

使ってみよう部分空間法！ - eingang -

堀田政二 (東京農工大学) 河原智一 (東芝) 坂野鋭 (NTT)

部分空間法研究会 2010

2010 年 7 月 26 日

目的と内容

パターン認識法の一つである部分空間法 (CLAFIC) を理解することを目指すとし、そのために必要な以下の内容について解説する：

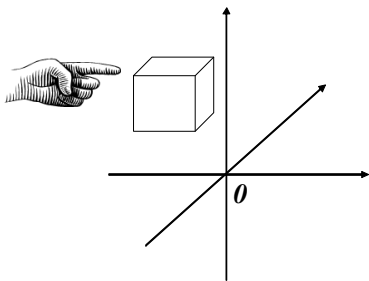
- 数学的準備 (ベクトルによる偏微分，ラグランジュ乗数法)
- 部分空間法 (CLAFIC の直感的な理解．相互部分空間法，非線型部分空間法への橋渡し)
- MATLAB/Octave を用いた手書き数字パターン認識の実験

本講演で使用するプログラムは以下のページからダウンロード可能：

<http://www.tuat.ac.jp/~s-hotta/ss2010>

線型部分空間とはなんですか？ 1/3

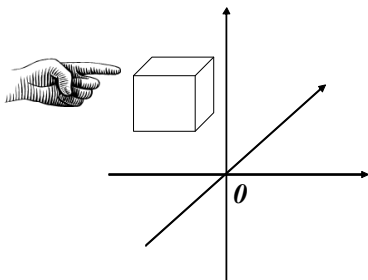
これは部分空間ですか？ [1]



- これは空間の一部で “部分空間” ではない
- 英語では subspace . sub は (身分や質が) 下 , 下位等の意味 (\Longleftrightarrow super)
- もとの空間よりも次元数 (空間に置ける座標軸の本数) が小さい空間 (上の図では 2 , 1 , 0 次元)

線型部分空間とはなんですか？ 1/3

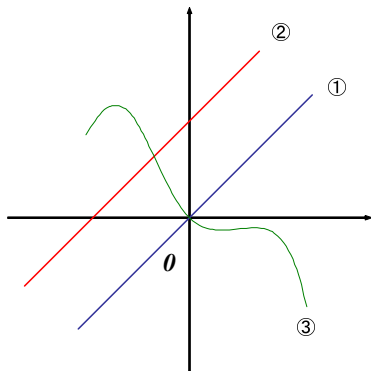
これは部分空間ですか？ [1]



- これは空間の一部分で “部分空間” ではない
- 英語では subspace . sub は (身分や質が) 下 , 下位等の意味 (\Longleftrightarrow super)
- もとの空間よりも次元数 (空間に置ける座標軸の本数) が小さい空間 (上の図では 2 , 1 , 0 次元)

線型部分空間とはなんですか？ 2/3

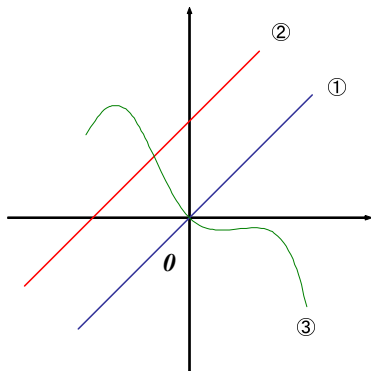
どれが線型部分空間ですか？



- ①が正解
- 足し算・引き算・スカラー倍が可能な集合

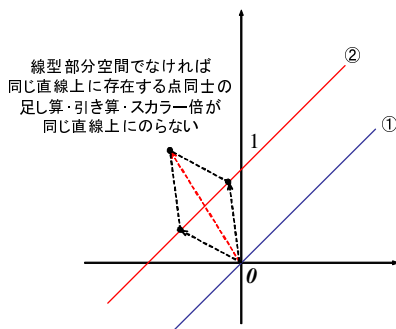
線型部分空間とはなんですか？ 2/3

どれが線型部分空間ですか？



- ①が正解
- 足し算・引き算・スカラー倍が可能な集合

線型部分空間とはなんですか？ 3/3



- ②と③は点同士の足し算，スカラー倍が外に飛び出す
- ②は集合上の任意の点を独自の原点と定めれば線型部分空間のように扱える (affine subspace, linear manifold, linear variety)

内積とノルム

本講演ではパターンを d 次元の縦実ベクトルで表現:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}, \quad \mathbf{x}^\top = (x_1, \dots, x_d)$$

