

例題ではじめる部分空間法 - パターン認識へのいざない -

堀田政二 (東京農工大学)

日本ロボット学会 第 69 回ロボット工学セミナー

2012 年 5 月 22 日

目的と内容

目的

部分空間法を通して初心者でもパターン認識を体現できるようにすることと，部分空間法を研究に利用してもらうこと

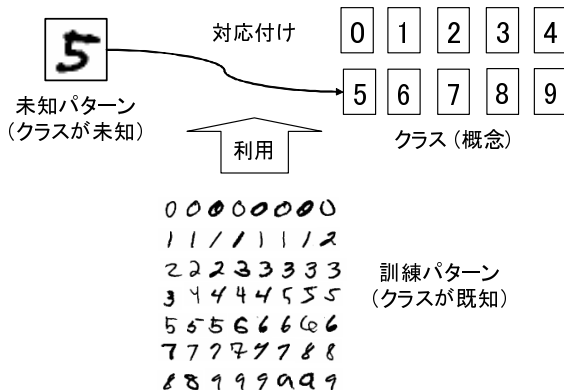
- 数学的準備 (ベクトルによる偏微分，ラグランジュ未定乗数法)
- 部分空間法，複合部分空間法，相互部分空間法
- MATLAB/Octave を用いた手書き数字・3次元物体認識の演習

本演習で使用するプログラムは下記ページから download 可能：

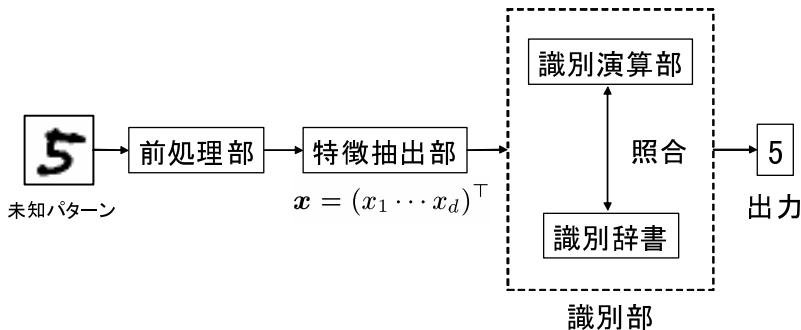
<http://www.tuat.ac.jp/~s-hotta/RSJ2012>

パターン認識とは

観測されたパターンを予め定められた複数の概念のうちの一つに対応させる (識別する) 処理 [1]

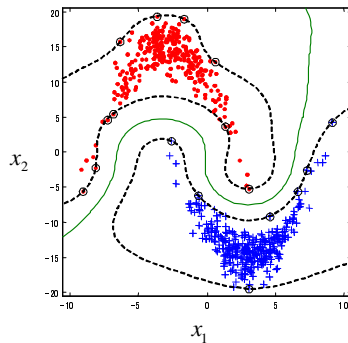
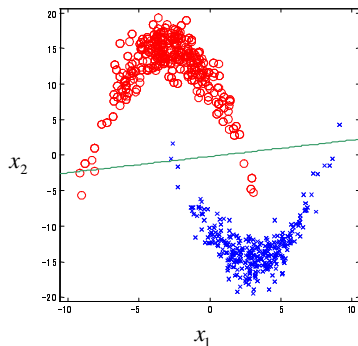


パターン認識の工学モデル [1]



- 本講演では，識別部に着目する

識別部の目的



- 未知パターンを精度良く識別できる (汎化能力の高い) 識別境界を引くこと

理想的な識別方法

ベイズ決定則 (Bayes decision rule)

$$\max_j P(\omega_j|\mathbf{x}) = P(\omega_k|\mathbf{x}) \Rightarrow \mathbf{x} \in \omega_k \quad \text{ここで } P(\omega_j|\mathbf{x}) = \frac{P(\omega_j)p(\mathbf{x}|\omega_j)}{\sum_j P(\omega_j)p(\mathbf{x}|\omega_j)}$$

- $P(\omega_j)$: パターンを観測する前のクラス ω_j の生起確率 (事前確率)
- $P(\omega_j|\mathbf{x})$: パターン \mathbf{x} を観測した後の ω_j の生起確率 (事後確率)
- $p(\mathbf{x}|\omega_j)$: ω_j における \mathbf{x} の分布 (確率密度関数)
- $p(\mathbf{x}|\omega_j)$ を精度良く推定することが難しい (大量データが得られれば別)

