修士学位論文

パターン認識のための特徴抽出法に関する研究

東北大学大学院工学研究科 電気・通信工学専攻
佐々木 裕児

目 次

図目次

表 目 次

第1章

序論

1.1　研究の背景

1.1.1　パターン認識とは

パターン認識とは，観測されたパターンをあらかじめ定められた複数の概念のうちの1つに対応させる処理である [1]．この「概念」をクラスと呼んでいる．例えばアルファベットの認識であれば、入力パターンを 26 個のクラスのいずれかに対応させる処理ということになる．パターンというと人間の視覚に入ってくる2次元のパターンを思い浮かべるかもしれないが，パターン認識で扱う対象は広い．例えば，音声のような時系列信号を処理して 50 音や単語に対応させる音声認識もパターン認識の 1 分野であるし，心電図の波形を分析して異常の有無を判定するのも同様である．

このようなパターン認識の技術は文字読み取り装置や音声認識装置などに応用され，多少の制約はあるもののすでに実用化されてさまざまな分野で使われている．最近では，インターネットにおける情報検索システムや，ITS における高速で高精度な道路状況判断システムなど，マルチメディア時代を迎えるにあたり、種々のメディアを効率よく処理しなくてはならない機会が急速に増えつつあり，パターン認識技術のさらなる高度化に対する要求や期待が高まってきている．このようにパターン認識の研究は今後もますます活発になっていき，その重要性も増してゆくものと思われる．

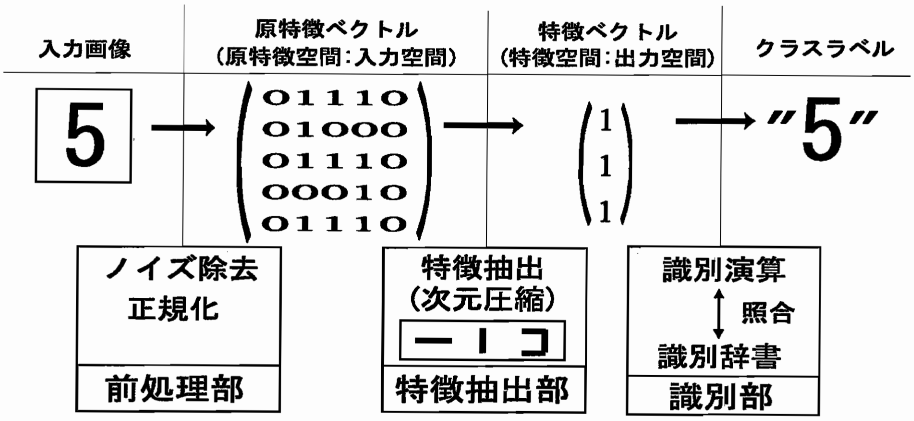


図 1.1: **認識系の流れ**．入力パターンが入力されると，前処理が行われ原特徴ベクトルが生成される．次に原特徴ベクトルから本質的な特徴を抽出し，これを特徴ベクトルとする．この特徴ベクトルを基に入力パターンの所属するクラスラベルを出力する．

1.1.2　パターン認識系の構成

計算機で「パターン認識系」を構成する場合，一般に図 1.1 の形をとる．パターンが入力されるとまず前処理部でノイズ除去，正規化などの処理を行い，これ原特徴ベクトルとする．続いて特徴抽出部では，膨大な情報をもつ原特徴ベクトルから識別に必要な本質的な特徴のみを抽出し，これを特徴ベクトルとする．本論分では，原特徴ベクトルの張る空間を原特徴空間，特徴ベクトルの張る空間を特徴空間と呼ぶことにする．この特徴ベクトルをもとに識別部では識別処理を行う．識別処理は入力パターンに対して複数のクラスのうちの１つを対応させせることによって行われる。そのため，あらかじめ識別辞書を用意し，抽出された特徴をこの辞書と照合することにより入力パターンの所属するクラスラベルを出力する．本論文ではこの辞書照合の部分を識別と呼び，パターンが入力されてから出力されるまでの前処理，特徴抽出処理，識別処理を総称して認識と呼ぶことにする．

1.1.3　特徴抽出

高次元入力パターンの各次元はしばしば相関があり，これらのパターンの本質的次元数は入力次元数よりもずっと低いことが多い．適切な次元圧縮を用いることで，パターンから余分な情報を除去し，より信頼でき

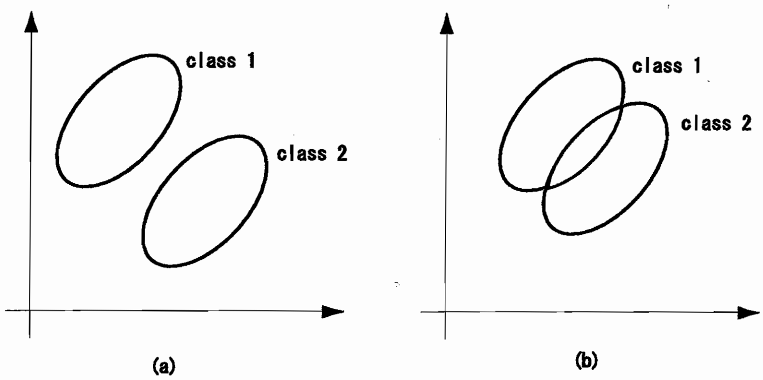


図 1.2: **特徴空間における分布**. 二次元特徴空間に2クラスが分布している場合を考える. (a) では分布の重なりが生じておらず, 適切な識別器を設計すれば分離することができる. 一方 (b) では, 分布の重なりが生じてしまっているため, どのような識別器を設計しようとも必ず誤認識されてしまう. 識別に有利な特徴空間は明らかに (a) である.

る識別が可能になる. いくつかの識別器において, とくに最近傍識別器 (k-NN) のような余分な情報の影響を受けやすい識別器では, 次元圧縮によってしばしば識別性能を向上できる. 従って, 高次元パターンをそのまま識別するよりも低次元部分空間に落として識別したほうが優れた識別性能を示すことが少なからずある. このように, ある基準にのっとって特徴空間の次元圧縮を行うことを特徴抽出と呼ぶ.

識別系の中でも特徴抽出は, 識別性能を左右する極めて重要な処理である. 例えば2次元特徴空間上に2つのクラスが分布している場合を考える. クラスの分布を特徴空間上で観測したところ図 1.2 (a) のようになったとする. この場合, 2つのクラスは完全に分離されているから, 識別部を適切に設計すれば誤識別を起こさない認識形を実現できるはずである. 一方, クラスの分布が図 1.2 (b) のようになったとする. この場合はクラスの分布間に重なりがあるため, 識別部をどのように設計しても誤認識が生じてしまう. このように, 適切な特徴空間を求めることができなければ識別部設計にいかに力を注いでも高精度の認識系は実現できない.

1.1.4　次元圧縮法

次元圧縮法は，教師あり，教師なしの 2 種類に大別できる．つまり，部分空間を構築するときにクラスラベルを用いるか否かである．最も知られた教師なし次元圧縮方は主成分分析 (PCA) であろう．PCA は二乗誤差最小の意味で最適なパターン表現法である．よって識別という意味では最適ではない．ほかの教師なし次元圧縮法としては独立成分解析 (ICA)[2]，自己組織化マップ (SOM) などがあるが，これらも同様である．

一方，教師あり次元圧縮法はクラスラベルを用いるので，識別により適した部分空間を構築できる．例えば，フィッシャーの線形判別分析は広く用いられている手法で，クラス間の分布を最も分離させるような，クラス内の分布をもっともまとまらせるような部分空間を求める．しかし，多クラス問題に適用したときにはフィッシャーの線形判別分析により求まった部分空間が，クラス間の分離という点で必ずしも十分な判別能力を持たないことがある．

それに対し Lotlikar らは，パターン認識における理想的な基準であるベイズエラーを用いて識別に最適な部分空間を求める，適応的線形次元圧縮法 [3] を提案した．実験により，フィッシャーの線形判別分析よりも識別に適した部分空間への線形写像を求めることに成功した．しかし，実問題で線形分離可能な場合は稀であり，線形な次元圧縮法では対応できない問題も多い．

1.2　本論文の目的

本研究では，適応的次元圧縮法 [3] を非線形に拡張することにより，パターン認識に最適な非線形な次元圧縮を行う方法を提案する．実際には高次元なヒルベルト空間へ非線形写像した像を，適応的次元圧縮法により低次元部分空間に線形写像することによって，非線形な次元圧縮を実現する．

1.3　本論文の構成

本論文の構成は以下のとおりである．

第 1 章 序論

序論であり，本研究の背景や目的を述べる．

第 2 章　適応的線形次元圧縮法

ベイズエラーを最小とする部分空間への線形写像を求める適応的線形次元圧縮法について述べ，フィッシャーの線形判別分析よりも識別に適した部分空間を求めることを実験的に示す.

第 3 章　適応的非線形次元圧縮法

第 2 章で述べた適応的線形次元圧縮法を非線形に拡張する方法について述べる．このとき，ヒルベルト空間における入力パターンの像の分布に基づいて 2 通りの手法を提案する.

第 4 章　評価実験

第 4 章で提案した手法の有効性を公開データベースを用いて確認する.

第 5 章　結論

結論であり本研究のまとめ，今後の課題を述べる.

第2章

適応的線形次元圧縮法

2.1 初めに

序論ではパターン認識のための特徴抽出, 次元圧縮の有効な方法として, 適応的線形次元圧縮法を紹介した. 本章では, 適応的次元圧縮法の評価基準であるベイズエラーについて述べ, ベイズエラーを最小にする部分空間への線形写像の学習方法を紹介する. そして実際に人工データ, 実データを用いて認識実験を行い, フィッシャーの線形判別分析よりも認識に適した特徴空間が求まることを示す.

2.2 ベイズエラー

ベイズエラーとは特徴量そのものの不完全さに起因する必然的な誤りであり, 識別器を工夫することでは解決できないものである. これは特徴空間上における分布の重なりとも考えられる (図 1.2 参照). 分布の重なりが生じている以上, 誤識別をするのは当然である.

ここで, ω_1 と ω_2 の 2 つのクラスに識別する 2 クラス問題を考える. ω_1 と ω_2 の事後確率をそれぞれ, $P(\omega_1|\boldsymbol{x})$, $P(\omega_2|\boldsymbol{x})$ とする. 観測値 $\boldsymbol{x}$ が誤っ

て識別される確率，誤り確率 $P_e(\boldsymbol{x})$ は次のようになる．

$$P_e(\boldsymbol{x}) = \begin{cases} \boldsymbol{x} \in \omega_1 \text{ と判定した時} & P(\omega_2|\boldsymbol{x}) \\ \boldsymbol{x} \in \omega_2 \text{ と判定した時} & P(\omega_1|\boldsymbol{x}) \end{cases} \tag{2.1}$$

そして，起こりうるすべての$\boldsymbol{x}$に対する平均誤り確率 P_e は

$$P_e = \int P_e(\boldsymbol{x})\, p(\boldsymbol{x}) d\boldsymbol{x} \tag{2.2}$$

となる．この場合，観測された $\boldsymbol{x}$ に対してベイズ決定則では次のような判定を行う．

$$\begin{cases} P(\omega_1|\boldsymbol{x}) > P(\omega_2|\boldsymbol{x}) \Rightarrow \boldsymbol{x} \in \omega_1 \\ P(\omega_2|\boldsymbol{x}) > P(\omega_1|\boldsymbol{x}) \Rightarrow \boldsymbol{x} \in \omega_2 \end{cases} \tag{2.3}$$

ベイズ決定則は平均誤り確率 P_e を最小にするものであり，ベイズ決定則により識別を行うとすると，式 (2.1) は次のようになる．

$$P_e(\boldsymbol{x}) = \min\{P(\omega_1|\boldsymbol{x}), P(\omega_2|\boldsymbol{x})\} \tag{2.4}$$

式 (2.2) に式 (2.4) を代入すると平均誤り確率 P_e は

$$P_e = \int \min\{P(\omega_1|\boldsymbol{x}), P(\omega_2|\boldsymbol{x})\} p(\boldsymbol{x}) d\boldsymbol{x} \tag{2.5}$$

と書き換えることができる．この P_e はこれ以上小さくできない誤り確率の限界であり，特徴量が本質的に持っている必然的な誤り確率である．これを**ベイズエラー**と呼ぶ．

これまで 2 クラスの場合を述べてきたが，これは多クラスの場合にも拡張できる．多クラスの場合のベイズ決定則は

$$\max_{i=1,2,\cdots,c} \{P(\omega_i|\boldsymbol{x})\} = P(\omega_k|\boldsymbol{x}) \Rightarrow \boldsymbol{x} \in \omega_k \tag{2.6}$$

となる．c はクラス数である．この時のベイズエラーは

$$P_e = 1 - \int \max_{i=1,2,\cdots,c} \{P(\omega_i|\boldsymbol{x})\} p(\boldsymbol{x}) d\boldsymbol{x} \tag{2.7}$$

と表される．

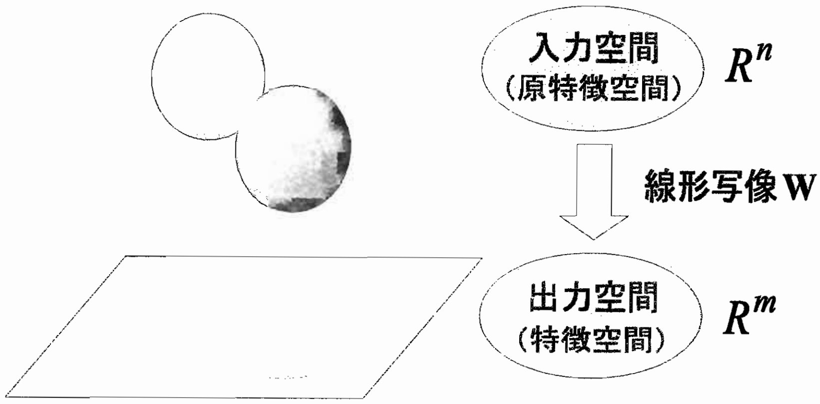


図 2.1: LL **法の問題の設定**. LL 法は入力空間 R^n からベイズエラーが最小となるような出力空間 R^m $(m \leq n)$ への線形写像 $\boldsymbol{W}$ を求める手法である. 図は 3 次元原特徴空間から 2 次元特徴空間への写像を行った例を表しており, この写像では特徴空間における 2 つのクラスの分布に重なりが生じている. LL 法ではこの重なり, すなわちベイズエラーを最小化するような線形写像を求める.

2.3　適応的線形次元圧縮法

適応的線形次元圧縮法は, ベイズエラーが最小となるようにな特徴空間への線形写像を求める手法である (図 2.1). パターン認識系における原特徴空間を次元圧縮法への入力という意味で入力空間と呼ぶ. 同様に特徴空間を次元圧縮法からの出力という意味で出力空間と呼ぶ. 以下, 適応的線形次元圧縮法を LL 法 (Linear Lotlikar method) と呼ぶことにする.

2.3.1　LL 法のベイズエラー

任意の分布においてベイズエラーを解析的に表すことは困難である. LL 法では多クラス, R^n の入力空間を想定し, 入力パターンの分布に次の 2 つの仮定をおくことで, 出力空間におけるベイズエラーの定式化を行う.

（a） クラス ω_i $(i = 1, \cdots, c)$ に属するパターン $\boldsymbol{x}$ は次式に従う.

$$\boldsymbol{x} \sim N\left(\boldsymbol{\mu}_i, \sigma \boldsymbol{I}_{(n\times n)}\right) \tag{2.8}$$

但し, $\boldsymbol{I}_{(n\times n)} \in R^{(n\times n)}$ は単位行列, $N(\cdot)$ は正規分布の確率密度関数 (pdf) であり, $\boldsymbol{\mu}_i$ はクラス i の平均ベクトルである. つまり, どのクラスも等し

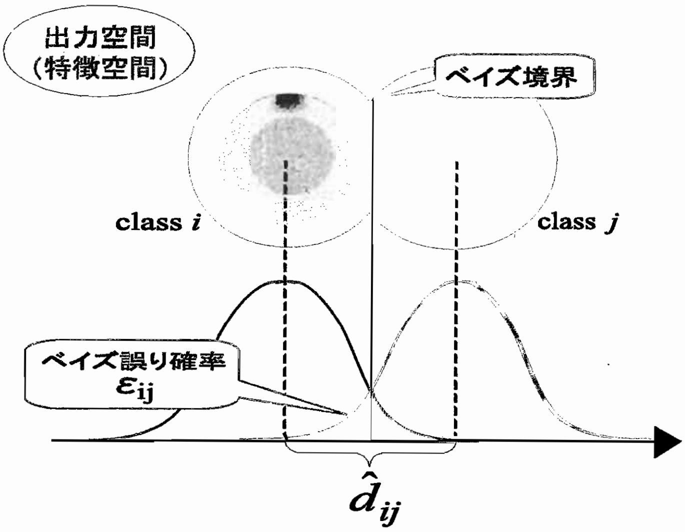


図 2.2: **出力空間における 2 クラス問題のベイズエラーとベイズ境界**. Lotlikar の仮定を満たすとき，ベイズ境界は超平面になり，その超平面は出力空間におけるそれぞれのクラスの平均 $\hat{\boldsymbol{\mu}}_i$，$\hat{\boldsymbol{\mu}}_j$ の中点 $\frac{\hat{\boldsymbol{\mu}}_i+\hat{\boldsymbol{\mu}}_j}{2}$ を通り，$\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j$ に直交する.