- 内積の例:

$$\mathbf{x}^\top \mathbf{x} = \sum_{i=1}^d x_i^2 = \|\mathbf{x}\|^2$$

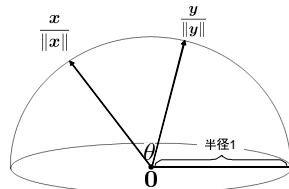
- \mathbf{x} のノルム: $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\top \mathbf{x}}$

- 二つのベクトルのなす角度:

$$\cos \theta = \frac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

- 正規化されたパターン同士のユークリッド距離:

$$\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|} - \frac{\mathbf{y}}{\|\mathbf{y}\|} \right\|^2 = 2(1 - \cos \theta)$$

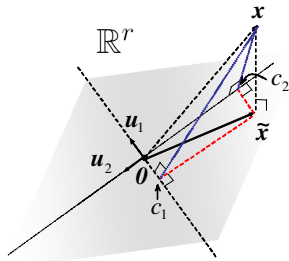


$n + 1$ 次元空間における n 次元単位球面

($n \geq 3$ のとき超球面と呼ぶ)

部分空間への正射影

r 次元線型部分空間を張る d 次元正規直交ベクトルを u_1, u_2, \dots, u_r とし, それらを並べた $d \times r$ の行列 (部分等長行列) を $U = (u_1 | u_2 | \dots | u_r)$ とする ($U^\top U = I$)



- \mathbb{R}^r での \tilde{x} の座標: $c = U^\top x = (u_1^\top x | u_2^\top x | \dots | u_r^\top x)^\top$
- \mathbb{R}^d での \tilde{x} の座標: $\tilde{x} = Uc = UU^\top x$ (x の U による展開)
- UU^\top は直交射影行列と呼ばれる

ベクトルによる偏微分

$\frac{\partial}{\partial \mathbf{x}}$ は \mathbf{x} に関する偏微分を表し, その第 i 成分が $\frac{\partial}{\partial x_i}$ となるベクトル

$$\frac{\partial}{\partial \mathbf{x}} = \left(\frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_d} \right)^\top$$

- 例 (1): 内積 $f = \mathbf{x}^\top \mathbf{x} = \sum_{i=1}^d x_i^2$ を \mathbf{x} で偏微分すると
$$\frac{\partial}{\partial \mathbf{x}} f = \nabla f = \left(\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_d} \right)^\top = (2x_1 \cdots 2x_d)^\top = 2\mathbf{x}$$
- 例 (2): $d \times d$ の対称行列を \mathbf{A} としたとき, 二次形式 $f = \mathbf{x}^\top \mathbf{A} \mathbf{x}$ を \mathbf{x} で偏微分すると $\frac{\partial}{\partial \mathbf{x}} f = 2\mathbf{A} \mathbf{x}$
- 例 (3): 双一次形式 $f = \mathbf{x}^\top \mathbf{A} \mathbf{y}$ を \mathbf{x} で偏微分すると $\frac{\partial}{\partial \mathbf{x}} f = \mathbf{A} \mathbf{y}$

\arg , \max , subject to の意味

それぞれどのような意味でしょうか？

■ $\max f(x)$

$f(x)$ の最大値

■ $\max_{\|x\|=1} f(x)$

$\|x\| = 1$ を満たす x が与える $f(x)$ の最大値

■ $\operatorname{argmax}_x f(x)$

$f(x)$ を最大にする x の集合．この \arg は偏角ではなく，引数という意味． $\operatorname{argmin}_x f(x)$ も同様に定義できる

■ $\text{subject to } \dots$

\dots のもとで，という意味．s.t. と略記する場合もある

■ $\max_x f(x), \text{ s.t. } \dots$

x に関する制約条件下での $f(x)$ の最大値

\arg , \max , subject to の意味

それぞれどのような意味でしょうか？

■ $\max f(x)$

$f(x)$ の最大値

■ $\max_{\|x\|=1} f(x)$

$\|x\| = 1$ を満たす x が与える $f(x)$ の最大値

■ $\operatorname{argmax}_x f(x)$

$f(x)$ を最大にする x の集合．この \arg は偏角ではなく，引数という意味． $\operatorname{argmin}_x f(x)$ も同様に定義できる

■ $\text{subject to } \dots$

\dots のもとで，という意味．s.t. と略記する場合もある

■ $\max_x f(x), \text{ s.t. } \dots$

x に関する制約条件下での $f(x)$ の最大値

\arg , \max , subject to の意味

それぞれどのような意味でしょうか？

■ $\max f(x)$

$f(x)$ の最大値

■ $\max_{\|x\|=1} f(x)$

$\|x\| = 1$ を満たす x が与える $f(x)$ の最大値

■ $\operatorname{argmax}_x f(x)$

$f(x)$ を最大にする x の集合．この \arg は偏角ではなく，引数という意味． $\operatorname{argmin}_x f(x)$ も同様に定義できる

