重畳エネルギー関数を用いた砂時計型MLPによる非線形主成分分析
—統計的基準による中間層素子数の選択法について—

高橋隆史†, 徳永隆治‡

† 筑波大学 電子・情報工学系 学振特別研究員(PD)
〒305-8573 筑波大学 電子・情報工学系 カオス研究室
e-mail: ume@chaos.is.tsukuba.ac.jp
‡ 筑波大学 電子・情報工学系

砂時計型多層パーセプトロン(MLP)の恒等写像学習を応用した非線形主成分分析においては、ボトルネックとなる中間層の素子数を適切に選択することが重要である。この問題に対する一つのアプローチとして、筆者らはこれまでに、中間層素子が寄与順に並んだ内部表現を自己組織する重畳エネルギー関数を提案し、これを用いたMLPでは学習後に任意の素子数のデータ表現を抽出可能となることを示している。本論文では、この重畳エネルギー関数を用いたMLPの中間層素子数選択のための指標として、平岡・吉澤によって提案された統計基準を適用した結果を報告する。

砂時計型**MLP**、恒等写像学習、主成分分析、次元圧縮、次元数選択

Nonlinear Principal Component Analysis by MLP using Superposed Energy
— On estimation of the intrinsic dimensionality of data —

Takashi TAKAHASHI†, Ryuji TOKUNAGA‡

† JSPS Research Fellow, Institute of Information Sciences and Electronics,
University of Tsukuba, Ibaraki 305–8573, Japan
e-mail: ume@chaos.is.tsukuba.ac.jp
‡ Institute of Information Sciences and Electronics, University of Tsukuba

In order to obtain good performance in the autoassociative learning of a sandglass-type MLP, the dimensionality of its bottleneck-layer must be selected appropriately. This paper reports that a simple energy function, called “superposed energy,” is applicable to nonlinear principal component analysis, and that the intrinsic dimensionality of nonlinear data is successfully estimated by applying the statistical criteria proposed by Hiraoka and Yoshizawa.

sandglass-type MLP, autoassociative learning, principal component analysis, dimensionality reduction

1　はじめに

多層パーセプトロン (MLP) の恒等写像学習を用いた次元圧縮に関して、3層パーセプトロンが線形の主成分分析と等価であること [1]、砂時計型 MLP によって非線形の主成分分析が実現できること [2, 3] などが知られており、理論・応用の両面から多様な研究が行なわれている。このような MLP で効率の良い次元圧縮を実現するためには、ボトルネックとなる中間層の素子数 (中間次元数) を適切に設定することが必要である。この問題に対して平岡・吉澤 [7] は、中間次元数の異なる砂時計型 MLP の中から最適な次元数のものを選択するための基準量を提案している。一方われわれは、この問題に対する別のアプローチとして、中間層出力に寄与順に並んだデータ表現を自己組織するためのエネルギー関数 (重畳エネルギー [4, 5]) を検討し、これを用いた場合には一つの MLP で学習を行なった後に中間次元数を任意に選択できることを示している。本研究では、重畳エネルギーを用いた砂時計型 MLP による非線形次元圧縮に平岡・吉澤の基準を適用した結果を報告する。従来の方法では複数の MLP で学習を行ない最適な次元数のものを選択する必要があるのに対して、重畳エネルギーの場合には一度の学習で得られたデータ表現から最適な次元数の成分を抽出できることを示す。

2　MLP による次元圧縮と中間次元数選択

2.1　砂時計型 MLP による次元圧縮

図1に示すような、二つの MLP を直列に接続した形式の MLP を考える。入力層と出力層の素子が同数 N であり、中間層の素子数が $K < N$ である場合、この MLP は砂時計型であるという。この砂時計型 MLP への入力データとして T 個の N 次元ベクトルから成るデータ集合