い共分散行列をもつ.

（b） クラスの事前確率が等しい

$$P(\omega_i) = \frac{1}{c} \tag{2.9}$$

但し，ω_i はクラスラベル，c はクラス数である.
以下，この 2 つの仮定を **Lotlikar の仮定** と呼ぶ.

LL 法の目的は m $(\leq n)$ 次元部分空間への線形写像 $\boldsymbol{W} \in R^{(n\times m)}$ を求めることである. 線形写像 $\boldsymbol{W}$ に $\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}$ なる制約条件を与える. すると共分散行列 $\sum = \sigma\boldsymbol{I}_{(n\times n)}$ の正規分布に従う確率変数の $\boldsymbol{W}$ による線形写像の結果は，共分散行列 $\hat{\sum} = \sigma\boldsymbol{I}_{(m\times m)}$ の正規分布に従う.

まず 2 クラス問題を考える. Lotlikar の仮定を満たすとき，ベイズ境界[1]は超平面になる. その超平面は出力空間におけるそれぞれのクラスの平均 $\hat{\boldsymbol{\mu}}_i$，$\hat{\boldsymbol{\mu}}_j$ の中点 $\frac{\hat{\boldsymbol{\mu}}_i+\hat{\boldsymbol{\mu}}_j}{2}$ を通り，$\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j$ に直交する (図 2.2). よって，クラス i のサンプルをクラス j と誤認識する確率 ϵ_{ij} は次のようになる.

$$\epsilon_{ij} = \frac{1}{\sqrt{2\pi}\sigma}\int_{\hat{d}_{ij}/2}^{\infty} \exp(-t^2/2\sigma^2)dt \tag{2.10}$$

[1]ベイズ境界：ベイズエラーを最小にする決定境界.

但し，$\hat{d}_{ij} = \| \hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j \|$，$\hat{\boldsymbol{\mu}}_i = \boldsymbol{W}^T \boldsymbol{\mu}_i$, $\hat{\boldsymbol{\mu}}_j = \boldsymbol{W}^T \boldsymbol{\mu}_j$ であり，$\hat{d}_{ij}$ は $\boldsymbol{W}$ 及び入力空間における平均 $\boldsymbol{\mu}_i$，$\boldsymbol{\mu}_j$ に依存する．ちなみに，入力空間において共分散行列が等しくない場合はベイズ境界は超平面にならず，分布の重なりを簡潔に表すことはできない．

3 クラス以上の場合のベイズ境界は複雑であり，解析が困難である．LL 法では各 2 クラスのペア間で誤識別確率 ϵ_{ij} を考えその和を目的関数 J とし，J を最小にする問題に置き換える．

$$J = 2\sum_{i=1}^{c}\sum_{j=i+1}^{c} \frac{1}{\sqrt{2\pi}\sigma} \int_{\hat{d}_{ij}/2}^{\infty} \exp(-t^2/2\sigma^2)dt \tag{2.11}$$

2.3.2　学習方法

目的関数 J を最小にする線形写像 $\boldsymbol{W}$ を勾配法により求める．目的関数 J の $\boldsymbol{W}$ に対する勾配は次式のようになる．

$$\frac{\partial J}{\partial \boldsymbol{W}} = -\sum_{i=1}^{c}\sum_{j=i+1}^{c} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(\hat{d}_{ij}/2)^2}{2\sigma^2}) \frac{\partial \hat{d}_{ij}}{\partial \boldsymbol{W}} \tag{2.12}$$

ここで，入力空間におけるクラス間の差を $\boldsymbol{v}_{ij} = \hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_j$ とおくと,$\hat{d}_{ij}$ は

$$\hat{d}_{ij} = \| \boldsymbol{W}^T \boldsymbol{v}_{ij} \| = \sqrt{\boldsymbol{v}_{ij}^T \boldsymbol{W}\boldsymbol{W}^T \boldsymbol{v}_{ij}} \tag{2.13}$$

と表され，

$$\frac{\partial \hat{d}_{ij}}{\partial \boldsymbol{W}} = \frac{1}{2\hat{d}_{ij}} \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^T \boldsymbol{W} \tag{2.14}$$

となる．式 (2.12) と式 (2.14) より次式が得られる．

$$\frac{\partial J}{\partial \boldsymbol{W}} = -\left(\sum_{i=1}^{c}\sum_{j=i+1}^{c} \exp(-\frac{(\hat{d}_{ij}/2)^2}{2\sigma^2}) \frac{1}{2\hat{d}_{ij}} \boldsymbol{v}_{ij}\boldsymbol{v}_{ij}^T \right) \boldsymbol{W} \tag{2.15}$$

但し，$\frac{1}{2\sqrt{2\pi\sigma^2}}$ は定数であるので省略した．$\boldsymbol{W}$ の更新式は次のようになる．

$$\boldsymbol{W}_{new} = \boldsymbol{W}_{old} - \eta \frac{\partial J}{\partial \boldsymbol{W}} \tag{2.16}$$

但し，η は学習定数である．Lotlikar の仮定を満たすために，$\boldsymbol{W}$ の更新毎に $\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}$ を満たすように変換する．適応的部分空間定理により，学習定数 η が十分に小さければ，$\boldsymbol{W}$ は必ず収束することが保証されている．

結局，$\boldsymbol{W}$ の学習方法は以下の手続きにより行われる．

STEP1　$\boldsymbol{W}$ の初期値をランダムに定める．

STEP2　J の値が収束するまで以下を繰り返す．

substep1　$\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}$ を満たすため以下の変換で $\boldsymbol{W}$ を更新する．

$$\boldsymbol{W} \Leftarrow \boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{W})^{-\frac{1}{2}} \tag{2.17}$$

substep2　式 (2.16) により $\boldsymbol{W}$ を更新する．

但し，平方根の逆行列 $(\boldsymbol{W}^T\boldsymbol{W})^{-\frac{1}{2}}$ は，$\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{F}\boldsymbol{A}\boldsymbol{F}^T$ の固有値分解を用いて，$(\boldsymbol{W}^T\boldsymbol{W})^{-\frac{1}{2}} = \boldsymbol{F}\boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{F}^T$ と定義する．$\boldsymbol{A}$ は対角行列で，$\boldsymbol{A}^{\frac{1}{2}}$ は要素の平方根をとったものであり，$\boldsymbol{A}^{-\frac{1}{2}} = (\boldsymbol{A}^{\frac{1}{2}})^{-1}$ である．

2.4　評価実験

人工データと実データを使用して，LL 法によって求まった出力空間の評価を行う．比較として教師あり次元圧縮法として広く用いられている，フィッシャーの線形判別分析 (Linear Fisher method:以下 LF 法と呼ぶ) により求まった出力空間の評価もあわせて行う．

2.4.1　人工データを用いた評価実験

人工データによる次元圧縮実験を行い，LL 法と LF 法を比較する．実験条件は表 2.1 の通りである．入力空間における使用サンプルの分布を図 2.3 に示す．

表 2.1: 実験条件

使用サンプル	人工データ
クラス数	6
サンプル数	50 個/$class$
入力空間の次元数	$n = 3$
出力空間の次元数	$m = 2$
クラス ω_i の分布	$N(\boldsymbol{\mu}_i, 0.04\boldsymbol{I})$
比較手法	LL 法，LF 法

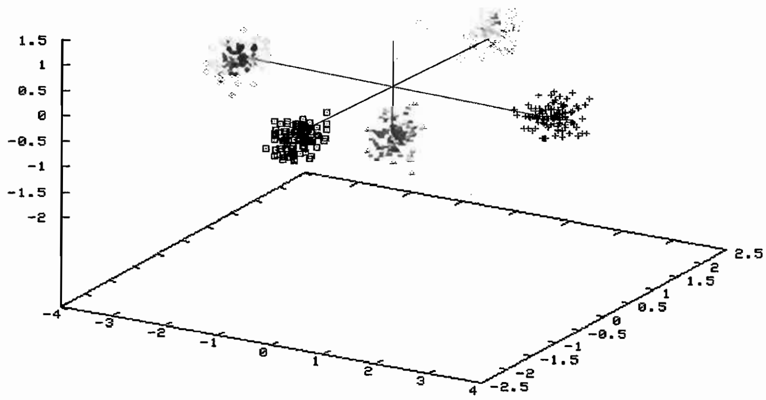


図 2.3: **入力空間における使用サンプルの分布**．色がクラスラベルを表す．それぞれのクラスの分布は等しい分散 $\sigma = 0.2$ をもつ正規分布である．各クラスの平均の，ゼロからの距離が赤色，緑色は 3，ピンク色，紫色は 2，黄色，水色は 1 である．

各クラスの平均は以下のようにおいた．

$$
\begin{aligned}
\boldsymbol{\mu}_1 &= [-3\ 0\ 0]^T \\
\boldsymbol{\mu}_2 &= [3\ 0\ 0]^T \\
\boldsymbol{\mu}_3 &= [0\ -2\ 0]^T \\
\boldsymbol{\mu}_4 &= [0\ 2\ 0]^T \\
\boldsymbol{\mu}_5 &= [0\ 0\ -1]^T \\
\boldsymbol{\mu}_6 &= [0\ 0\ 1]^T
\end{aligned}
$$

LL 法，LF 法を用いて求めた出力空間における使用サンプルの分布を図 2.5 に示す．分布中の実線は出力空間における各クラスの標準偏差の 2 倍の軸をもつ楕円である (95%のサンプルが円の内側に収まる)．結果は LF 法では黄色のクラスと水色のクラスが完全に重なってしまっているが，LL 法は完全に分離できている．

LF 法を多クラス問題に適用した際に，クラス間距離が大きい，もともと分離されているクラス間の距離が広がり，クラス間の距離が小さい，も

ともと分離されていないクラス間の距離が縮まるという問題点が指摘されている．本実験に用いたデータに対する実験結果より LF 法のこの問題点が確認された．一方，LL 法では識別に最適な出力空間が得られることが確認された．

2.4.2　実データを用いた評価実験

実データを用いた次元圧縮実験を行い，LL 法と LF 法の比較を行う．実データには，カルフォルニア州立大学アービン校の公開データベース[2]の一つである Image Segmentation Database を使用した．実験条件は表 2.2 の通りである．

表 2.2: 実験条件

使用サンプル	Image Segmentation Database
クラス数	7
訓練サンプル数	30 個/$class$
テストサンプル数	300 個/$class$
入力空間の次元数	$n = 18$
出力空間の次元数	$m = 2$
識別器	k-NN 法
比較手法	LL 法，LF 法

ハイパーパラメータである k-NN 法の k，Lotlikar の仮定における分散 σ は，5-fold 交差確認法により訓練サンプルより自動的に求めた．5-fold 交差確認法は以下の手順で行われる．

STEP1　訓練用サンプルをクラス内で 5 分割する．

STEP2　1 つのハイパーパラメータの組合せに対し 5 分割したうちの 1 つを評価用，4 つを訓練用として，5 通り学習，識別実験を行う．

STEP3　5 回の学習，識別実験の評価用サンプルに対する認識率の平均が最もよかったハイパーパラメータの組合せを採用する．

求まった出力空間の評価は，出力空間におけるテストサンプルの分布の重なり具合と，k-NN 法による認識率を用いて行う．注意すべきは前処理と

[2] http://www.ics.uci.edu/ mlearn/MLRepository.html

して次式で与えられるクラス内変動行列 $\boldsymbol{S}_w$ が単位行列となるようにサンプルを白色化しておくことである．(式 (2.8) を満たすためである)．

$$S_w = \sum_{i=1}^{c} \sum_{\boldsymbol{x}_j \in \omega_i} (\boldsymbol{x}_j - \boldsymbol{\mu}_i)(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T \tag{2.18}$$

但し，$\boldsymbol{x}_j$ はクラス ω_i に属するサンプルである．

表 2.3 に k-NN 法による認識率の一覧を示す．図 2.6 に出力空間におけるテストサンプルの分布を示す．分布中の実線は出力空間における各クラスの標準偏差の 2 倍の軸をもつ楕円である．LF 法では人工サンプルを用いた実験の考察で述べた通り，もともとクラス間の距離が近いクラスの分布の重なりが大きいのに対し，LL 法ではもともとクラス間距離の小さいクラス間を，分布の重なりの少ないより認識に適した出力空間へと射影していることが分かる．k-NN 法による認識でも，テストサンプルに対する認識率で 9.2%の差が現われた．

表 2.3: 実験結果 k-NN 法による認識率

	訓練サンプル	テストサンプル
LL 法	82.6%	82.8%
LF 法	79.1%	73.6%

2.5　まとめ

有効な教師あり次元圧縮法である LL 法の基準であるベイズエラーについて述べ，ベイズエラーを最小化する LL 法の学習則について述べた．識別実験により，LF 法よりも認識に適した出力空間が求まることを示した．

しかし，LL 法は線形な次元圧縮法であるため，非線形な分離境界を持つような分布に対して，識別に有効な出力空間が求めることができない．例として，2 次元入力空間における分布が 図 2.4 のようになる 2 クラス問題から，1 次元出力空間を求めることを考える．どのような 1 次元部分空間 (直線) への射影を考えようと，線形射影では出力空間において分布の重なりが生じてしまう．従って，より識別に適した出力空間を求めるためには，非線形な次元圧縮法が必要であると考えられる．

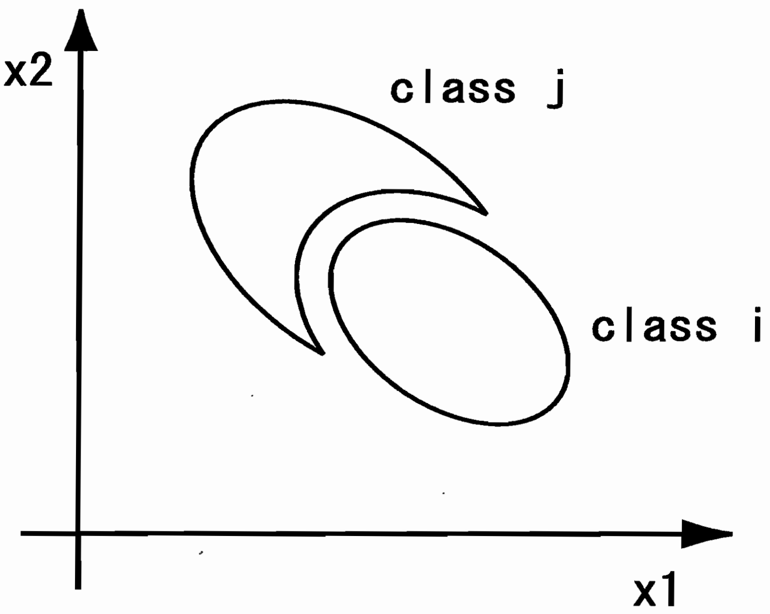
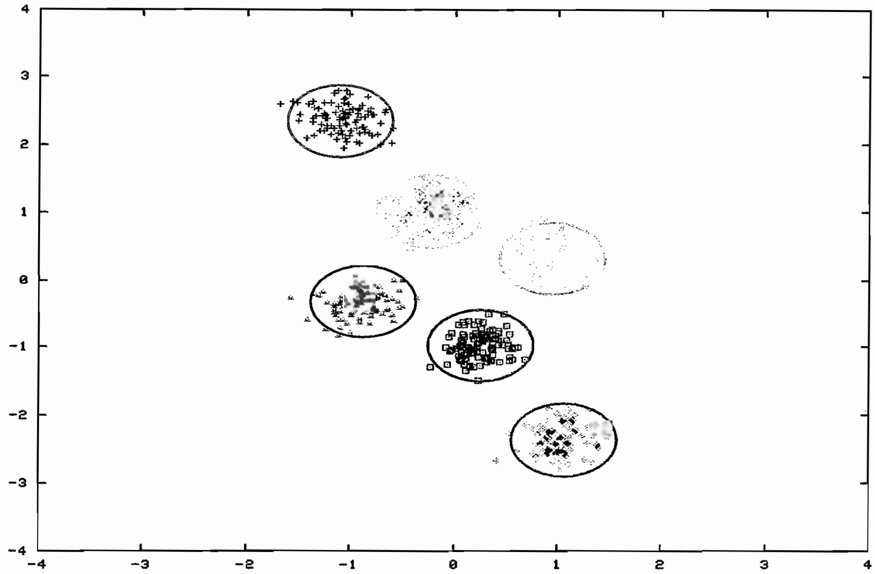
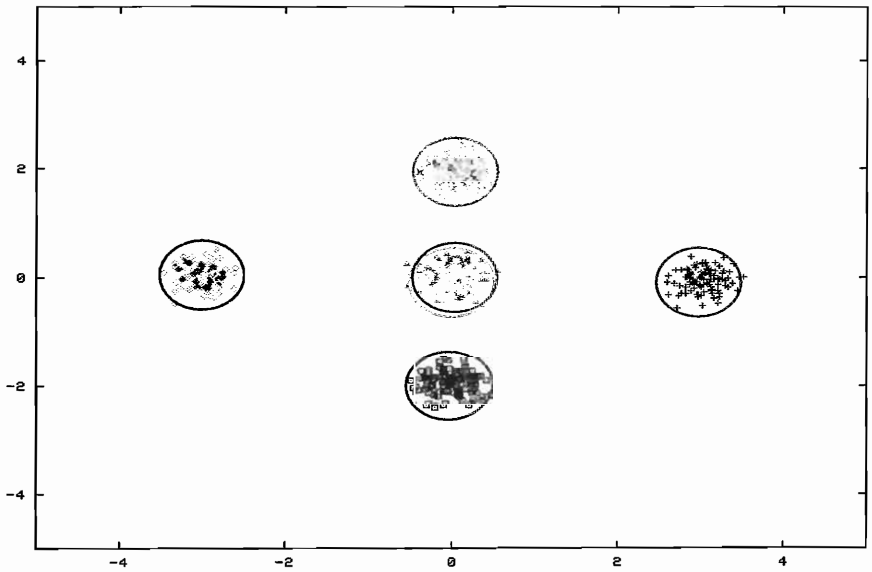