■ $\text{subject to } \dots$

\dots のもとで，という意味．s.t. と略記する場合もある

■ $\max_x f(x), \text{ s.t. } \dots$

x に関する制約条件下での $f(x)$ の最大値

\arg , \max , subject to の意味

それぞれどのような意味でしょうか？

■ $\max f(x)$

$f(x)$ の最大値

■ $\max_{\|x\|=1} f(x)$

$\|x\| = 1$ を満たす x が与える $f(x)$ の最大値

■ $\operatorname{argmax}_x f(x)$

$f(x)$ を最大にする x の集合．この \arg は偏角ではなく，引数という意味． $\operatorname{argmin}_x f(x)$ も同様に定義できる

■ $\text{subject to } \dots$

\dots のもとで，という意味．s.t. と略記する場合もある

■ $\max_x f(x), \text{ s.t. } \dots$

x に関する制約条件下での $f(x)$ の最大値

\arg , \max , subject to の意味

それぞれどのような意味でしょうか？

■ $\max f(x)$

$f(x)$ の最大値

■ $\max_{\|x\|=1} f(x)$

$\|x\| = 1$ を満たす x が与える $f(x)$ の最大値

■ $\operatorname{argmax}_x f(x)$

$f(x)$ を最大にする x の集合．この \arg は偏角ではなく，引数という意味． $\operatorname{argmin}_x f(x)$ も同様に定義できる

■ $\text{subject to } \dots$

\dots のもとで，という意味．s.t. と略記する場合もある

■ $\max_x f(x), \text{ s.t. } \dots$

x に関する制約条件下での $f(x)$ の最大値

$\arg, \max, \text{subject to}$ の意味

それぞれどのような意味でしょうか？

- $\max f(x)$
 $f(x)$ の最大値
- $\max_{\|x\|=1} f(x)$
 $\|x\| = 1$ を満たす x が与える $f(x)$ の最大値
- $\operatorname{argmax}_x f(x)$
 $f(x)$ を最大にする x の集合．この \arg は偏角ではなく，引数
 という意味． $\operatorname{argmin}_x f(x)$ も同様に定義できる
- $\text{subject to } \dots$
 \dots のもとで，という意味．s.t. と略記する場合もある
- $\max_x f(x), \text{ s.t. } \dots$
 x に関する制約条件下での $f(x)$ の最大値

$\arg, \max, \text{subject to}$ の意味

それぞれどのような意味でしょうか？

- $\max f(x)$
 $f(x)$ の最大値
- $\max_{\|x\|=1} f(x)$
 $\|x\| = 1$ を満たす x が与える $f(x)$ の最大値
- $\operatorname{argmax}_x f(x)$
 $f(x)$ を最大にする x の集合．この \arg は偏角ではなく，引数
 という意味． $\operatorname{argmin}_x f(x)$ も同様に定義できる
- $\text{subject to } \dots$
 \dots のもとで，という意味．s.t. と略記する場合もある
- $\max_x f(x), \text{ s.t. } \dots$
 x に関する制約条件下での $f(x)$ の最大値

$\arg, \max, \text{subject to}$ の意味

それぞれどのような意味でしょうか？

- $\max f(x)$
 $f(x)$ の最大値
- $\max_{\|x\|=1} f(x)$
 $\|x\| = 1$ を満たす x が与える $f(x)$ の最大値
- $\operatorname{argmax}_x f(x)$
 $f(x)$ を最大にする x の集合．この \arg は偏角ではなく，引数
 という意味． $\operatorname{argmin}_x f(x)$ も同様に定義できる
- $\text{subject to } \dots$
 \dots のもとで，という意味．s.t. と略記する場合もある
- $\max_x f(x), \text{ s.t. } \dots$
 x に関する制約条件下での $f(x)$ の最大値

$\arg, \max, \text{subject to}$ の意味

それぞれどのような意味でしょうか？

- $\max f(x)$
 $f(x)$ の最大値
- $\max_{\|x\|=1} f(x)$
 $\|x\| = 1$ を満たす x が与える $f(x)$ の最大値
- $\operatorname{argmax}_x f(x)$
 $f(x)$ を最大にする x の集合．この \arg は偏角ではなく，引数
 という意味． $\operatorname{argmin}_x f(x)$ も同様に定義できる
- $\text{subject to } \dots$
 \dots のもとで，という意味．s.t. と略記する場合もある
- $\max_x f(x), \text{ s.t. } \dots$
 x に関する制約条件下での $f(x)$ の最大値