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

を与えた場合、前段部分 (“encoder”) は各データベクトル $\boldsymbol{x}_t$ を K 次元のデータ表現に変換し、後段部分 (“decoder”) はそれをさらに N 次元に変換し出力する、という操作を行うことになる。$\boldsymbol{x}_t$ に対する出力を $\boldsymbol{y}_t$ と表すとき、教師信号に入力データ自身を用いてエネルギー関数

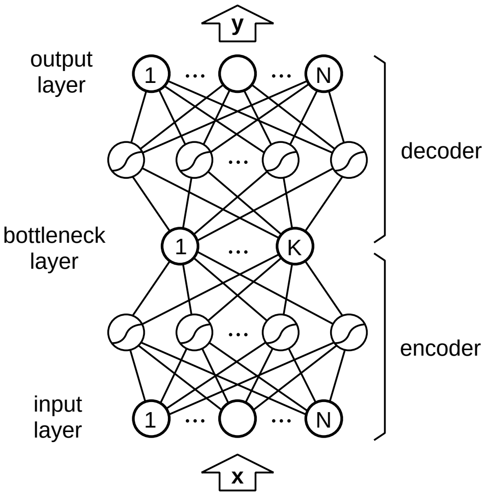


図 1: 5層砂時計型 MLP

$$E = \frac{1}{T}\sum_{t=1}^{T} ||\boldsymbol{x}_t - \boldsymbol{y}_t||^2 \tag{2}$$

に対する学習を行えば、この MLP の出力はデータ集合に対する恒等写像を近似する。したがって、学習後に K 次元のデータ表現すなわち中間層出力を取り出すことで、データの次元数を N から K に減らす次元圧縮を実現することができる。

上記の場合、MLP の出力 $\boldsymbol{y}_t$ は、入力 $\boldsymbol{x}_t$ を K 次元のある曲面 S に射影して $\boldsymbol{x}_t$ を近似したものとなる。このような砂時計型 MLP の恒等写像学習は、「ノイズを含む観測値から真値の満たす関係を推測し、観測値を補正する」という問題と考えることができ、以下のような推定問題と等価となる [6]。

N 次元空間中に、次元数 L が未知のある曲面 S^* を考える。このとき、S^* 上にデータ集合

$$\left\{\boldsymbol{z}_t \in R^N\right\}_{t=1}^{T} \tag{3}$$

が与えられ、それにノイズの加わったデータ

$$\left\{\boldsymbol{x}_t = \boldsymbol{z}_t + \boldsymbol{n}_t\right\}_{t=1}^{T} \tag{4}$$

が観測される。ただし、$\boldsymbol{n}_t$ は正規分布 $N(\mathbf{o}, \sigma^2 I)$ に従うノイズとする。観測値 $\{\boldsymbol{x}_t\}$ から、S^* と真の値 $\{\boldsymbol{z}_t\}$ を推定せよ。

この場合、推定値 $\boldsymbol{y}$ が真値 $\boldsymbol{z}$ に近ければ良いわけであるから、

$$E^* = \mathrm{E}\left[||\boldsymbol{y}-\boldsymbol{z}||^2\right] \tag{5}$$

を推定の良さの基準とすることができる。真値ののった曲面 S^* を適当な方法で推定した曲面 S が良い近似となっていれば、E^* は小さくなる。したがって、砂時計型MLPによる次元圧縮においては、いかにして E^* の小さいMLPを構成するかが問題となる。特に、S^* の次元数 L が未知の場合には、E^* が小さくなるように曲面 S の次元つまり中間層の素子数 K を選択することが必要となる。

2.2 平岡・吉澤の基準

砂時計型MLPにおいて最適な中間次元数を決定するという問題は一種のモデル選択の問題であるから、AICやMDLといった統計基準を用いることも考えられる。しかし、平岡らは砂時計型MLPに対してはこれらの基準が不適切であることを指摘し、上記の E^* の推定値を求めてこれを基準に次元数を選択するという新たな方法を提案している[7]。以下、その導出過程を簡単に示す。