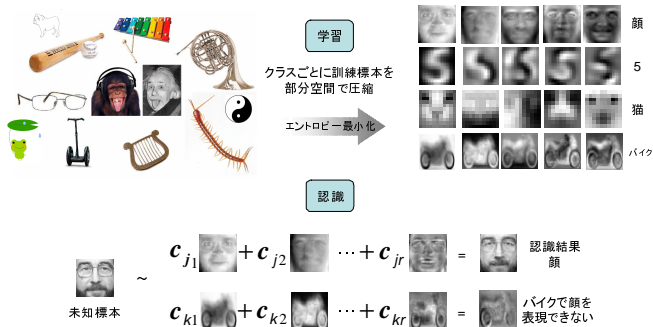
部分空間法に着目する理由

本講演では以下の理由から，部分空間法 [2, 3, 4] について取り上げる：

- 直感的に理解しやすい (初心者に向いている)
- 簡単に実装できる
- 様々なパターン変動を少ない辞書サイズで表現できるため高速・高精度に識別可能 (実用的)
- 線型代数，確率，最適化等の復習が効率的にできるので，教育的な観点からも初心者には良い題材

部分空間法とは

- 学習: 驚きの小さい低次元の空間 (部分空間) をクラスごとに求める
- 認識: 未知パターンを最も驚き (誤差) の小さいクラスへ分類する



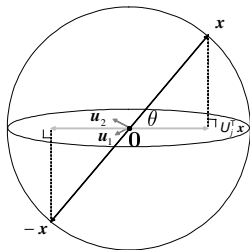
エントロピー最小化と幾何学的解釈

- 学習: クラス毎にエントロピーが最小となるような部分空間軸を求める

$$\min_{\mathbf{u}_i} - \sum_{i=1}^d p(\mathbf{u}_i) \log p(\mathbf{u}_i)$$

- 認識: 次元削減 + 内積演算

$\sum_{i=1}^r w_i (\mathbf{u}_i^\top \mathbf{x})^2$ が最大となるクラスへ \mathbf{x} を分類



部分空間法の開発者



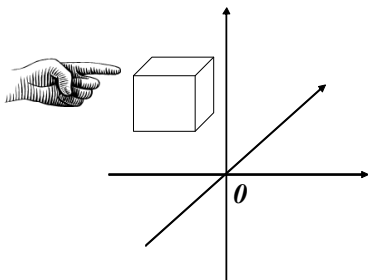
- 飯島泰蔵 (1925-)
- 東京工業大学電気工学科卒
- 視覚パターン認識に関する統一的基礎理論の構築
- 超高性能 OCR “ASPET/71” の研究開発
- 電総研 & 東芝による国家プロジェクト (通産省)
- 北陸先端大学院大学名誉教授等を歴任



- 渡辺慧 (1910-1993)
- 東京帝国大学理学部物理学科卒
- ド・ブロイのもとでドクトル・デタ (国家博士) 取得
- エントロピー概念の情報理論への応用 (シャノンより先)
- “ルネサンス人の最後の一人” とも称される
- ハワイ大学名誉教授, 国際時間学会会長等を歴任

線型部分空間とはなんですか？ 1/3

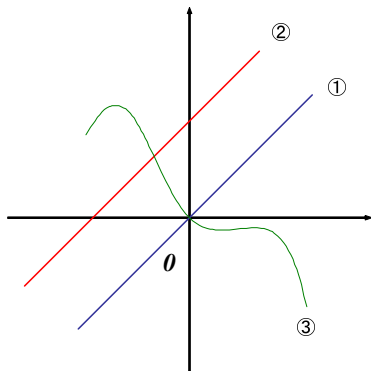
これは部分空間ですか？ [5]



- 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)

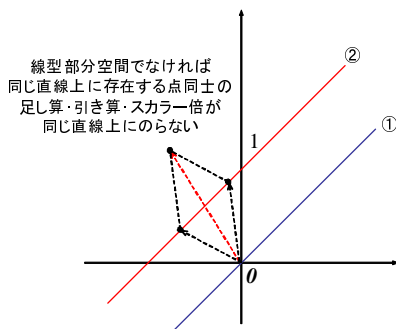
線型部分空間とはなんですか？ 2/3

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:

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

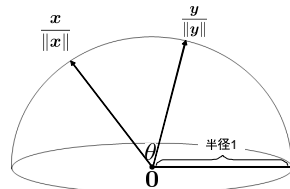
- \mathbf{x} のノルム: $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\top \mathbf{x}}$

- 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:

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

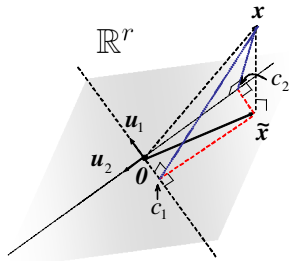


$n + 1$ 次元空間における n 次元単位球面

($n \geq 3$ 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 $f = \mathbf{y}^\top \mathbf{A} \mathbf{x}$ を \mathbf{x} で偏微分すると $\frac{\partial}{\partial \mathbf{x}} f = \mathbf{A}^\top \mathbf{y}$