$\arg, \max, \text{subject to}$ の意味

それぞれどのような意味でしょうか？

- $\max f(x)$
 $f(x)$ の最大値
- $\max_{\|x\|=1} f(x)$
 $\|x\| = 1$ を満たす x が与える $f(x)$ の最大値
- $\operatorname{argmax}_x f(x)$
 $f(x)$ を最大にする x の集合．この \arg は偏角ではなく，引数
 という意味． $\operatorname{argmin}_x f(x)$ も同様に定義できる
- $\text{subject to } \dots$
 \dots のもとで，という意味．s.t. と略記する場合もある
- $\max_x f(x), \text{ s.t. } \dots$
 x に関する制約条件下での $f(x)$ の最大値

ラグランジュ未定乗数法の解法レシピ

制約条件のもとで関数の極値を求める方法の一つ [2, 3] . 主成分分析, SVM の導出等, 知っていれば多くのパターン認識に関する問題が解ける

問題設定

制約条件 $g(x) = 0$ のもとで関数 $f(x)$ の極値を求めよ

- ラグランジュ乗数 λ を用いてラグランジュ関数を導入

$$L = f(x) - \lambda g(x)$$

- 制約条件のもとで関数が極値をとる点は次式を満たす¹ :

$$\frac{\partial}{\partial x} L = \nabla f - \lambda \nabla g = \mathbf{0}, \quad \frac{\partial L}{\partial \lambda} = 0$$

- 上記から $d + 1$ 個の方程式が得られる . 一方, 未知数は x_1, \dots, x_d, λ の $d + 1$ 個なので, 方程式の解を求めることができる

¹付録 1 参照

ラグランジュ未定乗数法の例

(部分空間法で頻出する例) A を半正定値対称行列とする. $u^\top u = 1$ の制約条件のもと, $u^\top A u$ の最大値を求めよ

$$\begin{aligned} \max_u \quad & u^\top A u \\ \text{s.t.} \quad & u^\top u - 1 = 0 \end{aligned}$$

- ラグランジュ乗数 λ を用いてラグランジュ関数を導入

$$L = u^\top A u - \lambda(u^\top u - 1)$$

- 制約条件のもとで関数が極値をとる点は次式を満たす:

$$\partial L / \partial u = 0, \quad \partial L / \partial \lambda = 0$$

- $\partial L / \partial u = 2Au - 2\lambda u = 0$ より, 解は以下を満たす:

$$Au = \lambda u$$

これは固有値問題². 求める解は最大固有値に対応する A の固有ベクトル (左から u^\top を掛けてみよう. $u^\top u = 1$ に注意)

²固有値, 固有ベクトルについては付録 2 を参照

部分空間法 - CLAss-Featuring Information Compression -

クラスらしさを部分空間で表現する方法 [4, 5] . 部分空間法は主に三つの観点から独立に見出された経緯がある

- クラスの特徴を統計的に抽出する [6]
- 視覚情報に関する理論から [7]
- 統計学からみて自然な発想 (縮退ガウス分布)

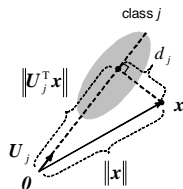
CLAFIC の識別則

- 問題設定：総数 n 個の d 次元訓練パターン x_i ($i = 1, \dots, n$) が与えられており，それらが C 個のクラスのいずれか一つに属するとする
- クラス j に属する訓練パターンを良く近似できる部分空間を張る正規直交ベクトルを並べた行列を U_j とする
- 未知パターン x が与えられたとき， x を良く近似できる部分空間が属するクラスを以下の識別則に基づき出力する：

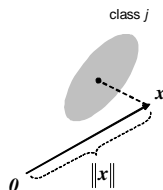
CLAFIC の識別則

$$\max_{j=1,\dots,C} \{\|U_j^\top x\|\} = \|U_k^\top x\| \Rightarrow x \in \text{class } k$$

CLAFIC における識別則の直感的な理解



部分空間法



最小距離法

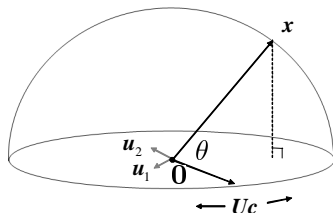
- CLAFIC における原点は各クラス共通
- 未知パターン x を部分空間を使って最も良く近似できるクラスへ分類する
 $\Rightarrow d_j^2 = \|x\|^2 - \|U_j^\top x\|^2$ が最小のクラスへ分類する．ただし， $\|x\|^2$ が全クラス共通なので，結局 $\|U_j^\top x\|^2$ が最大となるクラスへ分類すれば良い