一般の非線形の場合を解析するのは困難であるから、S や S^* が平面とみなせるものとして線形の場合を考え、真値 $\boldsymbol{z}$ の共分散行列の固有値を大きいものから順に

$$\lambda_1 \geq \ldots \geq \lambda_L > \lambda_{L+1} = \cdots = \lambda_N = 0 \tag{6}$$

とおく。このとき、$K \geq L$ で S が S^* を包含するならば、

$$\begin{aligned} E &= (N-K)\sigma^2 && (7)\\ E^* &= K\sigma^2 && (8) \end{aligned}$$

が成り立つ。一方、$K < L$ の場合、S が S^* の上位 K 個の固有ベクトルの張る部分空間に一致するならば、

$$\begin{aligned} E &= \lambda_{K+1} + \cdots \lambda_L + (N-K)\sigma^2 && (9)\\ E^* &= \lambda_{K+1} + \cdots \lambda_L + K\sigma^2 && (10) \end{aligned}$$

が成り立つ。これらの関係式をもとに、E^* の推定値を構成する。まずノイズの分散 σ^2 が既知の場合、E^* の推定値 E_{known} は次式で与えられる。

$$E_{known} = E + (2K-N)\sigma^2 \tag{11}$$

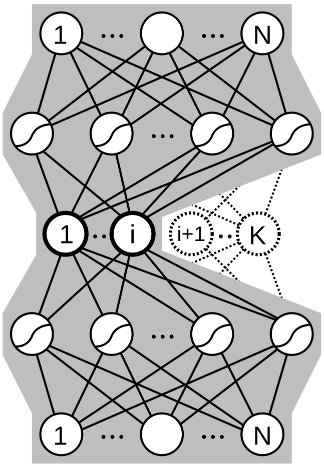


図 2: 5層砂時計型MLPと部分パーセプトロン

次に、σ^2 が未知の場合、式(7)からこれを

$$\tilde{\sigma}^2 = \frac{1}{N-K}E \tag{12}$$

と推定し、これを式(11)に代入すれば

$$E_{unknown} = \frac{K}{N-K}E \tag{13}$$

が得られる[1]。

いずれの推定値も、σ^2 の大きさに対する妥当な仮定の下で、$K = L$ において最小値をとることが示されている[2]。したがって、様々な中間次元数のMLPで学習を行い、それらの E^* の推定値を求めてやれば、最適な中間次元数をもつMLPを選択することができる。

3 重畳エネルギーへの平岡・吉澤の基準の適用

重畳エネルギーについて概説し、これを用いた砂時計型MLPの恒等写像学習に対する平岡・吉澤の基準の適用法を検討する。

3.1 重畳エネルギー

---

[1] σ^2 が未知の場合の E^* の推定法としてモンテカルロ法を用いる方法も検討されている[7]が、ここでは説明を省略する。

[2] ただし、$E_{unknown}$ は式(7)の成立を仮定して導かれているため、$K < L$ においては E^* とのずれが大きくなる。3.2節でこの問題について議論する。

重畳エネルギーは、MLPの中間層素子のうち一部のみを使用したときのMLPの出力誤差を考え、様々な素子数で得られる誤差を加え合わせたものである。図2のMLPにおいて、K個の中間層素子のうち1番目からi番目$(i = 1, 2, \ldots, K)$までを用いて構成される部分的なパーセプトロンを第i部分パーセプトロンと呼ぶ。第i部分パーセプトロンによる恒等写像の近似写像をf_iと表し、入力$\boldsymbol{x}$に対する出力を$f_i(\boldsymbol{x})$と表記する。このとき、式(1)のデータ集合に対してすべての素子を用いたときのMLPの誤差が式(2)で与えられるのと同様に、各部分パーセプトロンについても平均二乗誤差

$$E_i = \frac{1}{T}\sum_{t=1}^{T} ||\boldsymbol{x}_t - f_i(\boldsymbol{x}_t)||^2 \qquad (14)$$