$\arg, \max, \text{subject to}$ 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 [6]．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． $\mathbf{u}^\top \mathbf{u} = 1$ の制約条件のもと， $\mathbf{u}^\top A \mathbf{u}$ の最大値を求めよ

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

- ラグランジュ乗数 λ を用いてラグランジュ関数を導入

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

- 制約条件のもとで関数が極値をとる点は次式を満たす：

$$\partial L / \partial \mathbf{u} = \mathbf{0}, \quad \partial L / \partial \lambda = 0$$

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

$$A\mathbf{u} = \lambda\mathbf{u}$$

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 \mathbf{u}^\top を掛けてみよう． $\mathbf{u}^\top \mathbf{u} = 1$ 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 (subspace method)

クラスらしさを部分空間で表現する方法 [2, 4] . 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 [7]
- 視覚情報に関する理論から [8]
- 統計学からみて自然な発想 (縮退ガウス分布)

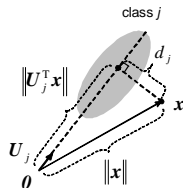
部分空間法の識別則

- 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：

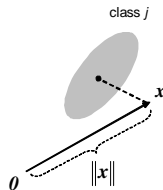
部分空間法の識別則

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

部分空間法における識別則の直感的な理解



部分空間法



最小距離法

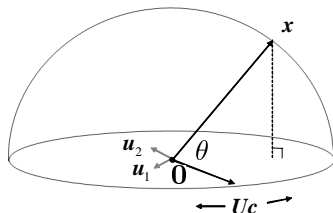
- 部分空間法における原点は各クラス共通
- 未知パターン x を部分空間を使って最も良く近似できるクラスへ分類する
 $\Rightarrow d_j^2 = \|x\|^2 - \|U_j^\top x\|^2$ が最小のクラスへ分類する．ただし， $\|x\|^2$ が全クラス共通なので，結局 $\|U_j^\top x\|^2$ が最大となるクラスへ分類すれば良い

$\cos \theta$ 最大化基準から見た部分空間法 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部分空間法 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

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

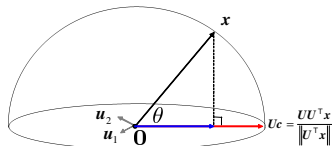
となる. これは \mathbf{x} を \mathbf{U} の張る部分空間へ正射影したパターンをノルム 1 に正規化したものに他ならない

$\cos \theta$ 最大化基準から見た部分空間法 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

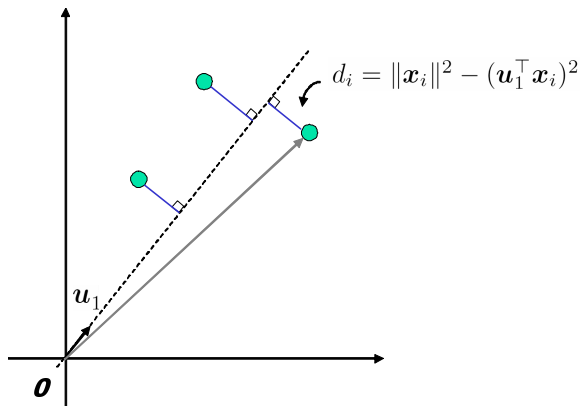
$$\cos \theta = \mathbf{x}^\top \mathbf{U} \mathbf{c} = \frac{\mathbf{x}^\top \mathbf{U} \mathbf{U}^\top \mathbf{x}}{\|\mathbf{U}^\top \mathbf{x}\|} = \frac{\|\mathbf{U}^\top \mathbf{x}\|^2}{\|\mathbf{U}^\top \mathbf{x}\|} = \|\mathbf{U}^\top \mathbf{x}\| = \lambda$$

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



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と元のパターンとの二乗誤差の総和が最小となるように選ぶ



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，二乗誤差 $d_i = \|x_i\|^2 - (u_1^\top x_i)^2$ の総和が最小となる u_1 を求めることを考えるが，これは $\|x_i\|^2$ が u_1 のとり方に依らないため $(u_1^\top x_i)^2$ が最大となる u_1 を求めることと等価である：

$$\begin{aligned} \max_{u_1} \quad & \sum_{i=1}^n (u_1^\top x_i)^2 = u_1^\top \left(\sum_{i=1}^n x_i x_i^\top \right) u_1 \\ \text{s.t.} \quad & u_1^\top u_1 - 1 = 0 \end{aligned}$$