$\cos \theta$ 最大化基準から見た CLAFIC 1/3

U を部分空間を張る正規直交ベクトルを並べた $d \times r$ の行列とする．ノルムが 1 に正規化された未知パターン x と U の線型結合パターン Uc とのなす角度の最大値を $c^\top c = 1$ の制約条件のもとで求めよ：

$$\begin{aligned} \max_{\mathbf{c}} \quad & \mathbf{x}^\top \mathbf{U} \mathbf{c} = \cos \theta \\ \text{s.t.} \quad & \mathbf{c}^\top \mathbf{c} - 1 = 0 \end{aligned} \quad (1)$$

なお $\|x\| = 1$, $\|Uc\| = \sqrt{(Uc)^\top (Uc)} = \sqrt{c^\top I c} = 1$, クラスの添え字 j を省略していることに注意



$\cos \theta$ 最大化基準から見た CLAFIC 2/3

- 式 (1) のラグランジュ関数を導入 $L = \mathbf{x}^\top \mathbf{U} \mathbf{c} - \frac{\lambda}{2} (\mathbf{c}^\top \mathbf{c} - 1)$
- $\partial L / \partial \mathbf{c} = \mathbf{U}^\top \mathbf{x} - \lambda \mathbf{c} = \mathbf{0}$ より, 解は以下を満たす:

$$\mathbf{c} = \frac{1}{\lambda} \mathbf{U}^\top \mathbf{x} \quad (2)$$

- 制約条件より

$$\mathbf{c}^\top \mathbf{c} = \frac{1}{\lambda^2} (\mathbf{U}^\top \mathbf{x})^\top (\mathbf{U}^\top \mathbf{x}) = \frac{1}{\lambda^2} \|\mathbf{U}^\top \mathbf{x}\|^2 = 1$$

となるから $\lambda = \|\mathbf{U}^\top \mathbf{x}\|$ であることがわかる

- 求めた λ を式 (2) に代入すれば

$$\mathbf{c} = \frac{\mathbf{U}^\top \mathbf{x}}{\|\mathbf{U}^\top \mathbf{x}\|}$$

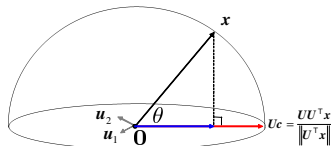
となる. これは \mathbf{x} を \mathbf{U} の張る部分空間へ正射影したパターンをノルム 1 に正規化したものに他ならない

$\cos \theta$ 最大化基準から見た CLAFIC 3/3

- 求めた $c = (\mathbf{U}^\top \mathbf{x}) / \|\mathbf{U}^\top \mathbf{x}\|$ を $\mathbf{x}^\top \mathbf{U} \mathbf{c}$ に代入すると

$$\cos \theta = \mathbf{x}^\top \mathbf{U} \mathbf{c} = \frac{\mathbf{x}^\top \mathbf{U} \mathbf{U}^\top \mathbf{x}}{\|\mathbf{U}^\top \mathbf{x}\|} = \frac{\|\mathbf{U}^\top \mathbf{x}\|^2}{\|\mathbf{U}^\top \mathbf{x}\|} = \|\mathbf{U}^\top \mathbf{x}\| = \lambda$$

したがって最大の $\cos \theta$ を与える部分空間上のパターンは $(\mathbf{U}^\top \mathbf{x}) / \|\mathbf{U}^\top \mathbf{x}\| \in \mathbb{R}^r$, または $(\mathbf{U} \mathbf{U}^\top \mathbf{x}) / \|\mathbf{U}^\top \mathbf{x}\| \in \mathbb{R}^d$ で与えられ, その $\cos \theta$ の値は $\|\mathbf{U}^\top \mathbf{x}\|$ である



相互部分空間法への拡張

未知パターン x を部分空間上のパターンに変更することで，相互部分空間法に容易に拡張できる： V を部分空間を張る正規直交ベクトルを並べた $d \times s$ の行列とする． V の線型結合パターン Vb と U の線型結合パターン Uc のなす角度の最大値を $b^\top b = 1, c^\top c = 1$ の制約条件のもとで求めよ：