を考えることができる[3]。重畳エネルギーFは、このE_iの重みつき和

$$F = \sum_{i=1}^{K} \beta_i E_i \qquad (15)$$

として定義される。ただし、β_iは適当な正の係数である。重畳エネルギーを用いると、データの近似の際に重要な成分ほど上位の中間層素子の出力に現れるようになり、寄与の大きさの順にならんだデータ表現が自己組織される。特に線形素子から成る3層MLPの恒等写像学習に適用した場合、i番目の中間層素子の出力がデータ集合のi番目の主成分に一致することが証明されている[4]。

3.2 平岡・吉澤の基準の適用

重畳エネルギーを用いた場合、すべてのE_iが小さくなるように学習が行なわれる。したがって、平岡・吉澤の基準を適用する最も単純なアプローチとしては、各部分パーセプトロンに対して式(11)あるいは(13)をそのまま適用し、$i = 1, 2, \ldots, K$の中から最も基準量の値が小さくなる中間次元数を選択する方法が考えられる。この場合、第i部分パーセプトロンに対する基準量は

$$\begin{aligned} E_{known}(i) &= E_i + (2i - N)\sigma^2 && (16) \\ E_{unknown}(i) &= \frac{i}{N - i}E_i && (17) \end{aligned}$$

と表せる。ノイズ分散σ^2が既知でありE_{known}を用いる場合については、この方法でも良好な結果が得られる。しかし、σ^2が未知の場合、通常のエネルギー関数を用いる際と共通の現象として、$i < L$においてE^*と$E_{unknown}$とのずれが大きくなるという問題がある。そこで、重畳エネルギーを用いて得られるパーセプトロンの性質を利用して$E_{unknown}$を修正する。

上述の$E_{unknown}$のずれは、σ^2を次元数iの大小にかかわらず

$$\tilde{\sigma}(i) = \frac{1}{N - i}E_i \qquad (18)$$

より推定していることに起因している。この場合、誤差E_iに関していずれのiにおいても式(7)が成立することを仮定しているが、実際には$i < L$では式(9)のようになる。そのため、$\tilde{\sigma}^2(i)$がσ^2を大きく見積もりすぎて$E_{unknown}$のずれが大きくなると考えられる。E_Kのみを最小化する従来の学習ではこのようなノイズ分散の過大評価を回避することは困難である。しかし、重畳エネルギー関数を用いる場合、任意の次元数iで部分パーセプトロンの真の誤差$E^*(i)$を推定する際に、$i = 1, 2, \ldots, K$のすべてのE_iの値を利用することができる。そこで、(18)にかえてσ^2を式

$$\tilde{\sigma}_{min} = \min_i \frac{1}{N - i}E_i \qquad (19)$$

より推定し、式(17)を次のように修正する。

$$E_{unknown}(i) = E_i + (2i - N)\tilde{\sigma}^2_{min} \qquad (20)$$

これによってノイズ分散が未知の場合の推定値が改善されることが期待できる。

3.3 係数β_iの設定

次に、上記の基準量を式(15)の係数β_iを適切に設定するための指針として利用することを考える。線形MLPの恒等写像学習の場合、β_iは正の定数であれば主成分分析と同一の結果が得られることが理論的に保証される[4]。しかし、非線形素子を含むMLPの場合、必要な学習回数などの学習効率がβ_iに左右される。本論文では、β_iの値を学習時に適応的に変更することを試みる。