- ラグランジュ未定乗数法から，解は行列 $D = \sum_{i=1}^n x_i x_i^\top$ 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，二乗誤差の総和が r 番目に小さい正規直交ベクトル u_r は 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 $U_j = (u_1 | \dots | u_r)$ を求め，それらを識別に用いる

次元数 d が大きい場合

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 [9]:

固有ベクトルの変換公式

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

次元が高い場合でも, これにより固有ベクトルを求めることが出来る (プログラムの `EVD.m` がこれに対応)

部分空間法はそもそも統計的手法ではない

- 飯島先生: テンプレートマッチングの拡張
- 渡辺先生: エントロピー最小化の枠組みから導出

部分空間法を統計的手法として導く [10, 11]

- ベイズ決定則との関連性が不明瞭

$$P(\omega_j | \mathbf{x}) = \frac{P(\omega_j)p(\mathbf{x}|\omega_j)}{\sum_j P(\omega_j)p(\mathbf{x}|\omega_j)}$$

これが分かれば …

- 部分空間法を深く理解可能に
- 統計的な側面からの拡張が可能に

ω_j における多次元ガウス分布

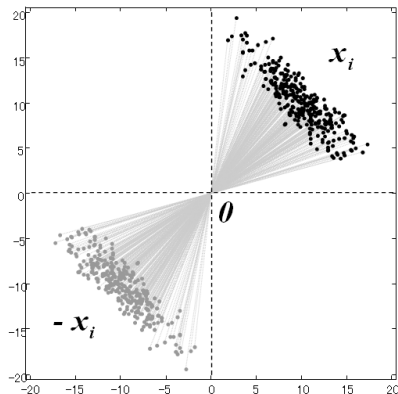
$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}_j|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j) \right\}$$

- $\boldsymbol{\mu}_j$: $d \times 1$ の母平均
- $\boldsymbol{\Sigma}_j$: $d \times d$ の母分散共分散行列

これらを与えられた訓練標本だけで最尤推定により推定すると部分空間法は導出できない



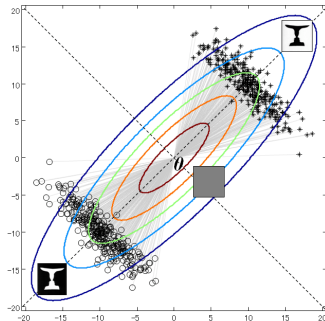
\mathbf{x}_i の零元に関する鏡像 $-\mathbf{x}_i$ を含めて最尤推定



鏡像を含めた訓練標本に (クラス毎に) ガウス分布を当てはめる

最尤推定の結果

$\hat{\mu}_j = 0, \hat{\Sigma}_j = \mathbf{R}_j$, ただし $\mathbf{R}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbf{x}_i \mathbf{x}_i^\top$: 自己相関行列



$$\mathcal{N}(\mathbf{x} | \hat{\mu}_j, \hat{\Sigma}_j) = \frac{1}{(2\pi)^{d/2} |\mathbf{R}_j|^{1/2}} \exp \left(-\frac{1}{2} \mathbf{x}^\top \mathbf{R}_j^{-1} \mathbf{x} \right)$$

以下をモデルとしたベイズ決定則を考える

$$\mathcal{N}(\mathbf{x}|\hat{\boldsymbol{\mu}}_j, \hat{\boldsymbol{\Sigma}}_j) = \frac{1}{(2\pi)^{d/2}|\mathbf{R}_j|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^\top \mathbf{R}_j^{-1}\mathbf{x}\right)$$



対数を取り事前確率が等しいと仮定した場合の識別関数

$$g_j(\mathbf{x}) = -\|\boldsymbol{\Lambda}_j^{-\frac{1}{2}}\mathbf{U}_j^\top \mathbf{x}\|^2 - \sum_{i=1}^d \ln \lambda_{ji}$$

ただし $\mathbf{R}_j = \mathbf{U}_j \boldsymbol{\Lambda}_j \mathbf{U}_j^\top$ で固有値は降順にソート

自己相関行列の固有値について考える

$$g_j(\boldsymbol{x}) = -\|\boldsymbol{\Lambda}_j^{-\frac{1}{2}} \mathbf{U}_j^\top \boldsymbol{x}\|^2 - \sum_{i=1}^d \ln \lambda_{ji}$$