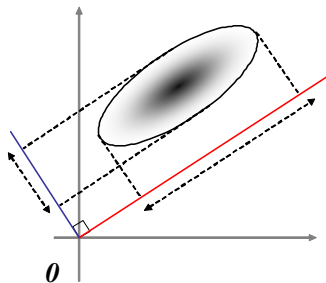
相互部分空間法への拡張

$$\begin{aligned} \max_{b,c} (Vb)^\top (Uc) \\ \text{s.t. } b^\top b - 1 = 0, c^\top c - 1 = 0 \end{aligned}$$

これもラグランジュ未定乗数法で解ける．詳しくは午後のチュートリアルを参照

Uをどのように求めるか？ 1/2

- 部分空間へ正射影したパターンのパラツキが大きい軸から r 本を選ぶ (下の例では赤の部分空間)



Uをどのように求めるか？ 2/2

- n 個の訓練パターン x_1, \dots, x_n が与えられた場合, はじめに $y_i = \mathbf{u}_1^\top \mathbf{x}_i$ のバラツキ (y_i^2) が最大となる \mathbf{u}_1 を求めることを考える:

$$\begin{aligned} \max_{\mathbf{u}_1} \quad & \sum_{i=1}^n (\mathbf{u}_1^\top \mathbf{x}_i)^2 = \mathbf{u}_1^\top \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top \right) \mathbf{u}_1 \\ \text{s.t.} \quad & \mathbf{u}_1^\top \mathbf{u}_1 - 1 = 0 \end{aligned}$$

- ラグランジュ未定乗数法から, 解は行列 $\mathbf{D} = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top$ の最大固有値 λ_1 に対応する固有ベクトルであることがわかる
- 一般に, バラツキの値が r 番目に大きい正規直交ベクトル \mathbf{u}_r は \mathbf{D} の r 番目に大きい固有値 λ_r に対応する固有ベクトルで与えられる
- クラスごとに r 次元の部分空間を張る固有ベクトルを並べた行列 $\mathbf{U}_j = (\mathbf{u}_1 | \dots | \mathbf{u}_r)$ を求め, それらを識別に用いる

次元数 d が大きい場合 (非線型部分空間法への橋渡し)

訓練パターンを並べた $d \times n$ の行列を $\mathbf{X} = (x_1 | \cdots | x_n)$ とする

- d が大きい場合, $d \times d$ の行列 $\mathbf{D} = \mathbf{X}\mathbf{X}^\top$ をメモリに格納するのは困難
- 固有値, 固有ベクトルを計算するのも大変
- $n \ll d$ のときは $n \times n$ の行列 $\mathbf{N} = \mathbf{X}^\top \mathbf{X}$ の固有ベクトルと固有値から \mathbf{U} を計算した方が効率的

固有ベクトルを求めるだけならば $\mathbf{X}^\top \mathbf{X}$ を n で割る必要はない
(固有値が n 倍される)

固有ベクトルの変換公式

行列 $\mathbf{N} = \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{n \times n}$ の r 個の 0 でない固有値を大きなものから順に $\lambda_1, \dots, \lambda_r$ ($\mathbf{D} = \mathbf{X}\mathbf{X}^\top$ の固有値と同じ) とし, それぞれに対応する固有ベクトルを $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbb{R}^n$ とする. i 番目に大きい固有値に対応する d 次元の固有ベクトル \mathbf{u}_i は以下で求めることができる [8]:

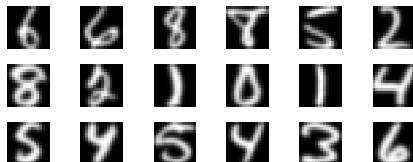
固有ベクトルの変換公式

$$\mathbf{u}_i = \frac{\pm \mathbf{X} \mathbf{v}_i}{\sqrt{\lambda_i}}$$

超高次元 (または無限次元) の固有ベクトルもこれで表現できる
⇒ 非線型部分空間法で利用できる

手書き数字パターン認識の実験

- 実験には <http://www.gaussianprocess.org/gpml/data/> で公開されている USPS 手書き数字データを使用
- 16×16 ピクセルの手書き数字パターンを 256 次元のベクトルにしたもので、未知・訓練パターン数はそれぞれ 4649
- 実験で使用するプログラムは `makedata.m` と `clafic.m`