予備的な実験の結果、真の次元数L以上の中間次元数をもつ部分パーセプトロンの係数β_iに関し

---

[3] 第K部分パーセプトロンはもとのパーセプトロン全体に一致し、その誤差E_Kが通常の誤差逆伝搬学習で用いられるエネルギー関数である。

ては、これらをすべて同一の値としても学習効率に影響のないことが確認できた。そこで、式(8)および(10)を考慮し、次のようなβ_iの設定法を考える。

- $E^*(i)$が$i\sigma^2$に近い場合

  このような部分パーセプトロンでは式(8)の関係が成立している、すなわち$i \geq L$であると考え、それらのβ_iの値を同一に設定する

- $E^*(i)$が$i\sigma^2$に比べて大きい場合

  式(10)の関係が成立している、すなわち$i < L$と考え、β_iを上記の値より小さくする

これを実現するためには、例えば

$$\beta_i \propto i/E^*(i) \tag{21}$$

とすればよい。実際には、E^*を適当な推定値で置き換えることになる。以降の計算機実験においては、

$$\beta_i \propto i/E_{unknown}(i) \tag{22}$$

および

$$\sum_{i}^{K} \beta_i = 1 \tag{23}$$

を満足するようにβ_iを設定するものとした。

4 実験

通常のエネルギー関数を用いた場合(従来型と呼ぶ)と重畳エネルギーを用いた場合(重畳型と呼ぶ)について、計算機実験を行なった結果を示す。

4.1 実験条件

実験には、文献[7]と同じくマニピュレータの姿勢を表すデータを用いた。表1に示すように、データの次元数$N = 4$、真の次元数Lは$1, 2, 3$の三通りである。これをもとに$\theta_1, \theta_2, \theta_3$の値を一様乱数として$z_t$を生成し、さらに$\sigma^2 = 0.0025$の正規乱数を加えて$\boldsymbol{x}_t$を作る。データ数$T = 1000$とし、$\{\boldsymbol{x}_t\}_{t=1}^T$を5層砂時計型MLPに与えて恒等写像学習を行う。MLPの素子数は入力層から順に

$$N - 16 - K - 16 - N$$

とし、第2層と第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た。重畳型の実験では、学習後に最適な中間次元数を選択できることを確認するため、中間次元数はLによらずすべて$K = 3$とした。一方従来型の場合、三種のLのそれぞれに対して$K = 1, 2, 3$で実験を行った。その他の実験条件は両者で同一である。学習回数は$5{,}000$回($L = 1, 2$)および$20{,}000$回($L = 3$)とした。また、重畳型の係数β_iについては、最初の500回の学習の間はすべて同一の値$\beta_i = \frac{1}{K}$に固定し、以降100回毎に式(22)および(23)を適用して変更するものとした。

表 1: 実験用データの真の関係

$$\boldsymbol{z} = [z_1, z_2, z_3, z_4]^T$$

$$L = 1 \quad \begin{cases} z_1 = 0.4\cos\theta_1 \\ z_2 = 0.4\sin\theta_1 \\ z_3 = z_1 + 0.4\cos(\theta_1 + \theta_1) \\ z_4 = z_2 + 0.4\sin(\theta_1 + \theta_1) \end{cases}$$

$$L = 2 \quad \begin{cases} z_1 = 0.4\cos\theta_1 \\ z_2 = 0.4\sin\theta_1 \\ z_3 = z_1 + 0.4\cos(\theta_1 + \theta_2) \\ z_4 = z_2 + 0.4\sin(\theta_1 + \theta_2) \end{cases}$$

$$L = 3 \quad \begin{cases} z_1 = 0.3\cos\theta_1 + 0.3\cos(\theta_1 + \theta_2) \\ z_2 = 0.3\sin\theta_1 + 0.3\sin(\theta_1 + \theta_2) \\ z_3 = z_1 + 0.3\cos(\theta_1 + \theta_2 + \theta_3) \\ z_4 = z_2 + 0.3\sin(\theta_1 + \theta_2 + \theta_3) \end{cases}$$