図 2.4: **入力空間において線形分離不可能な分布の例**．図のように非線形な分離境界面を持つ問題に対して LL 法を適用した場合，クラス間の距離をできるだけ離すような，x_1 から 135° 傾いた直線への線形写像が得られる．このような線形写像を行った場合，必ず分布の重なり，つまりベイズエラーが生じてしまう．

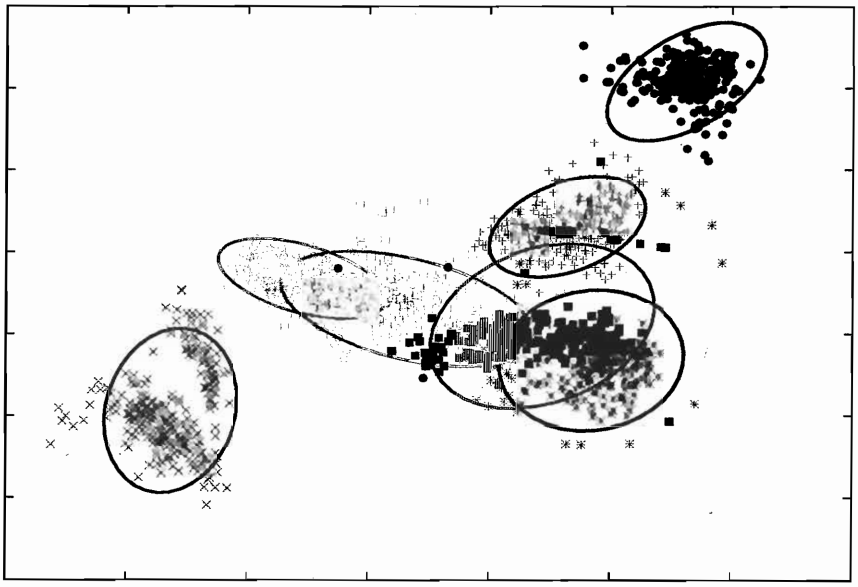


(a) LL法

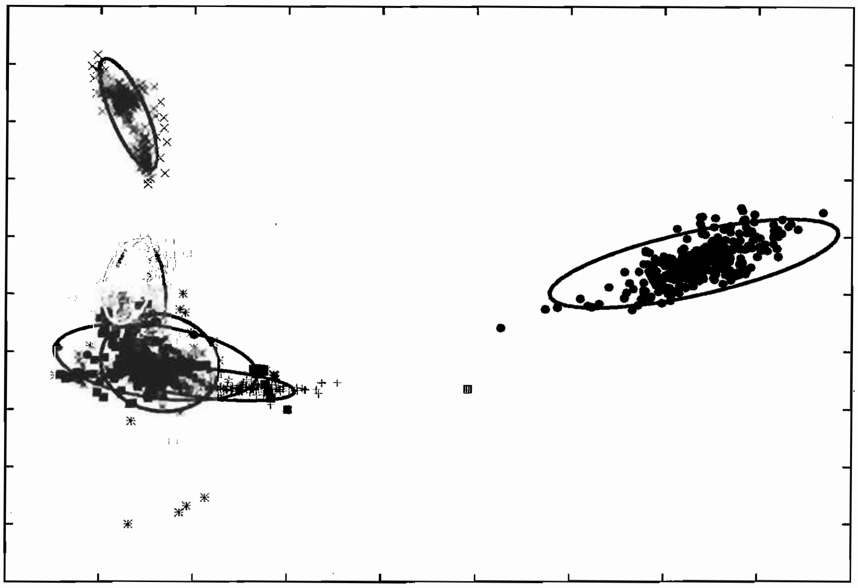


(b) LF法

図 2.5: LL 法，LF 法を用いて求めた出力空間におけるサンプルの分布．分布中の実線は出力空間における各クラスの標準偏差の 2 倍の軸をもつ楕円である (95%のサンプルが円の内側に収まる)．結果は LF 法では黄色のクラスと水色のクラスが完全に重なってしまっているが，LL 法は完全に分離できている．



(a) LL法



(b) LF法

図 2.6: Image Segmentation の出力空間におけるテストサンプルの分布．分布中の実線は出力空間における各クラスの標準偏差の 2 倍の軸をもつ楕円である．LF 法ではもともとクラス間の距離が近いクラスの分布の重なりが大きいのに対し，LL 法ではよりクラス間の分布の分離した認識に適した出力空間が得られているのが分かる．

第3章

適応的非線形次元圧縮法

3.1 はじめに

第2章ではLL法が線形次元圧縮法であるために，次元圧縮によって分布の重なりが生じてしまう場合があるという問題点を指摘した．本章ではその問題を解決するために，LL法を非線形に拡張した適応的非線形次元圧縮法を提案する．適応的非線形次元圧縮法は以下の手順によって行われる．流れ図を図3.1に示す．

STEP1 非線形変換によって原特徴ベクトルを原特徴空間よりもはるかに高次元なヒルベルト空間[4]へ非線形写像する．

STEP2 ヒルベルト空間における像にLL法を適用し，ヒルベルト空間からベイズエラーを最小にする出力空間への線形写像を求める．

LL法と提案手法である適応的非線形次元圧縮法は，**STEP1**を実行するかどうか，すなわち，原特徴空間に対する出力空間を求めるのか，ヒルベルト空間に対する出力空間を求めるのか，という点で異なる．

一般に，線形分離可能性[1]は次元数が高くなるほど高まる．図2.4は線形分離不可能な例である．これをヒルベルト空間へ射像することで，線形分離可能になると期待される．すなわち，高次元なヒルベルト空間に対するベイズエラーが最小となる出力空間の方が，原特徴空間に対するベイズ

[1]特徴空間における2つのクラスの分布が，一本の判別直線によって分割できる可能性

図 3.1: **非線形適応的次元圧縮法の基本構成**．入力パターンをはるかに高次元なヒルベルト空間へ写像し，ヒルベルト空間における像に LL 法を適用することで識別に最適な出力空間への写像を求める．

エラーが最小となる出力空間よりも，よりベイズエラーの少ない出力空間を求められると期待される．

しかし，難しい問題を線形分離可能にするためには，高次元な空間に写像しなければならないので，結果的に膨大な計算量が必要となってしまう．本論文では，それを解決する方法として Vapnik らが提案した **カーネルトリック**[5] を用いる．

3.2　カーネルトリック

原特徴ベクトル $\boldsymbol{x}$ を非線形写像 ϕ によってヒルベルト空間に写像した像を $\phi(\boldsymbol{x})$ とする．

$$\phi : \boldsymbol{x} \mapsto \phi(\boldsymbol{x}) \tag{3.1}$$

$\phi(\boldsymbol{x})$ は無限次元にもなりうるものであるが，もしヒルベルト空間におけるサンプル $\phi(\boldsymbol{x}_1)$，$\phi(\boldsymbol{x}_2)$ の内積が ϕ を陽に用いることなく $\boldsymbol{x}_1$ と $\boldsymbol{x}_2$ だけから簡単に計算できるとしたら，計算量は圧倒的に小さくなる．式で書くと

$$< \phi(\boldsymbol{x}_1), \phi(\boldsymbol{x}_2) >= K(\boldsymbol{x}_1, \boldsymbol{x}_2) \tag{3.2}$$

となる K があればよい．このような K をカーネルと呼ぶ．K は計算が易しいものが望ましいが，一般に与えられた ϕ に対して K が簡単になるとは限らない．そこで逆に適当に選んだ K が，$\phi(\boldsymbol{x})$ の内積の形で表せるか否かが判断できないだろうか．実際の ϕ は知る必要がないので，その保証だけが得られればよい．これについては，関数解析で Mercer の定理 [6], [7] が知られている．Mercer の定理を満たす K は，式 (3.2) を満たす非線形写像 ϕ が存在することが保証される．Mercer の条件を満たす K の中で

図 3.2: **KL 法の構成.** KL 法はヒルベルト空間における像の分布が Lotlikar の仮定を満たすとし，LL 法を用いて出力空間への線形写像を求めることにより，非線形な次元圧縮を実現する．本論文ではカーネルトリックにより非線形写像 ϕ を計算することなく出力空間への写像を求める．

よく用いられるものとしては，ガウスカーネル (gauss kernel)

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2) = \exp\left(\frac{-\parallel \boldsymbol{x}_1 - \boldsymbol{x}_2 \parallel^2}{2\sigma_g^2}\right) \tag{3.3}$$

多項式カーネル (polynomial kernel)

$$K(\boldsymbol{x}_1, \boldsymbol{x}_2) = (1 + \boldsymbol{x}_1^T \boldsymbol{x}_2)^p \tag{3.4}$$

等が挙げられる．本研究ではこの 2 種類のカーネルを用いる．

最適化問題においてヒルベルト空間における内積を K で置き換えることにより，ϕ を陽に計算することなく最適化を行うことができる．このように，高次元空間に写像しながらその計算を避けて K の計算だけで済ますテクニックを **カーネルトリック** と呼ぶ．

3.3　適応的カーネル次元圧縮法

ヒルベルト空間における像が Lotlikar の仮定 (a)(b) を満たすと仮定する．この条件をもとに LL 法を非線形に拡張した手法を，適応的カーネル次元圧縮法と呼ぶ (Kernel Lotlikar method: 以下 KL 法 と呼ぶ)．KL 法の流れ図を図 3.2 に示す．KL 法はヒルベルト空間に ϕ で写像した像に対し LL 法を適用することで出力空間への線形写像 W_ϕ を求める．線形写像を求める際，カーネルトリックを用いて写像 ϕ を実際に計算することなくベイズエラーが最小となる出力空間を学習する．

3.3.1　学習方法

クラス i に属するヒルベルト空間における訓練サンプルの像を行列で次のように記す.

$$\boldsymbol{X}_{\phi}^{i} \equiv [\phi(\boldsymbol{x}_{1,i}),\cdots,\phi(\boldsymbol{x}_{N_i,i})] \tag{3.5}$$

N_i はクラス i のサンプル数である. $\boldsymbol{X}_{\phi}^{i}$ を合わせたものとして,

$$\boldsymbol{X}_{\phi} \equiv [\boldsymbol{X}_{\phi}^{1},\cdots \boldsymbol{X}_{\phi}^{c}] \tag{3.6}$$

と記す. c はクラス数である.

ヒルベルト空間において, 訓練サンプル $\boldsymbol{X}_{\phi}$ が張る空間から, ヒルベルト空間の部分空間への線形射影を考える. すると, ヒルベルト空間から出力空間への線形写像 $\boldsymbol{W}_{\phi}$ は, $\boldsymbol{X}_{\phi}$ が張る空間から出力空間への線形写像として表わせる. すなわち, $\boldsymbol{X}_{\phi}$ の各訓練サンプルの線形和として表すことができる.

$$\boldsymbol{W}_{\phi} \equiv \boldsymbol{X}_{\phi}\boldsymbol{W} \tag{3.7}$$

これにより求めるパラメータが $\boldsymbol{W}_{\phi}$ から $\boldsymbol{W} \in R^{(N\times m)}$ となる. ただし N は全訓練サンプル数であり, m は出力次元数である.

ヒルベルト空間におけるサンプルの像が, Lotlikar の仮定 (a), すなわちそれぞれのクラスの平均を中心として単位行列を共分散行列に持つ正規分布にしたがっていると仮定する. このとき, クラス i の標本がクラス j に誤識別される確率は LL 法と同様に

$$\epsilon_{ij} = \frac{1}{\sqrt{2\pi\sigma}}\int_{\hat{d}_{ij}/2}^{\infty}\exp(\frac{t^2}{2\sigma^2})dt \tag{3.8}$$

と表される. $\hat{d}_{ij}$ は出力空間における平均間の距離である. ヒルベルト空間におけるクラス i の平均 $\mu_{\phi i}$ は, $\mu_{\phi i} = \frac{1}{N_i}\boldsymbol{X}_{\phi}^{i}\mathbf{1}_{N_i}$ と表わせる. 但し, $\mathbf{1}_{N_i}$ はすべての要素が 1 である N_i 次元の縦ベクトルである. 従って, $\hat{d}_{ij}$ は次のように計算できる.

$$\begin{aligned}\hat{d}_{ij}^{2} &= \left|\left|\boldsymbol{W}_{\phi}^{T}\left(\frac{1}{N_i}\boldsymbol{X}_{\phi}^{i}\mathbf{1}_{N_i} - \frac{1}{N_j}\boldsymbol{X}_{\phi}^{j}\mathbf{1}_{N_j}\right)\right|\right|^2 \\ &= \left|\left|\boldsymbol{W}^{T}\boldsymbol{V}_{ij}\right|\right|^2 \end{aligned} \tag{3.9}$$

但し,

$$\boldsymbol{V}_{ij} \equiv \frac{1}{N_i}\boldsymbol{K}_{(\cdot,i)}\mathbf{1}_{N_i} - \frac{1}{N_j}\boldsymbol{K}_{(\cdot,j)}\mathbf{1}_{N_j} \tag{3.10}$$

であり，$\boldsymbol{K}_{(\cdot,j)}$ は $\boldsymbol{K}=\boldsymbol{X}_\phi^T\boldsymbol{X}_\phi$ より次のようにおいた．

$$\begin{aligned}\boldsymbol{K} &\equiv \begin{bmatrix} \boldsymbol{K}_{(1,1)} & \cdots & \boldsymbol{K}_{(1,c)} \\ \vdots & \cdots & \vdots \\ \boldsymbol{K}_{(c,1)} & \cdots & \boldsymbol{K}_{(c,c)} \end{bmatrix} \\ &\equiv [\boldsymbol{K}_{(\cdot,1)},\cdots,\boldsymbol{K}_{(\cdot,c)}] \\ &\equiv [\boldsymbol{K}_{(1,\cdot)}^T,\cdots,\boldsymbol{K}_{(c,\cdot)}^T]^T \end{aligned} \tag{3.11}$$

$\boldsymbol{K}_{(i,j)} \in R^{(N_i\times N_j)}$ の各要素はクラス i の各訓練サンプルとクラス j の各訓練サンプルをヒルベルト空間において内積をとった値になっている．この訓練サンプルのヒルベルト空間における内積を各要素として持つ行列 $\boldsymbol{K}$ をカーネル行列と呼ぶ．$\boldsymbol{K}$ はカーネルトリックを用いて実際に ϕ を計算することなく求めることができる．これより，目的関数

$$J=\sum_{i=1}^{c}\sum_{j=i+1}^{c}\epsilon_{ij} \tag{3.12}$$

の勾配は次式のようになる．

$$\frac{\partial J}{\partial \boldsymbol{W}} = -\sum_{i=1}^{c}\sum_{j=i+1}^{c}\frac{1}{\hat{d}_{ij}}\exp\left(\frac{(\hat{d}_{ij}/2)^2}{2\sigma^2}\right)\boldsymbol{V}_{(ij)}\boldsymbol{V}_{(ij)}^T\boldsymbol{W} \tag{3.13}$$