データに関する注意と makedata.m の内容

- オリジナルの USPS データは未知パターンと訓練パターンの収集方法が異なるため、前述のウェブサイトで公開されているものは以下のような修正が施されている：
 - 未知パターンと訓練パターンをランダムに混ぜた後、同数の未知パターンと訓練パターン集合に分割
 - 画素値が $[-1, +1]$ となるようにスケーリング
- ただし、上記のデータは原点が 0 である保障がないことと、画像が横向きに保存されていること、ならびにクラスラベルが pair-wise な形式で保存されていることから、以下のような修正を makedata.m を用いて行う：
 - 画像を縦方向に変換
 - 画素値が $[0, +1]$ となるようにスケーリング
 - クラスラベルを 0 から 9 となるように修正
- 修正を施したデータは usps.mat という名前で保存

clafic.m の内容

clafic.m を実行すると以下の結果が表示される

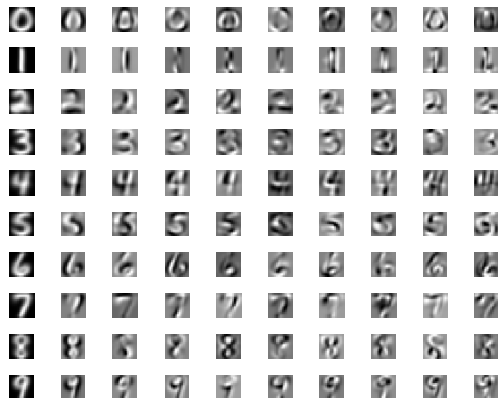
- Figure 1 : 各クラスで求められた固有値の大きい上位 10 個に対応した固有ベクトル
- Figure 2 : imgnum 番目の未知パターンの U_j による展開を画像化したもの ($r = 13$)
- 識別率と 1 パターンあたりの平均識別時間

なお, プログラム先頭の r や imgnum の値を変えると結果が変わるので, いろいろと値を変えてみよう

発表資料とプログラムで使われている記号の対応表

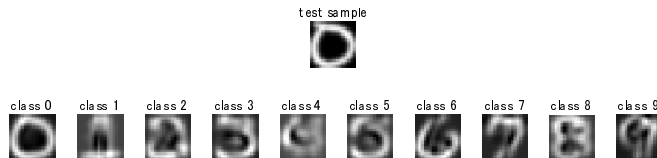
意味	発表資料	プログラム
クラス数	c	nclass
次元数	d	d
未知・訓練パターン数	n	ndata
第 i 訓練パターン	\mathbf{x}_i	tra _i (:,i)
第 i 訓練パターンのラベル		tra _i _label(i)
第 i 未知パターン	\mathbf{x}	test(:,i)
第 i 未知パターンのラベル		test_label(i)
クラス j の部分等長行列 \mathbf{U}_j	\mathbf{U}_j	C(j).U
ベクトル \mathbf{x} と \mathbf{y} の内積	$\mathbf{x}^\top \mathbf{y}$	x'*y
ベクトル \mathbf{x} のノルム	$\ \mathbf{x}\ = \sqrt{\mathbf{x}^\top \mathbf{x}}$	norm(x)

Figure 1 のキャプチャ画面



- 各クラスのパターンの変動が観察できる

Figure 2 のキャプチャ画面



- 未知パターンの各 U_j による展開 ($U_j U_j^T x$)
- 正しいクラスでは未知パターンを良く近似できる

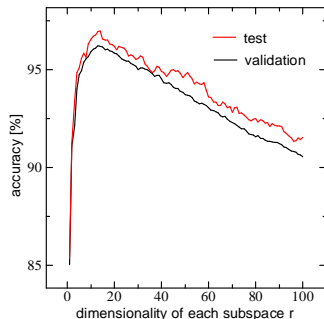
実験条件

CPU 1.86GHz , メモリ 2GB , 32bit の Windows , MATLAB (R14) を使用

識別法	
ベイズ決定則	単峰ガウス分布 (正則化あり)
マハラノビス距離	正則化あり
線型判別分析	正則化なし
最小距離法	各クラスの重心との距離で識別
最近傍決定則	最近傍パターンのラベルを出力
部分空間法 ($r = 13$)	CLAFIC
部分空間法 (累積寄与率)	CLAFIC
線型 SVM	one-against-all
非線型 SVM	one-against-all, RBF Kernel

部分空間の次元数の決め方

- 累積寄与率 $\bar{\lambda} = \sum_{i=1}^r \lambda_i / \sum_{j=1}^{\text{rank}(\mathbf{D})} \lambda_j$ でクラスごとに決定
- 分割学習法：訓練データを二つにわけ，一方を未知パターンとみなして r をクラス共通で決定 (下図参照)