ただし $\theta_1, \theta_2, \theta_3 \in [0, \pi)$

4.2 実験結果

以上の条件で、$L = 1, 2, 3$の三通りで入力データに加えるノイズの値およびMLPの重みの初期値を変えながら10試行(従来型では$K = 1, 2, 3$のそれぞれで10試行づつ計30試行)の学習を行った。学習後のMLPを用いて計算した真の誤差E^*およびその推定値の10試行の平均を図3および4に示す。ただし、誤差E^*の計算の際には、学習に用いたデータとは別に真値の集合$\{\hat{z}_t\}_{t=1}^T$およびデータ

表 2: 学習終了時の β_i の値

L	β_1	β_2	β_3
1	0.334527	0.356336	0.309138
2	0.021528	0.505350	0.473122
3	0.041326	0.358627	0.600048

集合 $\{\hat{\boldsymbol{x}}_t\}_{t=1}^{T}$ を用意し、$\hat{\boldsymbol{x}}_t$ を入力したときの MLP 出力 $\hat{\boldsymbol{y}}_t$ と $\hat{\boldsymbol{z}}_t$ との誤差から

$$E^* = \frac{1}{T}\sum_{t=1}^{T} ||\hat{\boldsymbol{x}}_t - \hat{\boldsymbol{y}}_t||^2 \qquad (24)$$

とした。また、重畳型の $E_{unknown}$ は式 (20) より算出した値を示す。さらに、E_{MC} は文献 [7] に示されたモンテカルロ法による推定値を表す。重畳型の場合、$K = 3$ で学習を行ったにもかかわらず三種の基準量 $E_{known}(i)$,$E_{unknown}(i)$,$E_{MC}(i)$ はいずれも $i = L$ で最小となっており、真の次元数 L の推定に成功している[4]。この結果は、従来型では K を変えながら複数の MLP で学習を行なう必要があるのに対して、重畳型では一度の学習結果から最適な中間次元数を選択できることを示している。また、$i < L$ においても $E_{unknown}(i)$ は $E^*(i)$ の良い推定となっており、式 (19) および (20) を用いたことによる $E_{unknown}$ の改善の効果が見られる。一方、表 2は重畳型の学習終了時点における β_i の値の 10 試行の平均を示している。$E^*(i)$ の大きさに対応して、それぞれの条件に適した係数値が得られていることがわかる。

5　むすび

本論文では、砂時計型 MLP の恒等写像学習による次元圧縮において最適な中間次元数を選択するという問題に対して、重畳エネルギー関数と平岡・吉澤の基準を用いる手法を検討し、計算機実験によってその有効性を示した。しかし、重畳エネルギーを用いる学習は従来の MLP の学習と比較して計算コストが大きくなる。この点を改善し、より計算量の少ない手法を検討することが今後の課題となる。

謝辞

本研究の遂行にあたり数々の有益な助言を頂いた埼玉大学の平岡和幸氏に感謝致します。

参考文献

[1] Baldi, P. and Hornik, K., “Neural networks and principal component analysis,” Neural Networks, vol.2, pp.53–58, 1989.

[2] Usui, S., Nakauchi, S. and Nakano, M., “Internal color representation acquired by a five-layer neural network,” Proc. Int. Conf. Artificial Neural Networks, vol.1, pp.867–872, 1991.

[3] DeMers, D. and Cottrell, G., “Non-linear dimensionality reduction,” Advances in Neural Information Processing Systems 5, pp.580–587, 1993.

[4] 高橋隆史, 徳永隆治, “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,” 信学論, vol.J80–D–II, no.5, pp.1267–1275, 1997.