但し，$\frac{1}{2\sqrt{2\pi\sigma^2}}$ は定数であるので省略した．$\boldsymbol{W}$ の更新式は次のようになる．

$$\boldsymbol{W}_{new}=\boldsymbol{W}_{old}-\eta\frac{\partial J}{\partial \boldsymbol{W}} \tag{3.14}$$

但し，η は学習定数である．Lotlikar の仮定を満たすために，$\boldsymbol{W}$ の更新毎に $\boldsymbol{W}_\phi^T\boldsymbol{W}_\phi=\boldsymbol{I}$ を満たすように変換する．結局，$\boldsymbol{W}$ の学習方法は以下の手続きにより行われる．

STEP1　訓練サンプルより K を計算する．

STEP2　$\boldsymbol{W}$ の初期値をランダムに定める．

STEP3　J の値が収束するまで以下を繰り返す．

substep1　$\boldsymbol{W}_\phi^T\boldsymbol{W}_\phi=\boldsymbol{I}$ を満たすため以下の変換を $\boldsymbol{W}$ に施す．

$$\boldsymbol{W} \Leftarrow \boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{K}\boldsymbol{W})^{-\frac{1}{2}} \tag{3.15}$$

substep2　式 (3.14) により $\boldsymbol{W}$ を更新する．

但し，平方根の逆行列 $(\boldsymbol{W}^T\boldsymbol{K}\boldsymbol{W})^{-\frac{1}{2}}$ は，$\boldsymbol{W}^T\boldsymbol{K}\boldsymbol{W} = \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^T$ の固有値分解を用いて $(\boldsymbol{W}^T\boldsymbol{K}\boldsymbol{W})^{-\frac{1}{2}} = \boldsymbol{F}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{F}^T$ と定義する．

未知パターン $\boldsymbol{x} \in R^d$ の出力空間への射影は $\boldsymbol{W}^T\boldsymbol{k}(\boldsymbol{x})$ である．但し，$\boldsymbol{k}(\boldsymbol{x}) = \boldsymbol{X}_\phi^T\phi(\boldsymbol{x})$ で，その第 s 要素は，$k(\boldsymbol{x})_s = \boldsymbol{K}(\boldsymbol{x}_s, \boldsymbol{x})$ $(s = 1, \cdots, N)$ として，カーネル関数を用いて計算できる．

3.3.2　評価実験

人工データと実データを使用して，KL 法によって求まった出力空間の評価を行う．比較手法として LL 法により求まった出力空間の評価もあわせて行う．

人工データを用いた評価実験

人工データによる次元圧縮実験を行い，非線形に拡張することでより識別に有利な出力空間が求まることを確認する．実験条件は表 3.1 の通りである．入力空間における使用サンプルの分布を図 3.3 に示す．これは図 2.4 のような分布を想定したものである．

表 3.1: 実験条件

使用サンプル	人工データ
クラス数	2
サンプル数	100 個/$class$
入力空間の次元数	$n = 2$
出力空間の次元数	$m = 1$
比較手法	**KL 法**，LL 法

KL 法，LL 法を用いて求めた出力空間におけるサンプルの分布を図 3.5 に示す．図 3.5 は横軸がサンプル番号，縦軸がサンプルの出力値となっている．結果は LL 法では出力空間において線形分離が不可能であるが，KL 法では線形分離可能な出力空間を求めることに成功した．これにより図 2.4 のように分布の識別境界が複雑に入り組んでいるときでも，ベイズエラーのより少ない出力空間を求められることが示された．

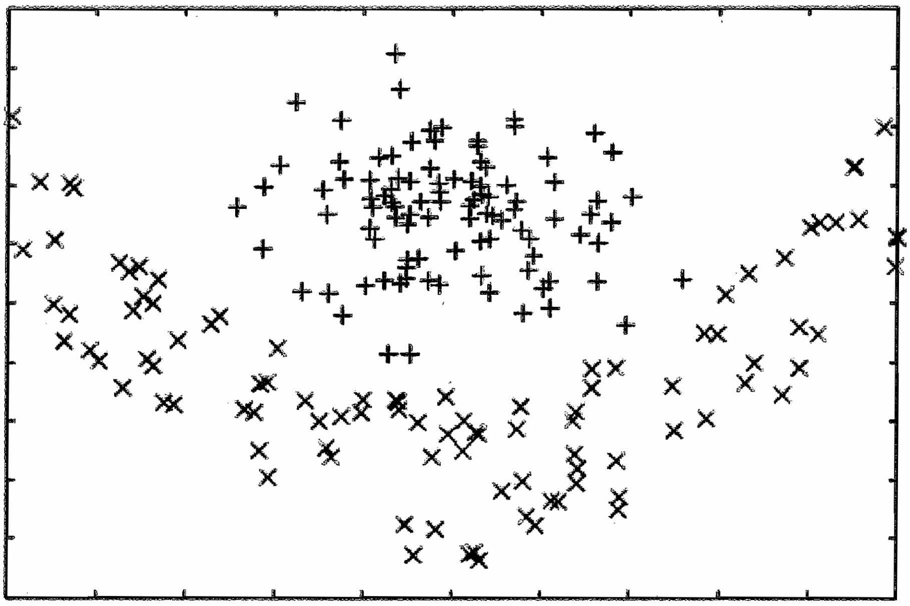


図 3.3: **入力空間における使用サンプルの分布．** 2 次元，2 クラスの人工データで，色が所属するクラスを表す．これは図 2.4 のような，非線形な分離境界面を持つ分布の例である．線形写像ではどのような写像を行っても出力空間において必ず分布の重なり，つまりベイズエラーが生じてしまう．

実データを用いた評価実験

実データを用いて次元圧縮実験を行い，KL 法，LL 法の比較を行う．実データとして前節で用いた Image Segmentation Database を用いた．実験条件は表 3.2 の通りである．

ハイパーパラメータである k-NN の k，Lotlikar の仮定における分散 σ，ガウスカーネルの σ_g，多項式カーネルの p は，5-fold 交差確認法により訓練サンプルより自動的に求めた．求まった出力空間の評価は，出力空間におけるサンプルの分布の重なり具合と，k-NN による認識率を用いて行う．第 2 章と同様に，前処理として入力空間におけるクラス内変動行列 $\boldsymbol{S}_w$ が単位行列となるようにサンプルを白色化した．但し，この白色化は入力空間においてクラス内変動行列 $\boldsymbol{S}_w$ を単位行列とする白色化であり，KL 法を用いた際，ヒルベルト空間における入力パターンの像のクラス内変動行列は単位行列とならない．つまり，Lotlikar の仮定を満たすための白色化とは意味が異なる．

表 3.2: 実験条件

使用サンプル	Image Segmentation Database
クラス数	7
訓練サンプル数	30 個/$class$
テストサンプル数	300 個/$class$
入力空間の次元数	$n = 18$
出力空間の次元数	$m = 2$
識別器	k-NN 法
比較手法	**KL 法**，LL 法

表 3.3 に k-NN 法による認識率の一覧を示す．図 3.6 に出力空間におけるテストサンプルの分布を示す．分布中の実線は出力空間における各クラスの標準偏差の 2 倍の軸をもつ楕円である．図 3.6 を見た限りでは KL 法，LL 法の差は見られなかった．これは，ヒルベルト空間における入力パターンの像の分布が Lotlikar の仮定を満たしていないため，最適な線形写像を求めることができていないことが原因と考えられる．k-NN 法による認識では，テストサンプルに対する認識率に 3.2%の差があった．

表 3.3: 実験結果 k-NN 法による認識率

	訓練サンプル	テストサンプル
KL 法	87.1%	86.0%
LL 法	82.6%	82.8%

3.4　適応的白色カーネル次元圧縮法

前節の方法はヒルベルト空間における像が Lotlikar の仮定 (a)(b) を満たしているという仮定があった．Lotlikar 仮定 (a) を満たさない場合，これをおおよそ満たすように近付けるにはまず訓練サンプルのクラス内分散をヒルベルト空間において正規化，つまり白色化すれば良い [8]，[9]．このように，訓練サンプルをヒルベルト空間において白色化し，LL 法の入力とする方法を適応的白色カーネル次元圧縮法と呼ぶ (White Kernel Lotlikar Method: 以下 WKL 法と呼ぶ)．WKL 法の流れ図を図 3.4 に示す．但し，

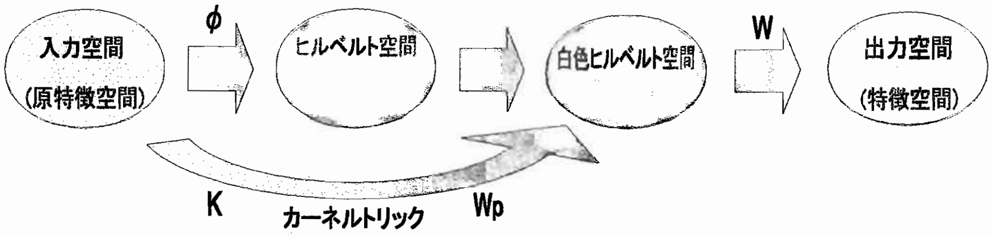


図 3.4: WKL **法の構成**．はじめにヒルベルト空間における像を白色化をする線形写像を求め，白色ヒルベルト空間に写像する．続いて，白色ヒルベルト空間における像に LL 法を適用することで，出力空間への線形写像 $\boldsymbol{W}$ を求める．但し，カーネルトリックを用いて写像 ϕ を実際に計算することなく白色ヒルベルト空間への写像が求まる．

白色ヒルベルト空間とは，ヒルベルト空間において白色化された訓練サンプルの張る空間である．WKL 法は以下の手順で行われる．

STEP1 非線形変換によって原特徴ベクトルを原特徴空間よりもはるかに高次元なヒルベルト空間へ非線形写像する．

STEP2 ヒルベルト空間に ϕ で写像した像を白色化をするような線形写像を求め，白色ヒルベルト空間に写像する．

STEP3 白色ヒルベルト空間における像に LL 法を適用することで，出力空間への線形写像 $\boldsymbol{W}$ を求める．

但し，カーネルトリックを用いて写像 ϕ を実際に計算することなく白色ヒルベルト空間への写像が求まる．

3.4.1 ヒルベルト空間における白色化

ヒルベルト空間における各クラスの分布をその重心が平均ゼロで一致するように重ねるため，クラス i における訓練サンプルの平均をゼロとするような次の行列を定義する．

$$\tilde{\boldsymbol{X}}_\phi^i \equiv [\phi(\boldsymbol{x}_{1,i}) - \boldsymbol{\mu}_{\phi i}, \cdots, \phi(\boldsymbol{x}_{N_i,i}) - \boldsymbol{\mu}_{\phi i}] \quad (3.16)$$

N_i はクラス i の標本数，$\boldsymbol{\mu}_{\phi i}$ はヒルベルト空間におけるクラス i の平均ベクトルである．$\tilde{\boldsymbol{X}}_\phi^i$ を合わせたものとして，

$$\tilde{\boldsymbol{X}}_\phi \equiv [\tilde{\boldsymbol{X}}_\phi^1, \cdots, \tilde{\boldsymbol{X}}_\phi^c] \quad (3.17)$$

と記す．c はクラス数である．このときヒルベルト空間におけるクラス内変動行列 $\boldsymbol{S}_{\phi w}$ は，

$$\boldsymbol{S}_{\phi w} = \tilde{\boldsymbol{X}}_\phi \tilde{\boldsymbol{X}}_\phi^T \tag{3.18}$$

となる．しかし，このクラス内変動行列 $\boldsymbol{S}_{\phi w}$ は，特徴空間における遥かに高次元の値を持ち直接特異値分解することはできない．そこで以下では，特異値分解の特性から，クラス内分散行列 $\boldsymbol{S}_{\phi w}$ を直接特異値分解することなく正規化，つまり白色化を行う方法を示す．

$\tilde{\boldsymbol{X}}_\phi^T \tilde{\boldsymbol{X}}_\phi$ の固有値を λ_i $(\lambda_1 \geq \cdots \geq \lambda_{d_\phi})$ とおくと，その固有値は r $(= rank(\tilde{X}_\phi))$ 個存在し，$\tilde{\boldsymbol{X}}_\phi \tilde{\boldsymbol{X}}_\phi^T$ の固有値と一致する．固有値 λ_i に対応する $\tilde{\boldsymbol{X}}_\phi^T \tilde{\boldsymbol{X}}_\phi$, $\tilde{\boldsymbol{X}}_\phi \tilde{\boldsymbol{X}}_\phi^T$ の正規直交固有ベクトルをそれぞれ $\boldsymbol{u}_i, \boldsymbol{v}_i$，特徴空間の次元数を d_ϕ, 全訓練標本の個数を N としたとき，行列 $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\Lambda}$ を，

$$\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r] \qquad (\in R^{(N \times r)}) \tag{3.19}$$

$$\boldsymbol{V} = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r] \qquad (\in R^{(d_\phi \times r)}) \tag{3.20}$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_r \end{bmatrix} \tag{3.21}$$

とおくと，$\tilde{\boldsymbol{X}}_\phi^T \tilde{\boldsymbol{X}}_\phi$, $\tilde{\boldsymbol{X}}_\phi \tilde{\boldsymbol{X}}_\phi^T$ はそれぞれ

$$\tilde{\boldsymbol{X}}_\phi \tilde{\boldsymbol{X}}_\phi^T = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T \qquad (\in R^{(d_\phi \times d_\phi)}) \tag{3.22}$$

$$\tilde{\boldsymbol{X}}_\phi^T \tilde{\boldsymbol{X}}_\phi = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T \qquad (\in R^{(N \times N)}) \tag{3.23}$$

と表され，このとき，特異値分解の特性から次の式が成り立つ．[8]

$$\boldsymbol{V}^T = \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{U}^T \tilde{\boldsymbol{X}}_\phi^T \tag{3.24}$$

ヒルベルト空間において，白色化された訓練サンプルを $\boldsymbol{X}_{\phi W}$ とすると，クラス内変動行列を正規化する行列演算は，次の式となる．

$$\boldsymbol{X}_{\phi W} = \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{V}^T \boldsymbol{X}_\phi \tag{3.25}$$

ここで，c 個のブロックを持つ対角行列であり，第 i ブロックが $N_i \times N_i$ 行列で全要素が $1/N_i$ であるような行列 $\boldsymbol{M}$ $(\in R^{(N \times N)})$ を考える．

$$\tilde{\boldsymbol{X}}_\phi = (\boldsymbol{X}_\phi - \boldsymbol{X}_\phi \boldsymbol{M}) = \boldsymbol{X}_\phi(\boldsymbol{I} - \boldsymbol{M}) \tag{3.26}$$

$$\boldsymbol{M} = \left[\begin{array}{ccc|c|ccc} \frac{1}{N_1} & \cdots & \frac{1}{N_1} & & & & \\ \vdots & \ddots & \vdots & \mathbf{0} & & \mathbf{0} & \\ \frac{1}{N_1} & \cdots & \frac{1}{N_1} & & & & \\ \hline & & & \ddots & & & \\ & \mathbf{0} & & \ddots & & \mathbf{0} & \\ & & & \ddots & & & \\ \hline & & & & \frac{1}{N_C} & \cdots & \frac{1}{N_C} \\ & \mathbf{0} & & \mathbf{0} & \vdots & \ddots & \vdots \\ & & & & \frac{1}{N_C} & \cdots & \frac{1}{N_C} \end{array}\right] \tag{3.27}$$

(3.25) に (3.24), (3.26) を代入すると，

$$\begin{aligned} \boldsymbol{X}_{\phi W} &= \boldsymbol{\Lambda}^{-1}\boldsymbol{U}^T(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{K} \\ &= \boldsymbol{W}_p\boldsymbol{K} \end{aligned} \tag{3.28}$$

と表すことができる．但し，$\boldsymbol{W}_p = \boldsymbol{\Lambda}^{-1}\boldsymbol{U}^T(\boldsymbol{I}-\boldsymbol{M})$ とした．一方，(3.23), (3.26) から

$$\begin{aligned} \boldsymbol{U\Lambda U}^T &= (\boldsymbol{I}-\boldsymbol{M})\boldsymbol{X}_\phi{}^T\boldsymbol{X}_\phi(\boldsymbol{I}-\boldsymbol{M}) \\ &= (\boldsymbol{I}-\boldsymbol{M})\boldsymbol{K}(\boldsymbol{I}-\boldsymbol{M}) \end{aligned} \tag{3.29}$$

となる． これは，$\boldsymbol{U}, \boldsymbol{\Lambda}$ はそれぞれ，$\boldsymbol{K}$ が与えられれば (3.29) の固有ベクトル，固有値として求まることを意味する．

よって $\boldsymbol{K}$ を定めるカーネル K と訓練サンプルが与えられれば，ϕ の具体的な形を知らずに $\boldsymbol{X}_{\phi W}$ が求まり，クラス内分散を正規化した白色ヒルベルト空間への非線形写像を求めることができることが示せた．

3.4.2　学習方法

白色ヒルベルト空間に対する最適な部分空間への写像の学習は，白色化された訓練サンプルを LL 法に適用し線形写像 $\boldsymbol{W} \in R^{(r \times m)}$ を求めることで行われる．但し，r は白色ヒルベルト空間の次元数，m は出力空間の次元数である．

結局，WKL 法は以下の手順で出力空間への写像を学習する．

STEP1　訓練サンプルより $\boldsymbol{K}$ を計算する．

STEP2　白色ヒルベルト空間への写像 $\boldsymbol{W}_p$ を求め $\boldsymbol{X}_{\phi W}$ を計算する.

