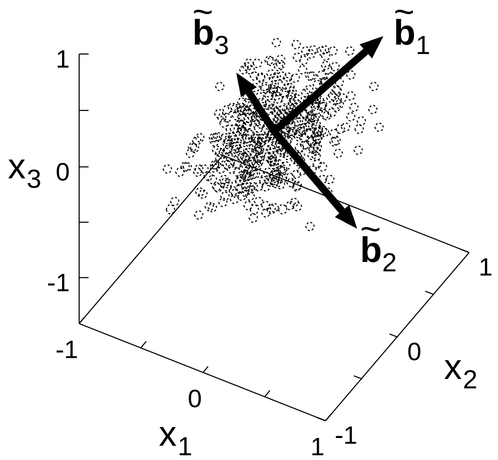


図 3: データ集合に重ねて表示した学習後の重みベクトル。破線の丸は入力データを表し、矢線は中間層素子の重みベクトル $\{\tilde{\boldsymbol{b}}_i\}_{i=1,2,3}$ を表す。

表 2: 通常の誤差逆伝搬学習における中間層の重みベクトルと KL 基底の方向余弦。

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

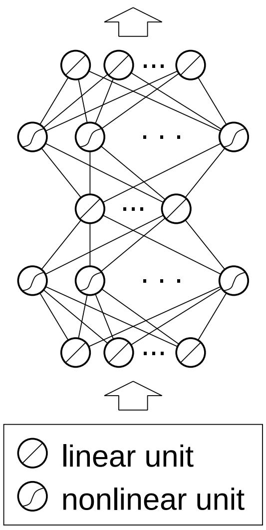
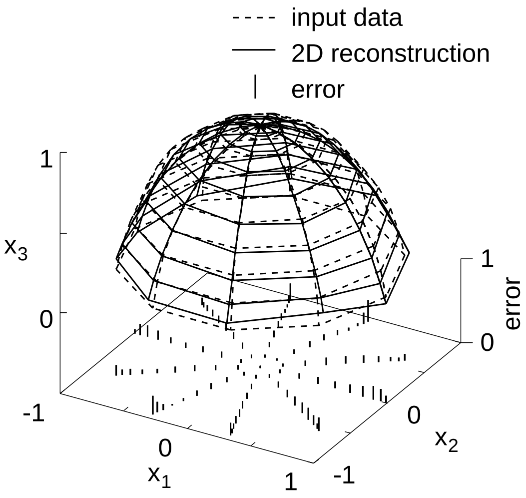
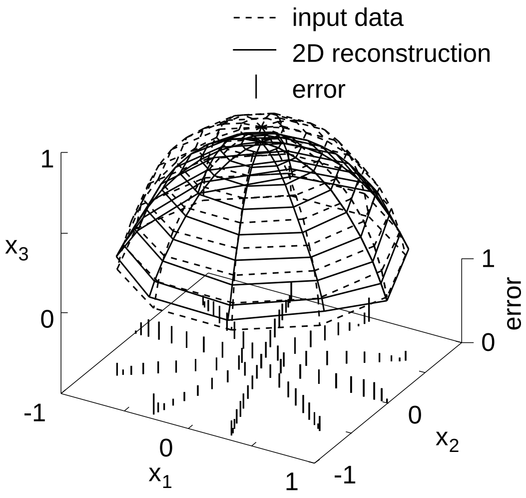


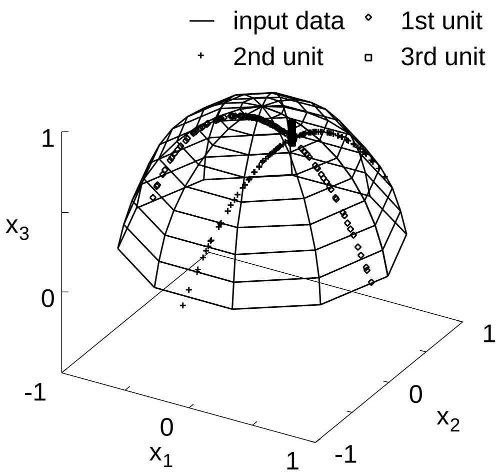
図 4: 5 層砂時計型非線形パーセプトロン



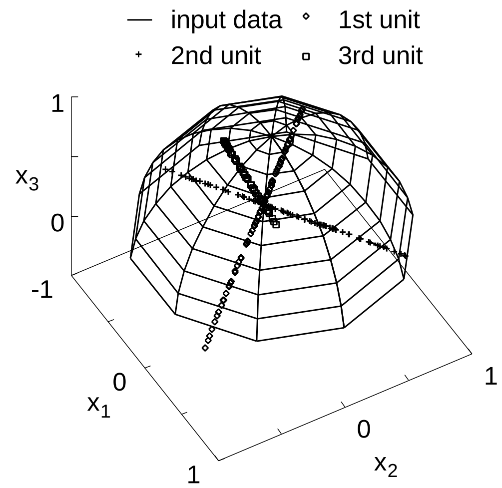
(a)



(b)



(c)



(d)

図 5: 中間層に 3 つの素子を用いた 5 層非線形パーセプトロンによる半球面の恒等写像学習： (a) 提案学習則を適用したパーセプトロンの出力。 (b) 同一のパーセプトロンで中間層第 2 素子までを用いて得られる出力。 (c) 中間層素子それぞれ一つだけを用いて計算される出力。 (d) 通常の誤差逆伝搬学習を適用した場合。 (出力の計算は (c) と同様)