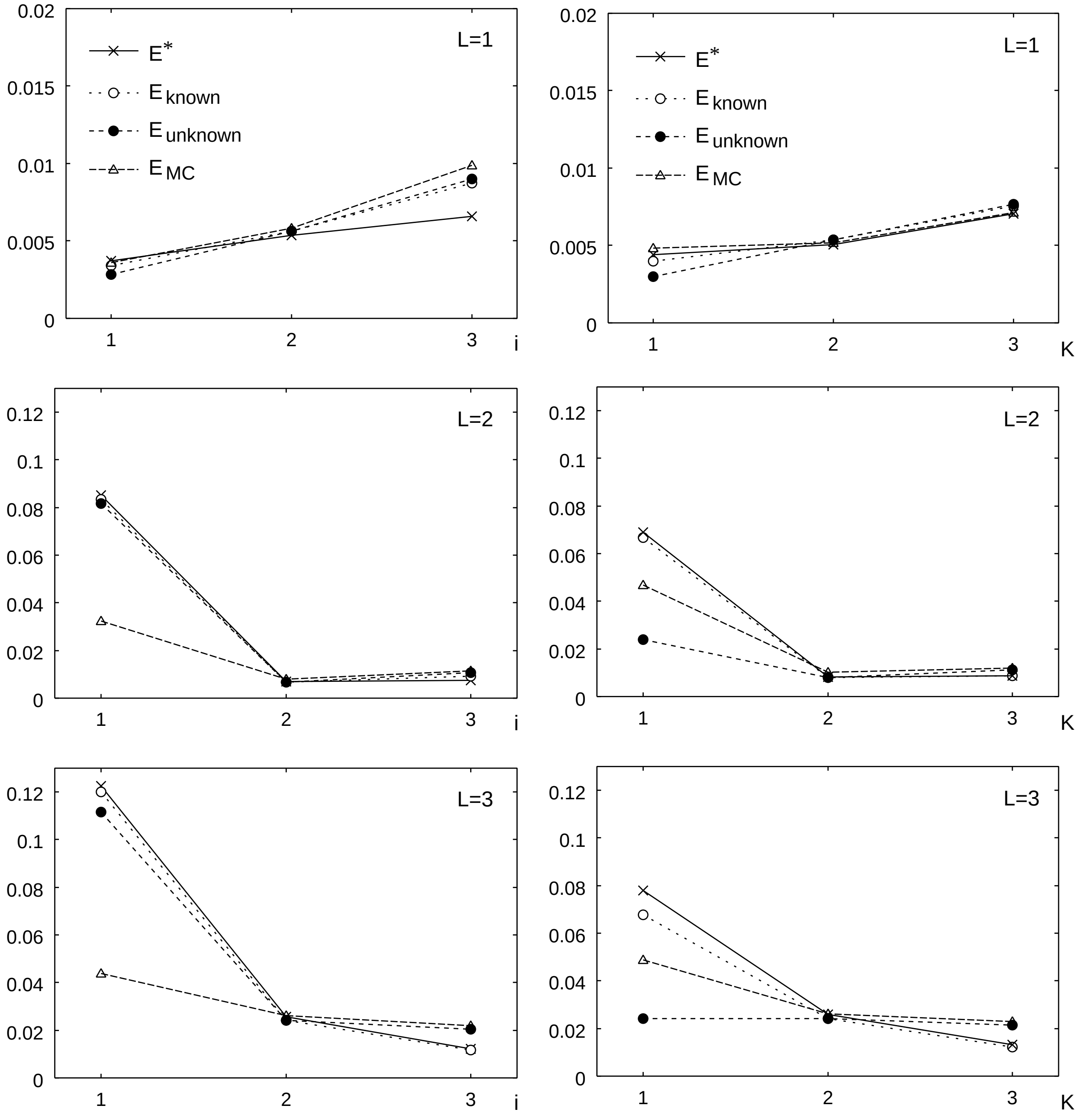
[5] 高橋隆史, 徳永隆治, “重畳エネルギー関数による多層パーセプトロンの冗長性削減,” 信学論, vol.J80-D-II, no.9, pp.2532–2540, 1997.

[6] 平岡和幸, 吉澤修治, “データ補正の教師無し学習,” 信学技報, NC97-4, 1997.

[7] 平岡和幸, 吉澤修治, “砂時計型ネットの中間次元数選択について,” 信学技報, NC97-139, 1998.