STEP3　$\boldsymbol{W}$ の初期値をランダムに定める.

STEP4　J の値が収束するまで以下を繰り返す.

substep1　$\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{I}$ を満たすため以下の変換を $\boldsymbol{W}$ に施す.

$$\boldsymbol{W} \Leftarrow \boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{W})^{-\frac{1}{2}} \tag{3.30}$$

substep2　$\boldsymbol{X}_{\phi W}$ を用いて式 (2.16) により $\boldsymbol{W}$ を更新する.

但し，平方根の逆行列 $(\boldsymbol{W}^T\boldsymbol{W})^{-\frac{1}{2}}$ は，$\boldsymbol{W}^T\boldsymbol{W} = \boldsymbol{F}\boldsymbol{A}\boldsymbol{F}^T$ の固有値分解を用いて，$(\boldsymbol{W}^T\boldsymbol{W})^{-\frac{1}{2}} = \boldsymbol{F}\boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{F}^T$ と定義する.

未知パターン $\boldsymbol{x} \in R^d$ の出力空間への射影は $\boldsymbol{W}^T\boldsymbol{k}(\boldsymbol{x})$ である．但し，$\boldsymbol{k}(\boldsymbol{x}) = \boldsymbol{X}_\phi^T\phi(\boldsymbol{x})$ で，その第 s 要素は，$k(\boldsymbol{x})_s = \boldsymbol{K}(\boldsymbol{x}_s, \boldsymbol{x})$ $(s = 1, \cdots, N)$ として，カーネル関数を用いて計算できる.

3.4.3　評価実験

実データを使用して，WKL 法によって求まった出力空間の評価を行う．比較手法として KL 法により求まった出力空間の評価もあわせて行う．実データとして前節で用いた Image Segmentation Database を再び用いた．実験条件は表 3.4 の通りである.

表 3.4: 実験条件

使用サンプル	Image Segmentation Database
クラス数	7
訓練サンプル数	30 個/$class$
テストサンプル数	300 個/$class$
入力空間の次元数	$n = 18$
出力空間の次元数	$m = 2$
識別器	k-NN 法
比較手法	**WKL 法**，KL 法

ハイパーパラメータである k-NN の k，Lotlikar の仮定における分散 σ，ガウスカーネルの σ_g，多項式カーネルの p は，5-fold 交差確認法により訓練サンプルより自動的に求めた．求まった出力空間の評価は，出力空間

におけるサンプルの分布の重なり具合と，k-NN による認識率を用いて行う．

前処理として入力空間におけるクラス内変動行列 $\boldsymbol{S}_w$ が単位行列となるようにサンプルを白色化した．但し，この白色化も入力空間においてクラス内変動行列 $\boldsymbol{S}_w$ を単位行列とする白色化であり，WKL 法を用いた際の，ヒルベルト空間における入力パターンの像のクラス内変動行列を単位行列とするためではない．つまり，Lotlikar の仮定を満たすための白色化とは意味が異なる．

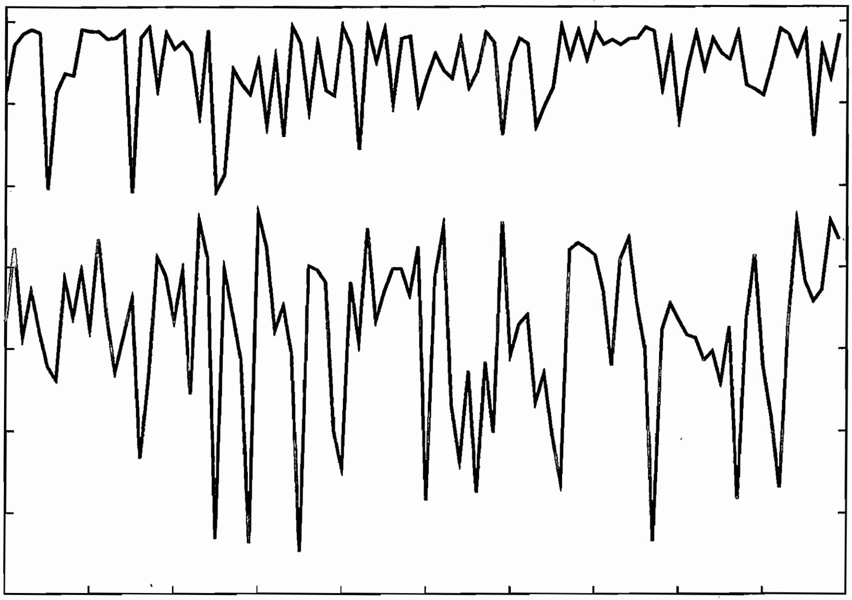
表 3.5 に k-NN 法による認識率の一覧を示す．図 3.7 に出力空間におけるテストサンプルの分布を示す．分布中の実線は出力空間における各クラスの標準偏差の 2 倍の軸をもつ楕円である．図 3.7 を見ると，WKL 法では黄色のクラスとピンク色のクラス，赤色のクラスとピンク色のクラスがそれぞれ分離され，ピンク色のクラスと水色のクラスと青色のクラスもクラス間の距離が広いような出力空間が得られたことが確認できる．k-NN 法による認識でも，テストサンプルに対する認識率が 6.9%向上した．

表 3.5: 実験結果 k-NN 法による認識率

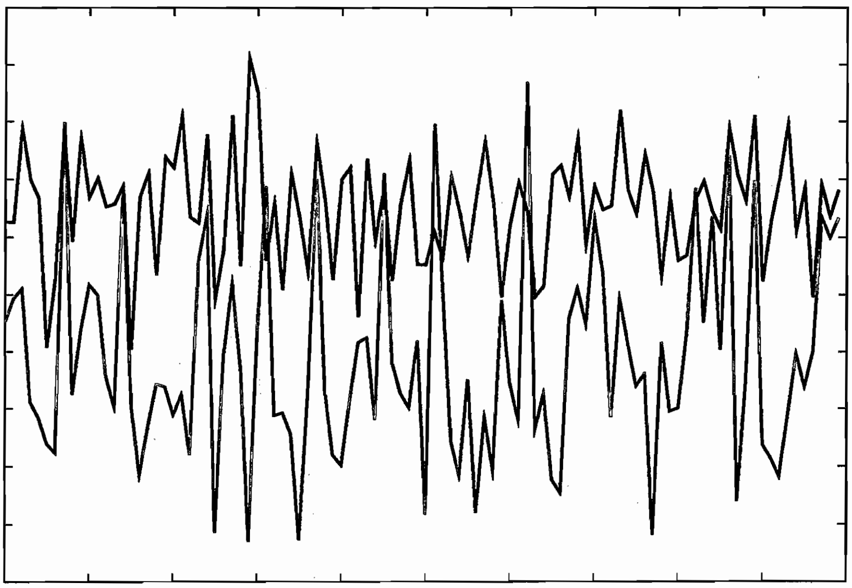
	訓練サンプル	テストサンプル
WKL 法	95.7%	92.9%
KL 法	87.1%	86.0%

3.5　まとめ

ヒルベルト空間を介することで非線形な教師つき次元圧縮を行う KL 法，WKL 法を提案した．KL 法では人工データを用いて第 2 章で指摘した，非線形な境界面をもつサンプルに対しても線形分離可能な出力空間を得られた．これにより，LL 法を非線形に拡張することにより複雑な分離境界面をもつ問題にも対応可能になったことが確認された．一方，WKL 法ではヒルベルト空間において白色化を行うことにより，より識別に有利な出力空間を得ることに成功した．公開データベースである Image Segmentation Database を用いた評価実験を行い，LL 法よりも約 10%の認識率の高い出力空間を得ることに成功した．

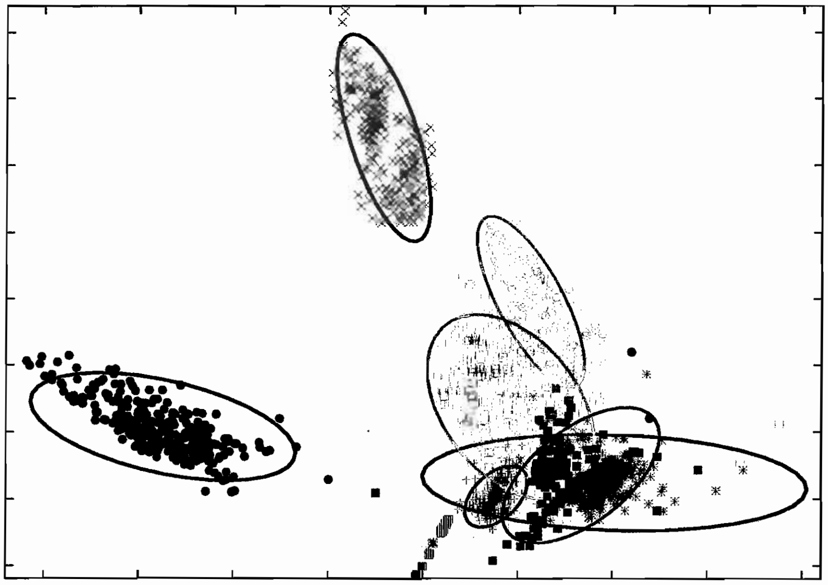


(a) KL法

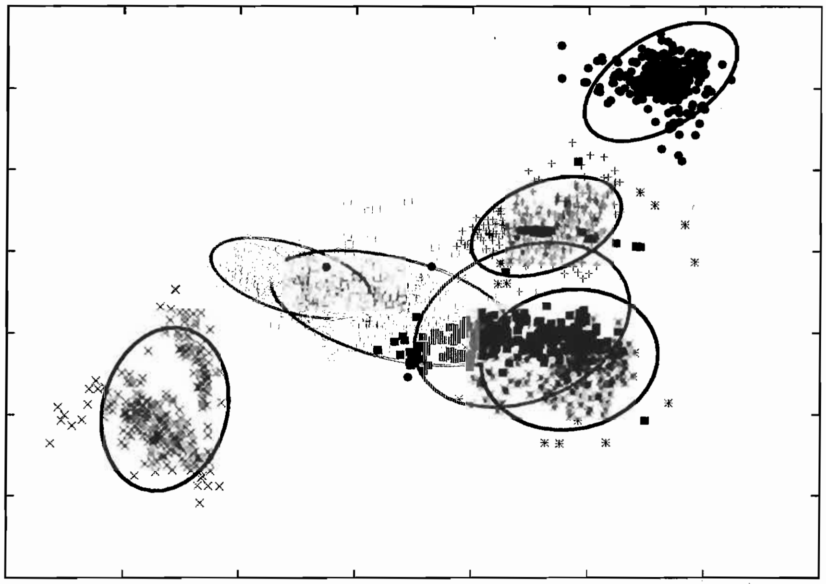


(b) LL法

図 3.5: **人工サンプル の出力空間における分布**．横軸がサンプル番号，縦軸がサンプルの出力値となっている．(b) の LL 法では出力空間において分布が重なっており線形分離が不可能であるのに対し，(a) の KL 法では分布が完全に分離されており線形分離可能な出力空間を求めることに成功したことがわかる．

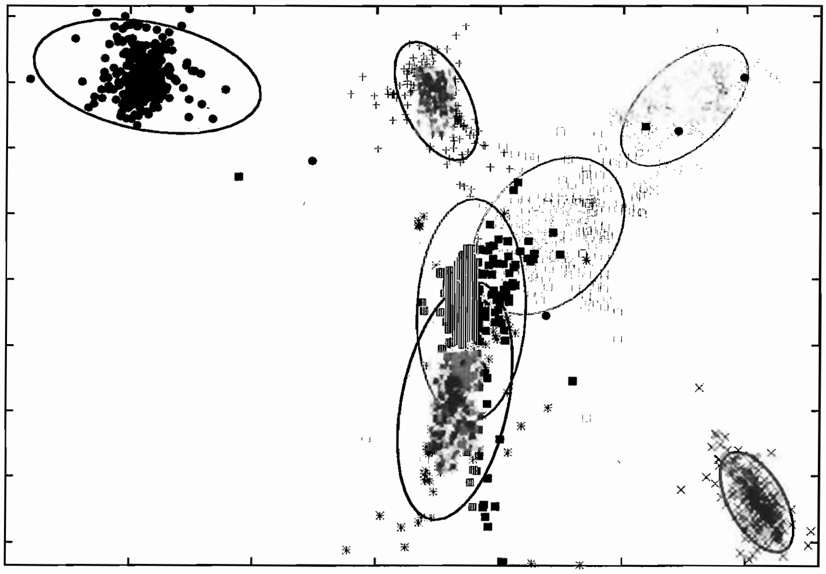


(a) KL法

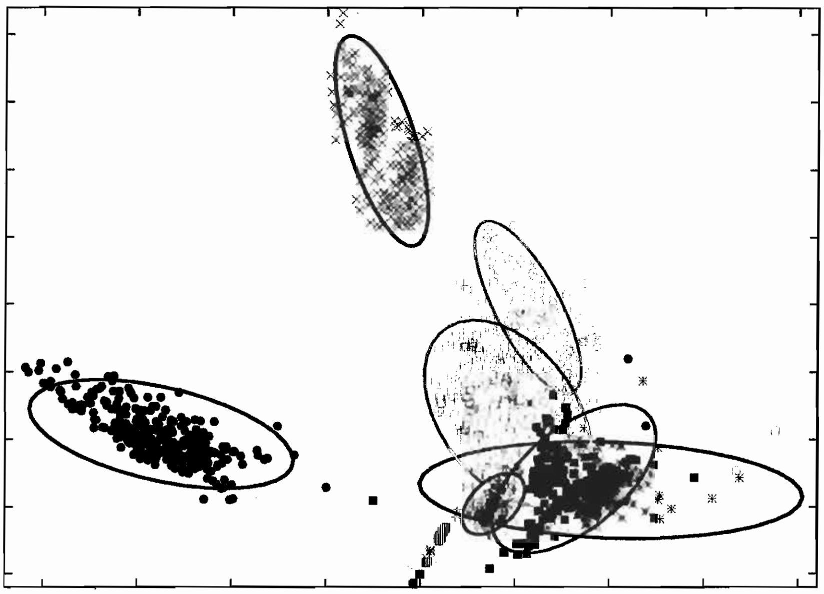


(b) LL法

図 3.6: Image Segmentation の出力空間におけるテストサンプルの分布．分布中の実線は出力空間における各クラスの標準偏差の 2 倍の軸をもつ楕円である．図を見た限りでは KL 法，LL 法の差はあまり見られなかった．



(a) WKL法



(b) KL法

図 3.7: Image Segmentation **の出力空間におけるテストサンプルの分布.** 分布中の実線は出力空間における各クラスの標準偏差の 2 倍の軸をもつ楕円である．(a) の WKL 法では WKL 法では黄色のクラスとピンク色のクラス，赤色のクラスとピンク色のクラスがそれぞれ分離され，ピンク色のクラスと水色のクラスと青色のクラスもクラス間の距離が広いような出力空間が得られたことが確認できる．

第4章

評価実験

4.1 はじめに

本章では提案手法であるKL法，WKL法に対して公開データベースを用いた実験を行う．過去に提案された教師つき次元圧縮法との比較を行うことで提案手法の有効性を確認する．また，実験結果より提案手法のもつ性質についても議論を行う．

4.2 実験条件

公開データベースとして，カリフォルニア州立大学アービン校の6つの公開データベース を使用する．詳細を表4.1に示す．但し，Multiple Featureは特徴量のうちZernikeを使用し，LED25，LED7の雑音はそれぞれ5%とした．

出力次元はBananaを除くデータベースで$m=2$とした．Bananaは非線形な分離境界面をもつような2次元人工データであり，入力空間が$n=2$であるので$m=1$とした．比較手法として，LL法，フィッシャーの線形判別分析[1](Linear Fisher method: 以下LF法と呼ぶ)，カーネル判別分析[10][11](Kernel Fisher merhod: 以下KF法と呼ぶ)を用いた．求まった出力空間の評価はk-NNによる認識率により行う．

データベースの各要素が正規化された値をもたない時，入力空間にお

表 4.1: 使用データベースの概要

データベース名	次元数	クラス数	訓練サンプル数/class	テストサンプル数/class
Image Segmentation	18	7	30	300
LED25	25	10	30	300
LED7	7	10	30	300
Multiple Feature	47	16	30	300
Pen Recognition	16	10	30	300
Banana	2	2	100	1000