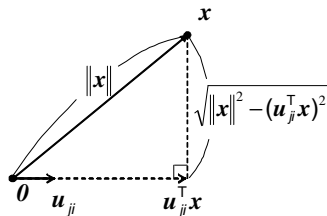
- 固有値は対応する固有ベクトルに \boldsymbol{x}_i と $-\boldsymbol{x}_i$ を正射影したものの分散 (これまでに指摘されていない)
- 普通の標本分散共分散行列の固有値よりも大きくなりがち
- 第二項は全クラスで固有値が等しければ省略可能



固有値の値は信頼しないで大小関係のみが重要であると仮定

固有値の代わりに重み $w_1 \geq w_2 \geq \dots \geq w_d > 0$ を導入

$$g_j(\mathbf{x}) = -\|\mathbf{W}^{-\frac{1}{2}} \mathbf{U}_j^\top \mathbf{x}\|^2 = -\sum_{i=1}^d \frac{1}{w_i} (\mathbf{u}_{ji}^\top \mathbf{x})^2$$



重みの小さな成分を重視 \Rightarrow 醜い家鴨の子の定理から望ましくない

$1/w_1 \leq 1/w_2 \leq \dots \leq 1/w_d > 0 \Rightarrow w_1 \geq w_2 \geq \dots \geq w_d > 0$ と
(特徴の持つ価値を) 変更してマイナスを掛け, w_i の大きい次元の
みで分類すれば …

部分空間法

識別関数: $S_j(\mathbf{x}) = \sum_{i=1}^r w_i (\mathbf{u}_{ji}^\top \mathbf{x})^2 = \|\mathbf{W}^{\frac{1}{2}} \mathbf{U}_j^\top \mathbf{x}\|^2 = \mathbf{w}^\top \mathbf{c}_j$

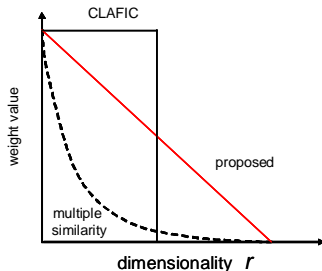
識別則: $S_j(\mathbf{x})$ が最大となるクラスへ \mathbf{x} を分類

ここまでの議論で明らかになったことや予想できること

- 重みが小さい成分を無視できる
- 重みは非増加であれば何でも良い
- 零元 (全クラス共通の原点) 付近は誤分類が起きやすい

重みの決め方

- CLAFIC (渡辺先生): $w_i = 1$ ($i = 1, \dots, r$)
- 複合類似度 (飯島先生): $w_i = \lambda_{ji} / \lambda_{j1}$
- 本講演: $w_i = r - i + 1$ ($i = 1, \dots, r$)



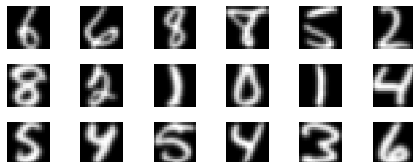
なぜ直線でよいのか？

$$\begin{aligned} S_j(\mathbf{x}) &= \sum_{i=1}^r (r - i + 1) (\mathbf{u}_{ji}^\top \mathbf{x})^2 \\ &= r (\mathbf{u}_{j1}^\top \mathbf{x})^2 + (r - 1) (\mathbf{u}_{j2}^\top \mathbf{x})^2 + \cdots + (\mathbf{u}_{jr}^\top \mathbf{x})^2 \\ &= \sum_{i=1}^r (\mathbf{u}_{ji}^\top \mathbf{x})^2 + \sum_{i=1}^{r-1} (\mathbf{u}_{ji}^\top \mathbf{x})^2 + \cdots + (\mathbf{u}_{j1}^\top \mathbf{x})^2 \end{aligned}$$

- 重みの大きい成分の強調と重みの小さい成分の抑制
- 部分空間の次元ごとの類似度の総和
- テンプレートベースのアンサンブル学習

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
- 演習で使用するプログラムは `makedata.m` と `WSC.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

WSC.m の内容

WSC.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 ($r = 13$)
- 認識率と 1 パターンあたりの平均認識時間