実験結果

識別時間は一つの未知パターンを分類するのに必要な平均時間

識別法	識別率 (%)	識別時間 (s)	辞書サイズ (KB)
ベイズ決定則	96.5	0.003	5140.2
マハラノビス距離	97.1	0.002	5140
線型判別分析	92.0	0.002	532
最小距離法	85.8	3×10^{-5}	20
最近傍決定則	97.4	0.01	9298
部分空間法 ($r = 13$)	96.9	6×10^{-4}	260
部分空間法 ($\bar{\lambda} = 0.95$)	94.3	5×10^{-4}	231
線型 SVM	93.9	136.8	3656
非線型 SVM	98.0	222	5352

部分空間法の特長

- 特徴抽出と識別を同時に行うことができる
- クラスの追加・削除が容易
- 高速な識別が可能
- 辞書サイズを部分空間の次元数 r で調整できる
- パラメータは r のみ (累積寄与率, 交差検定で決める)
- 学習により識別率を向上できる
- 理論的な拡張が容易
 - 複合類似度法, 混合類似度法
 - 直交部分空間法, 学習部分空間法
 - カーネル非線形部分空間法
 - 相互部分空間法, 相互投影距離法, tangent distance
 - k -subspace clustering, (fuzzy) k -varieties clustering
 - 上記の組合せ

部分空間法の難点とその回避法

- 次元数が小さい場合 (例えば $d = 2$) に識別率が低下 [9]
- 複雑な決定境界を持つパターン分布では識別率が低下

- カーネル非線形部分空間法

- local subspace classifier (入力近傍に限定した投影距離法)

これらは計算時間, メモリ容量およびパラメータ数が大きくなる傾向がある

- クラス数が増加すると識別率が低下
 - カーネル非線形部分空間法

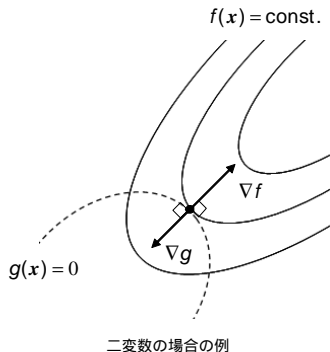
参考文献 I

- [1] 数学セミナー編集部, 教えて欲しい数学の疑問 1, 日本評論社, 1996.
- [2] 金谷健一, これなら分かる最適化数学 -基礎原理から計算手法まで -, 共立出版, 2005.
- [3] C.M. Bishop, Pattern recognition and machine learning, Springer, 2006.
元田 浩, 栗田多喜夫, 樋口知之, 松本祐治, 村田 昇 監訳, パターン認識と機械学習 - ベイズ理論による統計的予測, 上下巻, シュプリンガー・ジャパン, 2008.
- [4] S. Watanabe, P.F. Lambert, C.A. Kulikowski, J.L. Buxton, and R. Walker, "Evaluation and selection of variables in pattern recognition," in Computer and Information Sciences II, p. 91, Academic Press, 1967.
- [5] E. Oja, Subspace methods of pattern recognition, Research Studies Press, 1983.
小川英光, 佐藤 誠 訳, パターン認識と部分空間法, 産業図書, 1986.
- [6] S. Watanabe, Knowing and guessing : A quantitative study of inference and information, John Wiley & Sons, New York, 1969.
村上陽一郎, 丹治信春 訳, 知識と推測 : 科学的認識論, 上下巻, 東京図書, 1987.
- [7] 飯島泰蔵, 視覚情報の基礎理論 - パターン認識問題の源流 -, コロナ社, 1999.

参考文献 II

- [8] 金谷健一, これなら分かる応用数学教室 -最小二乗法からウェーブレットまで-, 共立出版, 2003.
- [9] 鷲沢嘉一, “正則化を用いた 2 次識別器,” MIRU 2007.

付録 1: ラグランジュ未定乗数法の直感的な理解

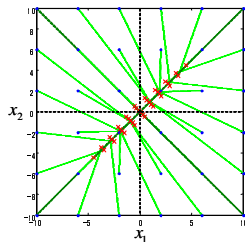


- 制約条件 $g(x) = 0$ と $f(x)$ の等高線の法線ベクトルが極値で平行

$$\nabla f = \lambda \nabla g$$

付録 2: 固有値 , 固有ベクトルの直感的な理解

$Au = \lambda u$ ($u \neq 0$) が成り立つとき u を A の固有ベクトル , λ を固有値とよぶ



例 : A を $A = \begin{pmatrix} 1/5 & 1/4 \\ 1/4 & 1/5 \end{pmatrix}$, x を任意の座標とする . 青い点を Ax で変換すると , 赤い \times へ移動する . 移動の軌跡を緑の線で表すと , 連続した二本の直線が現れる (この直線上にある点の位置ベクトルの方向が変化してないから) . この二本の直線の方角を与えるものが固有ベクトル , その直線上の移動量が固有値