けるクラス内変動行列 $\boldsymbol{S}_w$ が単位行列となるようにサンプルを白色化した．白色化の処理を行ったのは，Image Segmentation Database, Multiple Feature, Pen Recognition である．他のデータベースは前処理を行わずにそのまま入力とした．但し，この白色化は入力空間においてクラス内変動行列 $\boldsymbol{S}_w$ を単位行列とする白色化であり，KL 法，WKL 法を用いた際の，ヒルベルト空間における入力パターンの像のクラス内変動行列を単位行列とするためではない．つまり，Lotlikar の仮定を満たすための白色化とは意味が異なる．

ハイパーパラメータである k-NN の k，Lotlikar の仮定における分散 σ，ガウスカーネルの σ_g，多項式カーネルの p は，5-fold 交差確認法により訓練サンプルより自動的に求めた．カーネルはハイパーパラメータと同様に，5-fold 交差確認法により訓練サンプルよりガウスカーネルと多項式カーネルの訓練サンプルに対する認識率の良い方を用いた．

4.3 実験結果

各データベースに対する実験結果を図 4.1，図 4.2，図 4.3，図 4.4，図 4.5，図 4.6 に示す．実験結果を見ても分かる通り，提案手法である WKL 法，KL 法が Image Segmentation，LED7，Pen Recognition，Banana に対する認識率が最もよく，従来手法に対して性能が向上しているといえる．特に WKL 法では LL 法に比べ平均して 10%以上も良い認識率が得られた．これは，ヒルベルト空間への射影することで各クラスの分離度が高まり，より識別に適した次元圧縮を行うことができた結果であると考えら

れる．

一方，LED25 では WKL 法，KL 法での認識率の向上が見られなかった．Multiple Feature では KL 法での認識率の向上が見られなかった．この2つのデータベースの共通する特徴として，LED25 では 25 次元，Multiple Feature では 47 次元の入力空間を持ち，認識率が向上したデータベースと比べ高次元な入力空間を持つ．よって，原特徴ベクトルをヒルベルト空間へ射影したときに，他のデータベースに比べて各クラスの分離度が原特徴空間よりも向上しなかったことが原因と考えられる．

以上より，ヒルベルト空間を介して出力空間を求める WKL 法，KL 法は，高次元なヒルベルト空間へ射影し，各クラスの分離度が高まったとき，より識別に適した出力空間を求めることができると考えられる．

4.4 まとめ

提案手法である KL 法，WKL 法に対して公開データベースを用いた実験を行い，過去に提案された教師つき次元圧縮法との比較を行うことで提案手法の有効性を確認した．実験より，クラス数に対して次元数が少ない，つまり高次元なヒルベルト空間に射影したときに各クラスの分離度が高まるとき，提案手法が従来手法よりもより識別に適した出力空間への写像を求めることができることがわかった．

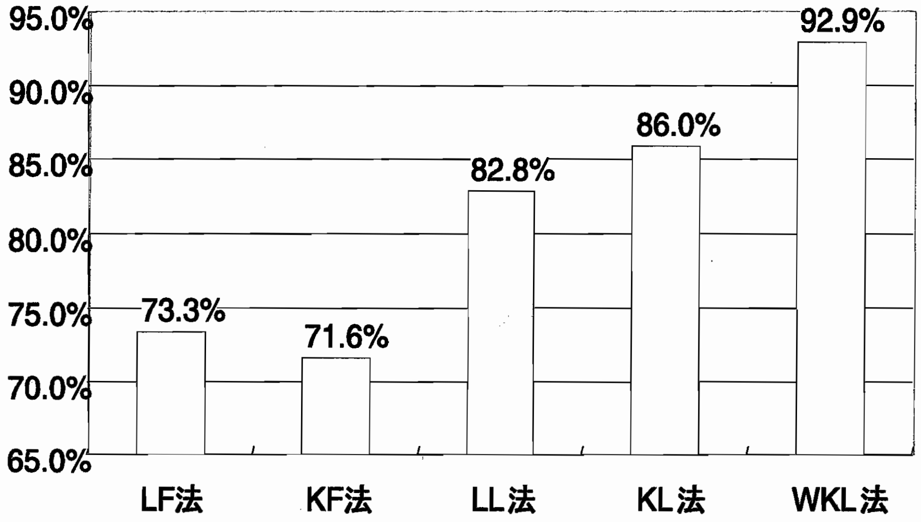


図 4.1: Image Segmentation のテストサンプルに対する各手法の認識率

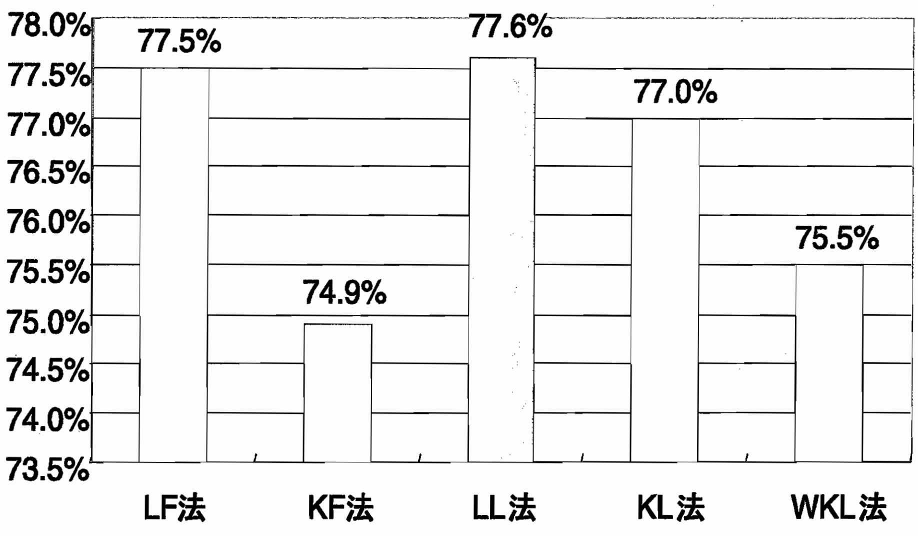


図 4.2: LED25 のテストサンプルに対する各手法の認識率

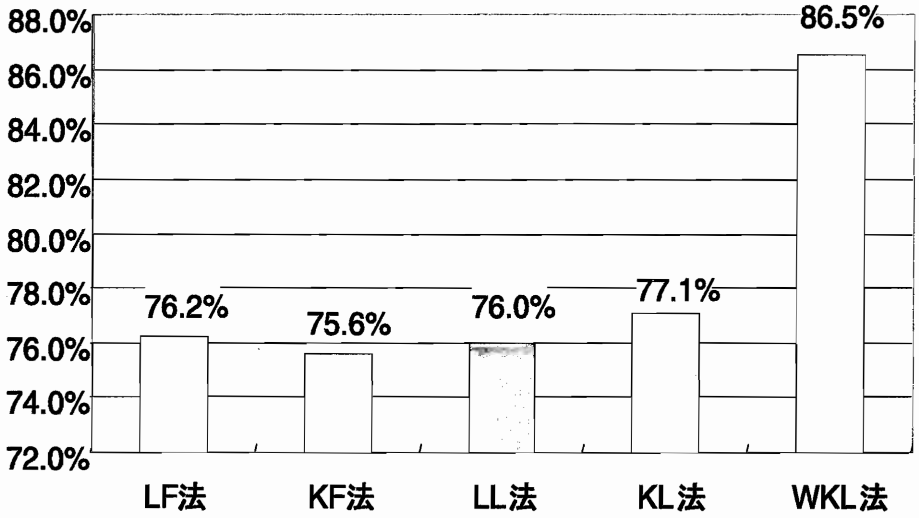


図 4.3: LED7 のテストサンプルに対する各手法の認識率

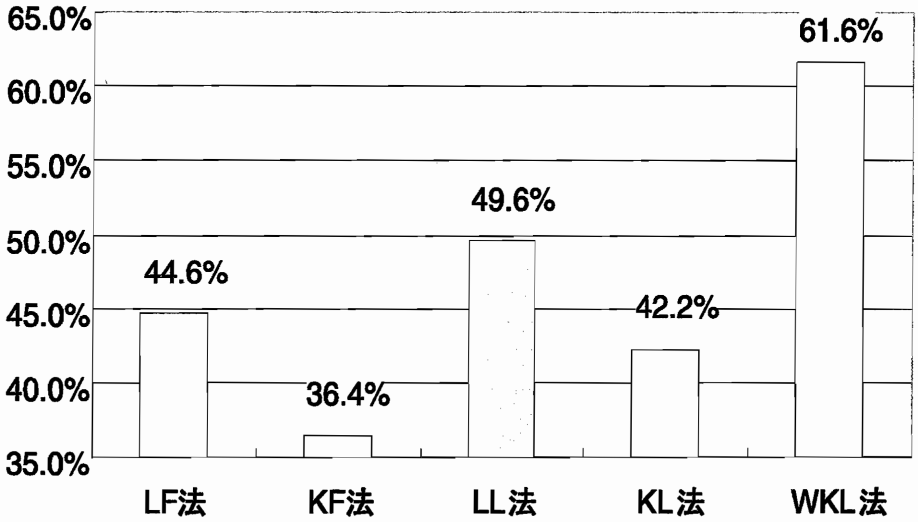


図 4.4: Multiple Feature のテストサンプルに対する各手法の認識率

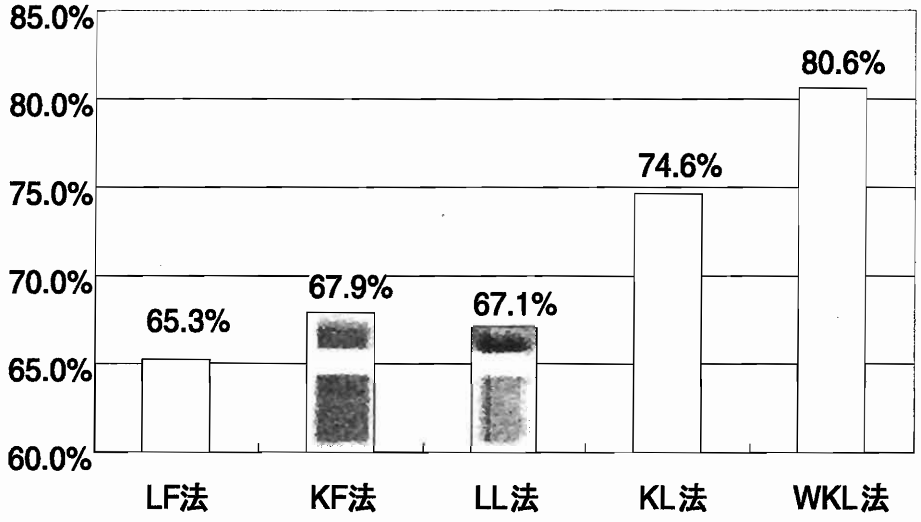


図 4.5: Pen recognition のテストサンプルに対する各手法の認識率

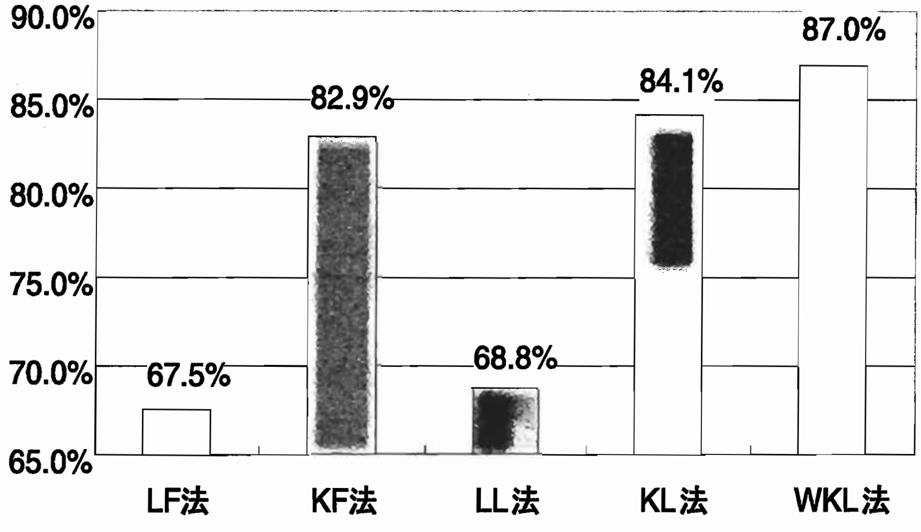


図 4.6: Banana のテストサンプルに対する各手法の認識率

第5章

結論

5.1 本論文の成果

本論文では，識別に適した次元圧縮を行うことを目指し，過去に提案された有効な次元圧縮法であるLL法をヒルベルト空間を介して次元圧縮を行うことにより非線形に拡張した．

第2章では，LL法について述べ，非線形な分離境界面を持つ問題に対しては識別に適した部分空間を求める事ができないという問題点を指摘した．

第3章では，LL法による次元圧縮をヒルベルト空間を介して行うことにより，非線形な教師つき次元圧縮を実現するKL法，WKL法を新たに提案した．次元圧縮を行う際，ヒルベルト空間への射影を実際にすることなく出力空間への写像を求めることにより，計算量を激減させるテクニックであるカーネルトリックについて述べた．また，人工データ，公開データベースを用いて提案手法の有効性を確認した．

第4章では，さまざまな公開データベースを用いて提案手法の有効性を確認した．実験により，過去に提案された次元圧縮法よりも，より識別に適した次元圧縮を行うことを確認した．また，ヒルベルト空間への写像によりクラスの分離度が高まるような問題や，非線形な分離境界面を持つ問題に対して最も性能が向上することを確認した．

5.2　今後の課題

本論文で提案した手法である KL 法，WKL 法は，LL 法を非線形に拡張したものである．LL 法には入力空間における各クラスの分布が，以下の Lotlikar の仮定 に従う必要があった．

(a) どのクラスも等しい共分散行列をもつ正規分布である．

(b) クラスの事前確率が等しい

提案手法を Lotlikar の仮定を満たさない問題に対して適用した場合，識別に最適な次元圧縮を行うことができないことがある．よって今後の方針としては，Lotlikar の仮定 (a)(b) を満たさない場合の出力空間におけるベイズエラーを定式化しなおし，新たな学習則を導出することにより，さまざまな条件に対して適用可能とすることが挙げられる．

また，より多くのデータベースに提案手法を適用し，有効性を確認する作業も行ってゆきたい．

謝辞

本研究を進めるにあたり，全般的な御指導とともにこの研究の機会を与えて下さった東北大学大学院工学研究科 阿曽弘具教授に心より感謝致します．

本論文をまとめるにあたり貴重な御意見を頂いた東北大学大学院工学研究科 川又政征教授，東北大学大学院情報科学研究科 静谷啓樹教授に深く感謝致します．

研究方針について親身に御指導頂き，ゼミを通して貴重な御意見を頂いた東北大学大学院工学研究科 大町真一郎助教授に心より感謝致します．

ゼミを通して貴重な御意見，御指導を頂いた東北大学大学院工学研究科 三宅章吾助教授に深く感謝致します．

常日頃から親身になって御意見，御助言を頂き，またこの研究分野に携わるきっかけを与えて下さった東北大学大学院工学研究科博士過程 加藤毅氏に心より感謝致します．

ゼミを通して鋭い御指摘をいただき，また発表における重要なポイントを的確に御指導賜わりました東北大学大学院工学研究科 佐藤俊治助手に深く感謝致します．

研究に必要な計算機環境を整えて頂いた東北大学大学院工学研究科 菅谷至寛助手に心より感謝致します．

また，東北大学大学院博士後期課程 下村正夫氏をはじめとする阿曽研究室の皆様には日頃から研究に限らず多岐に渡ってお世話になりました．深く感謝致します．

最後に，辛いときに支えとなってくれた友人，長い間温かく見守ってくれた両親，家族に心より感謝致します．

参考文献

[1] 石井健一郎, 上田修功, 前田英作, 村瀬洋："わかりやすいパターン認識," オーム社出版局 (1998).

[2] J. Bell and J. Sejnowski: "The independent components of natural scenes are edge filters," Vison, **37**, pp. 3327–3338 (1997).

[3] R.Lotlikar and R.Kothari: "Adaptive linear dimensionality reduction for classfication," Pattern Recognition, **33**, pp. 185–194 (2000).

[4] A. Ruiz and P. teruel: "Nonlinear kernel-based statistical pattern analysis," IEEE Trans. Neural Networks, **12**, pp. 16–32 (2001).

[5] V. Vapnik: "Statistical learning theory," John Wiley and Sons, Inc (1998).

[6] R. Coourant, D. Hilbert, 斎藤利弥, 丸山滋弥："Methods of Mathematical Physics," 東京出版 (1959).