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

認識率の求め方

混同行列: クラスラベルが i である未知パターンをクラス j に分類した頻度を ij 要素に持つ行列

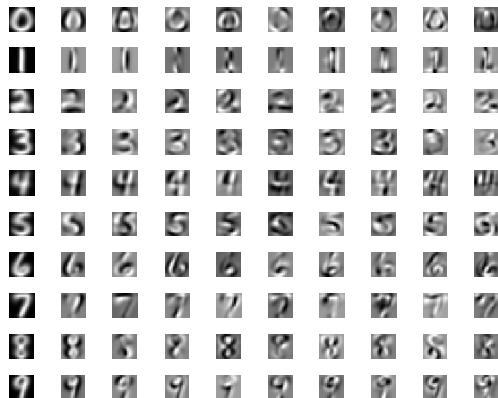
class	0	1	2	3	4	5	6	7	8	9	total
0	680	0	2	1	5	7	72	1	17	1	786
1	0	642	0	0	1	0	2	0	2	0	647
2	10	1	377	15	11	6	6	9	19	0	454
3	5	0	2	369	2	19	0	4	15	2	418
4	4	15	7	0	356	1	9	1	8	42	443
5	10	0	3	18	10	286	8	0	15	5	355
6	32	7	9	0	3	3	355	0	5	0	414
7	2	3	1	0	7	0	0	348	2	39	402
8	4	9	5	15	5	7	0	1	273	12	331
9	0	10	0	1	45	0	0	25	2	316	399
total	747	687	406	419	445	329	453	389	358	417	4649

本講演では, 対角要素の総和を未知パターンの総数で割って百分率で表したものを認識率と呼ぶ

発表資料とプログラムで使われている記号の対応表

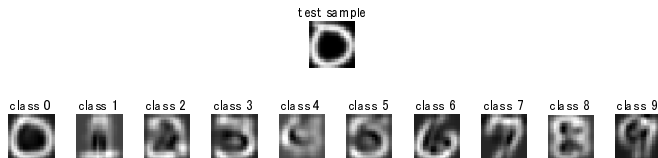
意味	発表資料	プログラム
クラス数	C	nclass
次元数	d	d
未知・訓練パターン数	n	ndata
第 i 訓練パターン	\mathbf{x}_i	trai(:,ii)
第 i 訓練パターンのラベル		trai_label(ii)
第 i 未知パターン	\mathbf{x}	test(:,ii)
第 i 未知パターンのラベル		test_label(ii)
クラス j の部分等長行列 \mathbf{U}_j	\mathbf{U}_j	C(j).U
ベクトル \mathbf{x} と \mathbf{y} の内積	$\mathbf{x}^\top \mathbf{y}$	$\mathbf{x}'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

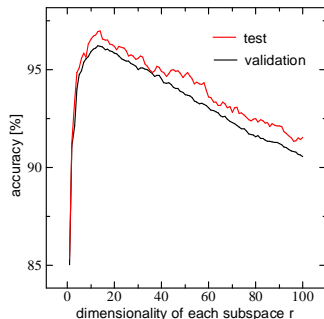
Figure 2 のキャプチャ画面



- 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

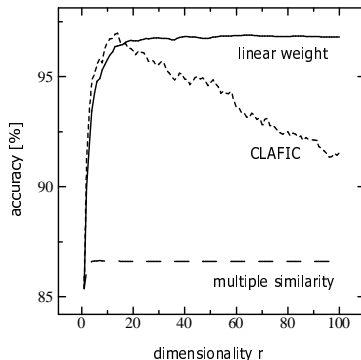
部分空間の次元数の決め方

- 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)



重みの違いによる認識率の推移

- 直線重みは次元数 r の影響を受けにくい



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

実験結果

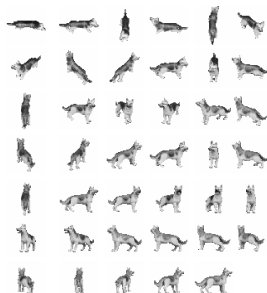
認識時間は一つの未知パターンを分類するのに必要な平均時間

識別法	認識率 (%)	認識時間 (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

同一クラスから複数の未知パターンが観測できる場合



単体のパターンを観測しても
何であるかはわからない



同じクラスから由来する複数のパターン



犬

- 複合部分空間法 (CSM): ベイズ決定則の拡張である複合決定問題から部分空間法を導出
- 相互部分空間法 (MSM): 部分空間同士の角度に基づく識別則を導出

部分空間法の拡張

- ベイズ決定則から部分空間法を導出できることから
- ベイズ決定則の拡張である複合決定問題からも部分空間法を導出でき
- その結果，複数の特徴量・複数の未知サンプルの分類が可能となる

Compound Subspace Method (CSM) [11]

$$P(\omega|\mathbf{X}) = \frac{P(\omega)p(\mathbf{X}|\omega)}{\sum_{\omega} P(\omega)p(\mathbf{X}|\omega)} \sim \sum_{k=1}^n \sum_{f=1}^T \sum_{i=1}^r {}_f w_i ({}_f \mathbf{u}_i^{\top} {}_f \mathbf{x}_k)^2$$

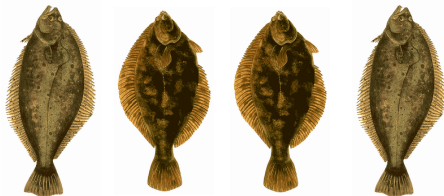
- 画像認識以外の問題にも適用可能な汎用的手法
- Multiple Kernel Learning と同程度の認識率で高速に学習・認識が可能