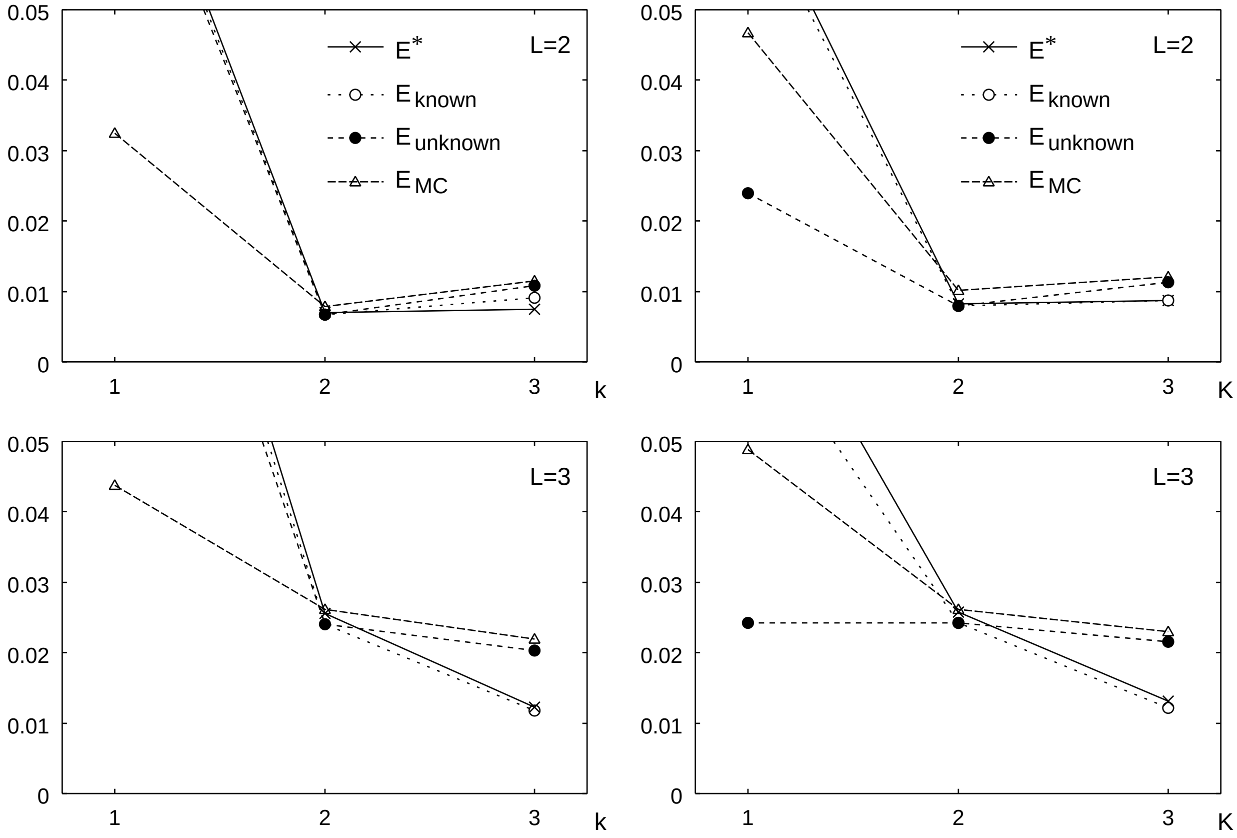
---

[4] $L = 2$ の実験では中間次元数が 2 の場合と 3 の場合とで E^* や基準量の差が小さく見えるが、試行毎の値のばらつきが小さいため実際には十分な差が得られている



(a) 重畳型　　　　(b) 従来型

図 3: 真の誤差 E^* およびその推定値



(a) 重畳型

(b) 従来型

図 4: 真の誤差 E^* およびその推定値(図 3 の $L = 2, 3$ のグラフを拡大したもの)