[7] 斎藤三郎："再生核の理論入門," 牧野書店 (2002).

[8] 津田宏治："ヒルベルト空間における部分空間法," 電子情報通信学会論文誌, **J82-D-2**, pp. 592–599 (1999).

[9] 前田英作, 村瀬洋："カーネル非線形部分空間法によるパターン認識," 電子情報通信学会論文誌, **J82-D-2**, pp. 600–612 (1999).

[10] G. Baudat and F. Anmouar: "Generalized discriminant analysis using a kernel approach," Neural Computation, **13**, pp. 2385–2404 (2000).

[11] S. Mika, B. Ratsch, J. Weston, B. Scholkopf and K. Muller: "Fisher discriminant analysis with kernels," Neural Networks for Signal Processing, **6**, pp. 41–48 (1999).

研究業績

学会発表予定

佐々木裕児, 加藤毅, 佐藤俊治, 三宅章吾, 阿曽弘具,
"パターン認識のための適応的非線形次元圧縮法,"
電子情報通信学会総合大会, 2003 年 3 月. 於東北大学

パターン認識のための特徴抽出法に関する研究

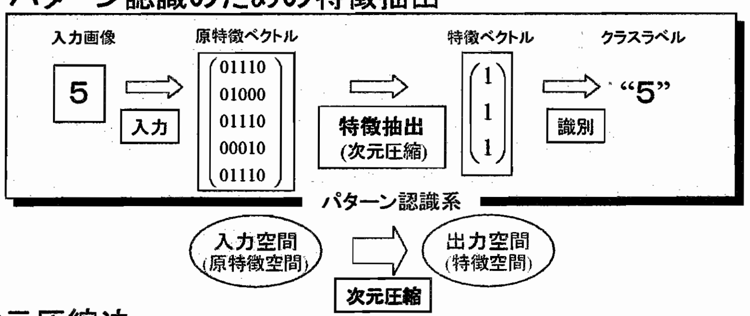
修士論文予備審査
2002年1月21日
阿曽研究室　佐々木　裕児

2003/1/21　　1

1.背景

2

- パターン認識のための特徴抽出



次元圧縮法

- 教師なし : PCA、ICA、SOM　→ 表現 (representation)
- 教師あり : Fisherの線形判別分析 → 判別 (discrimination)

$\boxed{\text{教師あり次元圧縮法により識別に有利な次元圧縮}}$

2003/1/21

1.背景

3

◆ 過去に提案された識別に有効な次元圧縮法
- 適応的線形次元圧縮法 (Lotlikar et al, 2000)
  - 識別に適した線形写像を求める 教師あり次元圧縮法
  - 実験により Fisherの線形判別分析よりも有効
- 問題点
  - 実問題で線形分離可能な場合は稀
  - 線形な次元圧縮法では対応できない問題

⇩

目的: 適応的線形次元圧縮法を非線形に拡張

2003/1/21

1.背景

4

◆ 線形次元圧縮法の非線形への拡張の例

線形次元圧縮法		非線形次元圧縮法
主成分分析		カーネル主成分分析
Fisherの線形判別分析		カーネル判別分析
適応的線形次元圧縮法	⇨	提案手法

2003/1/21

本発表の流れ

5

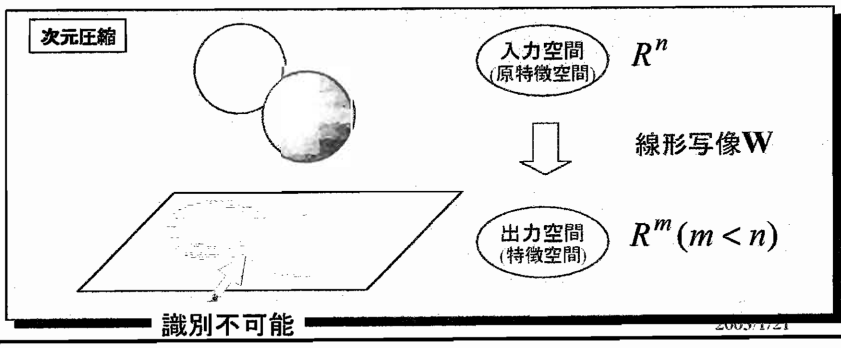
1. 背景
2. 適応的線形次元圧縮法(従来法)
3. 適応的非線形次元圧縮法(提案手法)
4. 評価実験
5. まとめと今後の課題

2003/1/21

2.適応的線形次元圧縮法(LL法)

6

- 適応的線形次元圧縮法 (Linear Lotlikar: LL法)
  - ベイズエラーを最小にする線形写像Wを求める
  - ベイズエラー
    - 特徴のもつ必然的な誤り
    - 特徴空間上での分布の“重なり度合い”

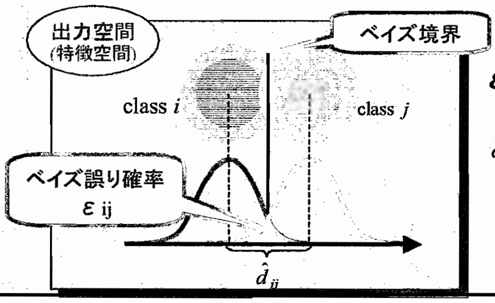


2003/1/21

2.適応的線形次元圧縮法(LL法)

7

- Lotlikarの仮定
  - 各クラスの分布に次の仮定をおく
    - (a) 等しい分散を持つ正規分布
    - (b) 事前確率が等しい
  - 線形写像行列Wに $\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I}$ の制約条件を与える
- 2クラス問題のベイズ境界とベイズエラー



$$\varepsilon_{ij}=\int_{\hat{d}_{ij}/2}^{\infty}\exp\left(-t^2/2\sigma^2\right)dt$$

$\hat{d}_{ij}$：出力空間における平均間のユークリッド距離

2003/1/21

2.適応的線形次元圧縮法(LL法)

8

- 多クラス問題のベイズエラー
  - 各ペア間における誤識別確率 ε_{ij} の和＝J を最小化

$$J=\sum_{i=1}^{c}\sum_{j\neq i}^{c}\varepsilon_{ij}\qquad \varepsilon_{ij}=\int_{\hat{d}_{ij}/2}^{\infty}\exp\left(-t^2/2\sigma^2\right)dt$$

- 学習アルゴリズム -目的関数 J を勾配法で最適化

Wの初期値をランダムに決める

⇩

Wの列ベクトルを直交化

$\mathbf{W}=\mathbf{W}(\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1/2}$

⇩ ⇧ 学習定数

Wの更新式

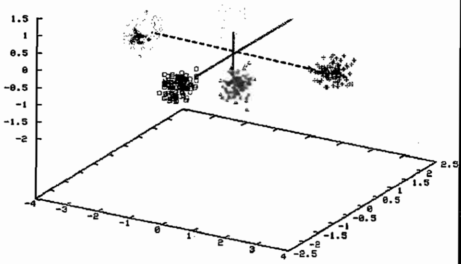
$$W_{new}=W_{old}-\eta\frac{\partial J}{\partial W}$$

2003/1/21

2. 実験1　線形適応的次元圧縮法(LL法)

9

◆ 実験条件



使用サンプル	人工サンプル
クラス数	6
サンプル数	100/クラス
入力次元	3
出力次元	2
比較手法	LL法 Fisherの線形判別分析
各クラスの分布	σ=0.2 の正規分布

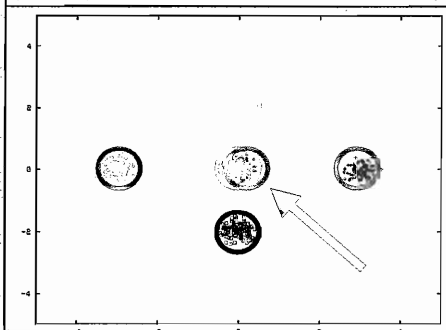
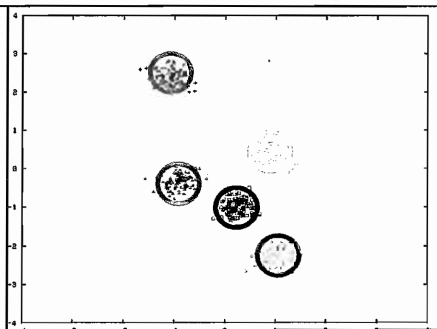
各クラスの平均

$$C_1=\begin{pmatrix}-3\\0\\0\end{pmatrix},C_2=\begin{pmatrix}3\\0\\0\end{pmatrix},C_3=\begin{pmatrix}0\\-2\\0\end{pmatrix},C_4=\begin{pmatrix}0\\2\\0\end{pmatrix},C_5=\begin{pmatrix}0\\0\\-1\end{pmatrix},C_6=\begin{pmatrix}0\\0\\1\end{pmatrix}$$

2003/1/21

2. 実験1　線形適応的次元圧縮法(LL法)

10

結果 - 出力2次元　6クラス -	
	
Fisherの線形判別分析(LF法)	適応的線形次元圧縮法(LL法)

2003/1/21

2. 実験2　線形適応的次元圧縮法（LL法）

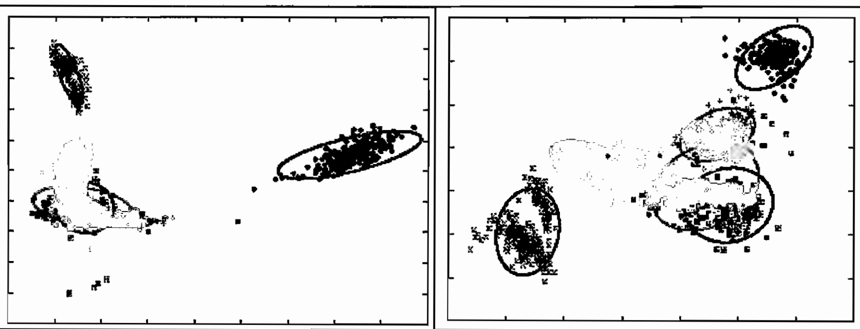
◆ 実験条件

使用サンプル	Imageデータ
クラス数	7（7種類のテクスチャ）
学習サンプル	30個/クラス
テストサンプル	300個/クラス
入力次元	18
出力次元	2

比較手法	**・LL法** ・Fisherの線形判別分析（LF法）
ハイパーパラメータ	・5-fold cross validation
評価方法	・k-NN法による認識率 ・出力空間のおけるテストサンプルの分布の重なり
前処理	クラス内変動行列が単位行列となるように白色化

2. 実験2　線形適応的次元圧縮法（LL法）

結果　- 出力2次元　7クラス

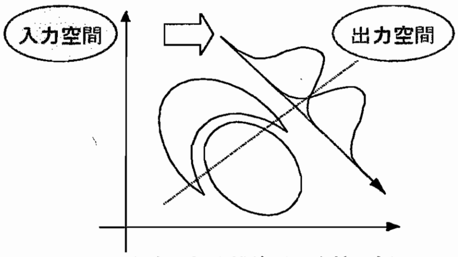


Fisherの線形判別分析(LF法)	適応的線形次元圧縮法(LL法)
73.6%	82.8%

2.適応的線形次元圧縮法　まとめ

13

- 2章のまとめ
  - LL法の紹介と追実験
    - Fisherの線形判別分析よりも**識別に有利な出力空間**が求まる事を確認
  - LL法でも分布が入り組んだ所は各クラスの分布が重なってしまった
    - 識別により有利な出力空間を求めるためには**非線形な変換**が必要



非線形な分離境界面を持つ例

2003/1/21

3.1 適応的非線形次元圧縮法

14

- 適応的非線形次元圧縮法（提案手法）

非線形写像　　線形写像

入力空間（原特徴空間） ⇒ ヒルベルト空間 ⇒ 出力空間（特徴空間）

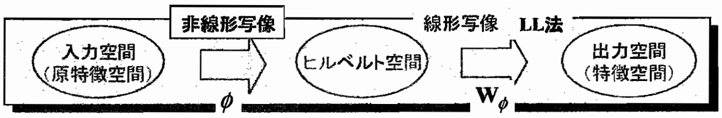
- ヒルベルト空間の像をLL法に適用
  - Lotlikarの制約条件
    - **(仮定a)　等しい分散を持つ正規分布**
    - (仮定b)　事前確率が等しい
  - ヒルベルト空間の像の分布が(仮定a)を
    - 満たす場合　：3.2 **適応的カーネル次元圧縮法 (KL法)**
    - 満たさない場合：3.3 **適応的白色カーネル次元圧縮法(WKL法)**

2003/1/21

3.2 適応的カーネル次元圧縮法(KL法)

15

- 適応的カーネル次元圧縮法 (Kernel Lotlikar: KL法)
  - ヒルベルト空間から出力空間への線形写像Wϕを求める



- クラス iに属するヒルベルト空間における訓練サンプル

  $\mathbf{X}_\phi^i \equiv [\phi(\mathbf{x}^{i,1}),\phi(\mathbf{x}^{i,2})...,\phi(\mathbf{x}^{i,Ni})]$

  Ni:クラスiの訓練サンプル数

- 訓練サンプル全体

  $\mathbf{X}_\phi \equiv [\mathbf{X}_\phi^1, \mathbf{X}_\phi^2,..., \mathbf{X}_\phi^c]$

  c:クラス数

- Wϕが訓練サンプルで張られる部分空間に含まれる

  $$\mathbf{W}_\phi = \mathbf{X}_\phi \mathbf{W}$$

  - 求めるパラメータ:

    $\mathbf{W} \in R^{N \times m}$

    N:訓練サンプル数
    m:出力次元数

2003/1/21

3.2 適応的カーネル次元圧縮法(KL法)

16

- ベイズエラー - 仮定を満たすのでLL法と同様

  $$\varepsilon_{ij} = \int_{\hat{d}_{ij}/2}^{\infty} \exp\left(-t^2/2\sigma^2\right)dt$$

  $\hat{d}_{ij}$:出力空間における平均間のユークリッド距離

  - 出力空間におけるクラス平均

    $$\hat{\boldsymbol{\mu}}_i = \mathbf{W}_\phi^T \underline{1/N_i \mathbf{X}_\phi^i \mathbf{1}_{Ni}}$$

    $$= 1/N_i \mathbf{W}^T \mathbf{X}_\phi^T \mathbf{X}_\phi^i \mathbf{1}_{Ni}$$

    $$= 1/N_i \mathbf{W}^T \mathbf{K}_{(\cdot,i)} \mathbf{1}_{Ni}$$

    1 を Ni個並べた縦ベクトル

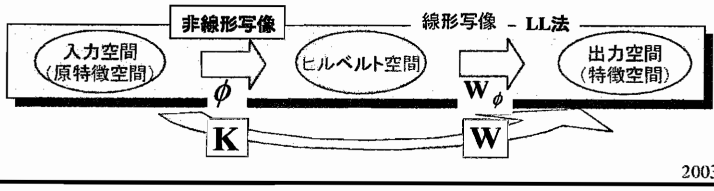
    クラスiの平均

    $\mathbf{K} = \mathbf{X}_\phi^T \mathbf{X}_\phi \in R^{N \times N}$

    :ヒルベルト空間における内積を並べた行列

    $\mathbf{K}_{(\cdot,i)} = \mathbf{X}_\phi^T \mathbf{X}_\phi^i \in R^{N \times N_i}$

    :$\mathbf{K} = \mathbf{X}_\phi^T \mathbf{X}_\phi \in R^{N \times N}$の部分行列

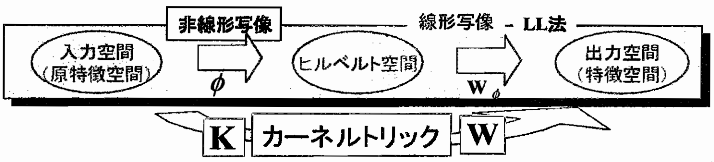


2003/1/21

3.2 適応的カーネル次元圧縮法(KL法)

17

- 行列Kの計算
  - K : カーネル行列 : ヒルベルト空間における内積の行列
- カーネルトリック
  - ヒルベルト空間における内積をカーネル k で置き換える
    $\phi(\mathbf{x}_1)^{\mathrm{T}}\phi(\mathbf{x}_2)=k(\mathbf{x}_1,\mathbf{x}_2)$
  - ϕを計算することなく内積を計算 – 計算量を削減
  - 出力空間におけるクラス平均をϕを計算することなく計算可能



2003/1/21

3.2 適応的カーネル次元圧縮法(KL法)

18

- カーネルの条件(Mercer の定理)
  - ϕが実在する k を選ぶ必要がある
    - ガウスカーネル $k(\mathbf{x}_1,\mathbf{x}_2)=\exp(\|\mathbf{x}_1-\mathbf{x}_2\|^2/\sigma^2_{\mathrm{kernel}})$
    - 多項式カーネル $k(\mathbf{x}_1,\mathbf{x}_2)=(\mathbf{x}_1^T\mathbf{x}_2+1)^p$
- 学習アルゴリズム 目的関数Jを最小化するWを勾配法で求める