複合決定問題 (compound Bayesian decision problem) [12]

$$P(\omega|\mathbf{X}) = \frac{P(\omega)p(\mathbf{X}|\omega)}{\sum_{\omega} P(\omega)p(\mathbf{X}|\omega)}$$

- 連続して観測された n 個の標本: $\mathbf{X} = (x_1|x_2|\cdots|x_n)$
- n 個の状態 (context): $\omega = (\omega(1), \dots, \omega(n))^{\top}$

各 x_i に対応する $\omega(i)$ を統計的独立性を仮定しないで決定する問題



ひらめ

カレイ

カレイ

ひらめ

複合決定問題 (compound Bayesian decision problem)

$$P(\omega|\mathbf{X}) = \frac{P(\omega)p(\mathbf{X}|\omega)}{\sum_{\omega} P(\omega)p(\mathbf{X}|\omega)}$$

- ω の組み合わせが c^n もある
- $p(\mathbf{X}|\omega)$ の推定も困難



下記の仮定と鏡像を含むガウス分布を用いて三種類の CSM を導出

- x_i が全て同じクラスに由来すると仮定
- 一つの未知標本から T 個の特徴量を抽出した場合
- 上記の組み合わせ

n 個の x_1, \dots, x_n が同じクラスに由来し, i.i.d. に従うと仮定

$$P(\omega_j | \mathcal{X}) = \frac{P(\omega_j) \prod_{i=1}^n p(\mathbf{x}_i | \omega_j)}{\sum_{j=1}^c P(\omega_j) \prod_{i=1}^n p(\mathbf{x}_i | \omega_j)}$$

クラス毎に鏡像を含めた単峰ガウス分布を用いれば

複数の同じクラスに由来する未知標本を分類するための部分空間法

$$S_j(\mathcal{X}) = \sum_{i=1}^n \sum_{l=1}^r w_l (\mathbf{u}_{jl}^\top \mathbf{x}_i)^2 = \sum_{i=1}^n \|\mathbf{W}^{\frac{1}{2}} \mathbf{U}_j^\top \mathbf{x}_i\|^2 = \sum_{i=1}^n S_j(\mathbf{x}_i)$$

x から T 種類の特徴量が抽出された時

$$P(\omega_j|\mathcal{F}) = \frac{P(\omega_j) \prod_{f=1}^T p_f({}_f\mathbf{x}|\omega_j)}{\sum_{j=1}^c P(\omega_j) \prod_{f=1}^T p_f({}_f\mathbf{x}|\omega_j)}$$

特徴量・クラスごとに鏡像を含めた単峰ガウス分布を用いれば

複数の特徴量を用いて分類するための部分空間法 (f 毎に正規化必要)

$$F_j(\mathcal{F}) \stackrel{\text{def}}{=} \sum_{f=1}^T \sum_{l=1}^{r_f} {}_f w_l ({}_f \mathbf{u}_{jl}^\top {}_f \mathbf{x})^2 = \sum_{f=1}^T \|{}_f \mathbf{W}^{\frac{1}{2}} {}_f \mathbf{U}_j^\top {}_f \mathbf{x}\|^2 = \sum_{f=1}^T {}_f S_j({}_f \mathbf{x})$$

n 個の未知標本 \mathcal{X} から T 種類の特徴量が抽出された時

$$P(\omega_j | \bar{\mathcal{X}}) = \frac{P(\omega_j) \prod_{f=1}^T \prod_{i=1}^n p_f({}_f\mathbf{x}_i | \omega_j)}{\sum_{j=1}^c P(\omega_j) \prod_{f=1}^T \prod_{i=1}^n p_f({}_f\mathbf{x}_i | \omega_j)}$$

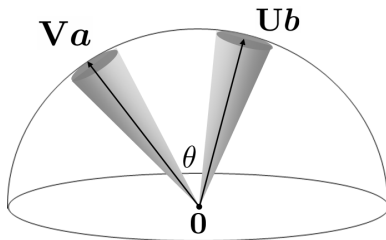
特徴量・クラスごとに鏡像を含めた単峰ガウス分布を用いれば

複数の未知標本を複数の特徴量を用いて分類するための部分空間法

$$C_j(\bar{\mathcal{X}}) \stackrel{\text{def}}{=} \sum_{f=1}^T \sum_{i=1}^n {}_fS_j({}_f\mathbf{x}_i)$$