訓練サンプルから K を計算する

⇩

Wの初期値をランダムに決める

⇩

$\mathbf{W}_\phi$の列ベクトルを直交化 : $\mathbf{W}=\mathbf{W}(\mathbf{W}^{\mathrm{T}}\mathbf{K}\mathbf{W})^{-1/2}$

⇩

Wの更新 : $W_{new}=W_{old}-\eta\dfrac{\partial J}{\partial W}$

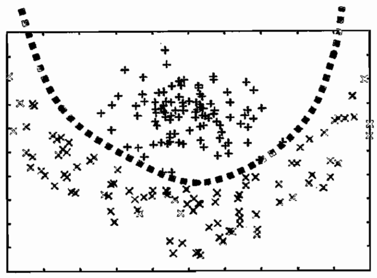
2003/1/21

3.2 実験 適応的カーネル次元圧縮法(KL法)

19

◆ 実験条件

使用サンプル	人工サンプル
クラス数	2
サンプル数	100/クラス
入力次元	2
出力次元	1
提案手法	**KL法**
比較手法	LL法
各クラスの分布	右図を参照
使用カーネル	ガウスカーネル

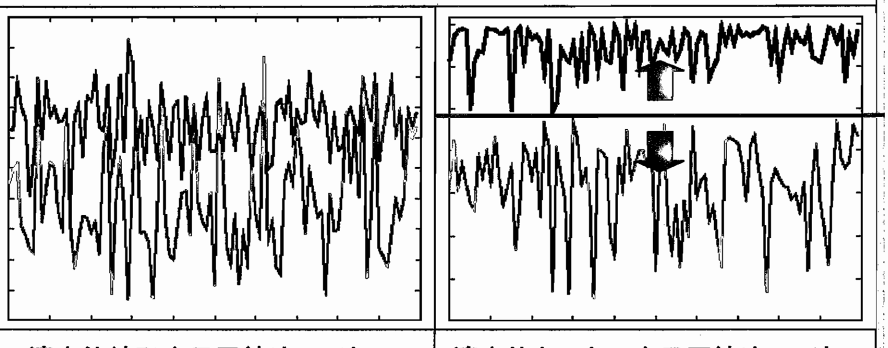


2003/1/21

3.2 実験 適応的カーネル次元圧縮法(KL法)

20

結果 -出力1次元 2クラス- 横軸:サンプル番号 縦軸:出力値

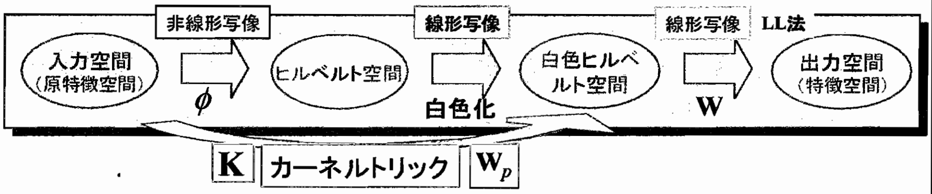


適応的線形次元圧縮法(LL法)	適応的カーネル次元圧縮法(KL法)
83.5%	100.0%

2003/1/21

3.3 適応的白色カーネル次元圧縮法(WKL法)

- 適応的白色カーネル次元圧縮法(White Kernel Lotlikar:WKL法)
  - ヒルベルト空間の像が仮定(a)を満たさない場合
    - **STEP1** 訓練サンプルをヒルベルト空間において白色化する.
    - **STEP2** 白色化した訓練サンプルをLL法に適用する.



3.3 適応的白色カーネル次元圧縮法(WKL法)

- STEP1 クラス内分散の白色化
  - ヒルベルト空間において白色化された訓練サンプル $\mathbf{X}_{\phi w}$

$$\mathbf{X}_{\phi w} = \mathbf{W}_\mathrm{p}\mathbf{K} \in R^{r\times N}$$

$$\boxed{\mathbf{W}_\mathrm{p} = \mathbf{\Lambda}^{-1}\mathbf{U}^\mathrm{T}(\mathbf{I}-\mathbf{M})}$$

$\mathbf{M}\in R^{N\times N}$: c個のブロックを持つブ ロック対角行列

第iブロックは$N_i \times N_i$行列で全要素が $1/N_i$

$\mathbf{\Lambda}\in R^{r\times r}$: $(\mathbf{I}-\mathbf{M})\mathbf{K}(\mathbf{I}-\mathbf{M})$の固有値からなる対角行列($r<N$)

$\mathbf{U}\in R^{N\times r}$: $(\mathbf{I}-\mathbf{M})\mathbf{K}(\mathbf{I}-\mathbf{M})$の固有ベクトルからなる行列

$\mathbf{K} = \mathbf{X}_\phi^T\mathbf{X}_\phi \in R^{N\times N}$: カーネル行列

- STEP2 LL法
  - 白色化した訓練サンプルをLL法により適用
  - 写像Wを求める $\mathbf{W}\in R^{r\times m}$

3.3 適応的白色カーネル次元圧縮法(WKL法)

23

◆ 学習アルゴリズム

訓練サンプルから K を計算する

⇩

Wpを求め, $X_{\phi w}$を計算する

⇩

Wの初期値をランダムに決める

⇩

Wの列ベクトルを直交化 ： $\mathbf{W} = \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1/2}$

⇩

Wの更新 : $W_{new} = W_{old} - \eta \frac{\partial J}{\partial W}$

STEP1
白色化

STEP2
LL法

2003/1/21

3.4 実験

24

◆ 実データを用いた提案手法の評価を行う

◆ 実験条件

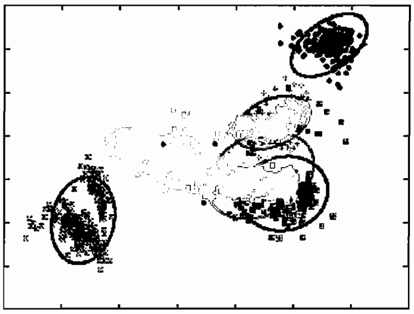
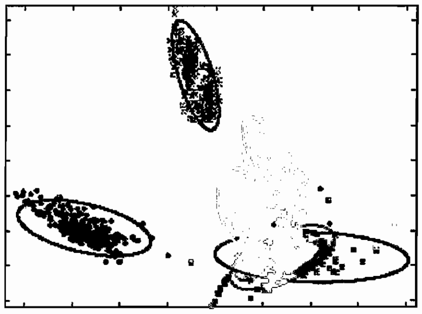
使用サンプル	Imageデータ
クラス数	7 (7種類のテクスチャ)
学習サンプル	30個/クラス
テストサンプル	300個/クラス
入力次元	18
出力次元	2

提案手法	**・WKL法** **・KL法**
比較手法	・LL法(従来法)
ハイパーパラメータ	・5-fold cross validation
使用カーネル	・ガウスカーネル ・多項式カーネル
評価方法	・k-NN法による認識率 ・出力空間のおけるテストサンプルの分布の重なり

2003/1/21

3.4 実験　KL法とLL法の比較

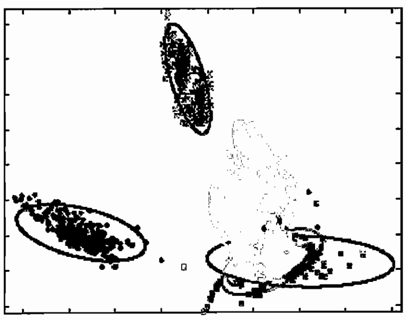
25

結果 - 出力2次元 7クラス	
	
LL法 (従来法)	KL法 （提案手法1）
82.8%	86.0%

2003/1/21

3.4 実験　WKL法とKL法の比較

26

結果 - 出力2次元 7クラス	
	
KL法 （提案手法1）	WKL法 （提案手法2）
86.0%	92.7%

2003/1/21

3.適応的非線形次元圧縮法

27

- 3章のまとめ
  - 適応的線形次元圧縮法を2通りの方法で非線形に拡張した
  - KL法
    - 非線形な境界面をもつ問題でも認識に適した出力空間を得た
  - WKL法
    - ヒルベルト空間の像の分布が仮定(a)を満たさない場合に対応
    - Imageに対して識別により適した出力空間を得た

2003/1/21

4.評価実験

28

- 提案手法の有効性を公開データベースを用いて確認
- 使用データベースと実験条件

データベース名	次元数	クラス数	訓練サンプル数/クラス	テストサンプル数/クラス
Image Segmentation	18	7	30	300
LED25	25	10	30	300
LED7	7	10	30	300
Multiple Features	47	10	30	300
Pen recognition	16	10	30	300
Banana	2 (1)	2	100	1000

提案手法	**・KL法** **・WKL法**
比較手法	・LL法 ・Fisherの線形判別分析(LF法) ・カーネル判別分析(KF法)
出力次元数	m=2
ハイパーパラメータ	・5-fold cross validation
使用カーネル	・ガウスカーネル ・多項式カーネル
評価方法	・k-NN法による認識率

2003/1/21

4.評価実験

29

◆ 実験結果

提案手法

データベース名	LF法	KF法	LL法	KL法	WKL法
Image	73.6	71.6	82.8	86.0	92.9
led7	76.2	75.6	76.0	77.1	86.5
led25	77.5	74.9	77.6	77.0	75.5
Multiple Feature	44.6	36.4	49.6	42.2	61.6
Pen recognition	65.3	67.9	67.1	74.6	80.6
banana	67.5	82.9	68.8	84.2	87.0

認識率第2位　認識率第1位

2003/1/21

4.評価実験 まとめ

30

- 第4章のまとめ
  - KL法、WKL法が良い性能を発揮したデータベース
    - Image (16,10), LED7 (7,10), Pen Recognition (16,10)
  - WKL法が良い性能を発揮したデータベース
    - Multiple Feature (47,10)
  - 性能が向上しなかったデータベース
    - LED25 (25,10)

(次元数，クラス数)

- 考察
  - 性能が向上したデータベースに比べ，向上しなかったデータベースは次元数が高い
  - ヒルベルト空間でのクラスの分離度が入力空間よりも向上しなかったことが原因と考えられる

ヒルベルト空間でのクラスの分離度が高まったとき
提案手法はより識別に適した出力空間を求める

2003/1/21

5.まとめと今後の方針

31

- まとめ
  - 特徴空間(ヒルベルト空間)を介することで非線形な教師付き次元圧縮を行うKL法、WKL法を提案した
  - 公開データベースを用いた評価実験により有効性を確認した

- 今後の方針
  - Lotlikarの仮定
    - (仮定a) 等しい分散を持つ正規分布
    - (仮定b) 事前確率が等しい
  - Lotlikarの仮定を満たさない場合へのLL法の拡張
    - ベイズエラーの定義し直し，新しい学習則の導出
    - 評価実験

2003/1/21

評価実験　訓練サンプルについて

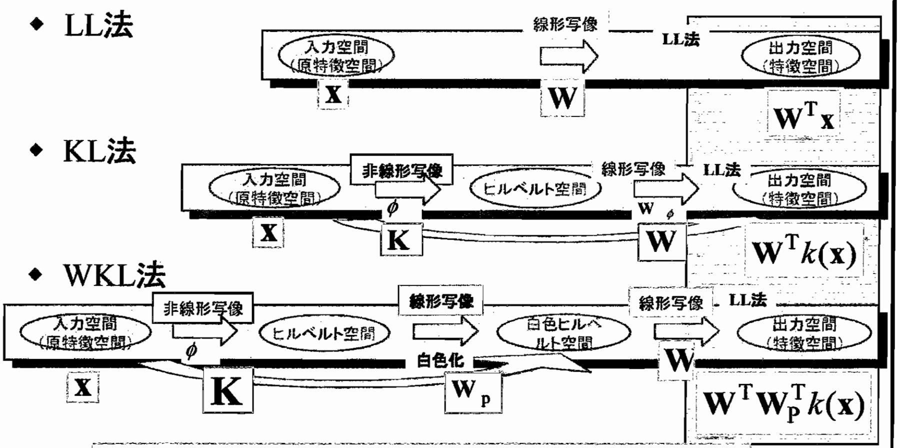
32

提案手法

テストサンプル	LF法	KF法	LL法	KL法	WKL法
Image	73.6	71.6	82.8	86.0	92.9
led7	76.2	75.6	76.0	77.1	86.5
led25	77.5	74.9	77.6	77.0	75.5
Multiple Feature	44.6	36.4	49.6	42.2	61.6
Pen recognition	65.3	67.9	67.1	74.6	80.6
banana	67.5	82.9	68.8	84.2	87.0
訓練サンプル	LF法	KF法	LL法	KL法	WKL法
Image	79.1	80.5	82.6	87.1	95.7
led7	83.0	82.5	82.5	81.5	83.5
led25	83.5	85.0	84.5	84.5	91.5
Multiple Feature	44.0	44.5	56.0	45.0	85.5
Pen recognition	69.5	74.0	70.0	80.5	98.0
banana	68.0	90.0	71.5	90.5	92.0

出力空間への写像について

33



$k(\mathbf{x}) = \mathbf{X}_\phi^T \phi(\mathbf{x})$

訓練サンプルの各列と$\mathbf{x}$とのヒルベルト空間における内積

2003/1/21

多項式カーネルについて $\boxed{k(\mathbf{x},\mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y})}$

- 多項式カーネルのϕの形（p = 2のとき）

  $k(\mathbf{x},\mathbf{y}) = (\mathbf{x}^T \mathbf{y} + 1)^p$

- xの次元数 = 2

$$\phi(\mathbf{x}) = \begin{pmatrix} 1 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \\ x_1^2 \\ x_2^2 \\ x_1 x_2 \end{pmatrix}$$

xの次元数 = n

$$\phi(\mathbf{x}) = \begin{pmatrix} 1 \\ \sqrt{2}x_1 \\ \ldots \\ \sqrt{2}x_n \\ x_1 x_1 \\ x_1 x_2 \\ \ldots. \\ x_n x_n \end{pmatrix}$$

2003/1/21

ガウスカーネルについて $k(\mathbf{x},\mathbf{y}) = \phi(\mathbf{x})^{\mathrm{T}} \phi(\mathbf{y})$

- ガウスカーネルの ϕ の形

  $k(\mathbf{x}_1,\mathbf{x}_2) = \exp\left(\|\mathbf{x}_1 - \mathbf{x}_2\|^2 / \sigma^2_{\mathrm{kernel}}\right)$

  $\phi(\mathbf{x}) = \exp\left(\|\mathbf{t} - \mathbf{x}\| / \sigma^2_{\mathrm{ker\,nel}}\right)$

  $\mathbf{x}$ を中心とするガウスカーネル関数への写像となる

- ヒルベルト空間における内積

$$\langle \mathbf{x},\mathbf{y} \rangle = \int \exp\left(\|\mathbf{t} - \mathbf{x}\|^2 / \sigma^2_{\mathrm{ker\,nel}}\right) \exp\left(\|\mathbf{t} - \mathbf{y}\|^2 / \sigma^2_{\mathrm{ker\,nel}}\right) d\mathbf{t}$$
$$= C \exp\left(\|\mathbf{x} - \mathbf{y}\|^2 / 2\sigma^2_{\mathrm{ker\,nel}}\right)$$

2003/1/21

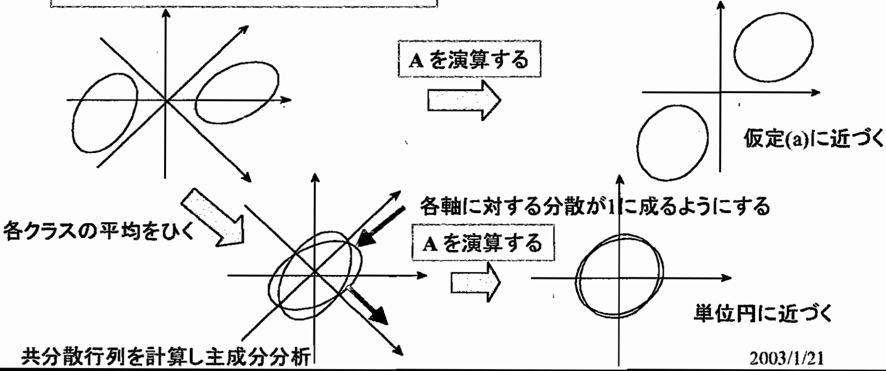
クラス内分散の白色化

36

- クラス内変動行列の白色化

$$S_w = \frac{1}{N} \sum_{i=1}^{c} \sum_{x_j \in \omega_i} (\mathbf{x}_j - \boldsymbol{\mu}_i)(\mathbf{x}_j - \boldsymbol{\mu}_i)^{\mathrm{T}}$$

$\mathbf{A}^{\mathrm{T}} \mathbf{S}_w \mathbf{A} = \mathbf{I}$ なる $\mathbf{A}$ を演算する



2003/1/21