相互部分空間法 (mutual subspace method, MSM) への拡張

複数の同一クラスに由来する未知パターンが与えられた場合，未知パターン集合を部分空間で表現することで，相互部分空間法 [13] に容易に拡張できる．



相互部分空間法では，部分空間同士の成す角度が最小 ($\cos \theta$ が最大) となるクラスへ未知パターン集合を分類する

問題の定式化

便宜上，以後はクラスを表す添字は省略．部分空間は自己相関行列の固有値分解から得られる固有ベクトルによって張られているとする

- $\mathbf{V} \in \mathbb{R}^{d \times r_d}$: あるクラスの訓練パターン集合から求めた $r_d < d$ 次元の部分空間を張る固有ベクトル
- $\mathbf{U} \in \mathbb{R}^{d \times r_i}$: 未知パターン集合から求めた $r_i < d$ 次元の部分空間を張る固有ベクトル
- 部分空間上のパターン $\mathbf{V}\mathbf{b}$, $\mathbf{U}\mathbf{c}$ について以下の問題を考える

相互部分空間法への拡張

$$\begin{aligned} \max_{\mathbf{b}, \mathbf{c}} (\mathbf{V}\mathbf{b})^\top (\mathbf{U}\mathbf{c}) &= \cos \theta \\ \text{s.t. } \mathbf{b}^\top \mathbf{b} - 1 &= 0, \mathbf{c}^\top \mathbf{c} - 1 = 0 \end{aligned} \quad (3)$$

部分空間法の場合と同様に，ラグランジュ未定乗数法で解いてみよう (59)

相互部分空間法の導出 1/2

- 式 (3) のラグランジュ関数を導入

$$L = (\mathbf{V}\mathbf{b})^\top (\mathbf{U}\mathbf{c}) - \frac{\lambda}{2}(\mathbf{b}^\top \mathbf{b} - 1) - \frac{\mu}{2}(\mathbf{c}^\top \mathbf{c} - 1)$$

- $\partial L / \partial \mathbf{b} = \mathbf{0}$, $\partial L / \partial \mathbf{c} = \mathbf{0}$ より, 解は以下を満たす:

$$\mathbf{V}^\top \mathbf{U}\mathbf{c} = \lambda \mathbf{b} \quad (4)$$

$$\mathbf{U}^\top \mathbf{V}\mathbf{b} = \mu \mathbf{c} \quad (5)$$

- 式 (4) から \mathbf{c} , あるいは式 (5) から \mathbf{b} を消去すると

$$\mathbf{V}^\top \mathbf{U}\mathbf{U}^\top \mathbf{V}\mathbf{b} = \lambda \mu \mathbf{b}, \quad \mathbf{U}^\top \mathbf{V}\mathbf{V}^\top \mathbf{U}\mathbf{c} = \lambda \mu \mathbf{c} \quad (6)$$

となり, 解は固有値問題を解くことで得られることがわかる

相互部分空間法の導出 2/2

- ところで式 (4) の左から b^\top , 式 (5) の左から c^\top を掛けると

$$b^\top V^\top U c = b^\top \lambda b = \lambda \quad (7)$$

$$c^\top U^\top V b = c^\top \mu c = \mu \quad (8)$$

となり, $\lambda = \mu$ であることがわかる

- したがって式 (6) から, $\lambda (= \mu)$ は $V^\top U U^\top V$ の最大固有値
(= $U^\top V V^\top U$ の最大固有値) の平方根であり, これが二つの部分空間の成す最大の $\cos \theta$ である
- 計算量の観点から $V^\top U U^\top V$ と $U^\top V V^\top U$ のうち, サイズの小さい行列の固有値を求めればよい

使用するデータ : ETH-80 dataset [14]



- 本来は 8 クラス問題 : 1 クラスあたり 10 物体 , 1 物体あたり 41 視点画像 (128×128 ピクセル)
- 各画像には背景除去のためのマスク画像が用意されている
- 今回は犬 , 牛 , 馬の物体を独立のクラスとみなす (30 クラス問題)

dog1 の例

dog1-000-000.png dog1-022-000.png dog1-022-090.png dog1-022-180.png dog1-022-270.png dog1-035-045.png



dog1-035-135.png dog1-035-225.png dog1-035-315.png dog1-045-000.png dog1-045-090.png dog1-045-180.png



dog1-045-270.png dog1-066-027.png dog1-066-063.png dog1-066-117.png dog1-066-153.png dog1-066-207.png



dog1-066-243.png dog1-066-297.png dog1-066-333.png dog1-068-000.png dog1-068-090.png dog1-068-180.png



dog1-068-270.png dog1-090-000.png dog1-090-022.png dog1-090-045.png dog1-090-068.png dog1-090-090.png



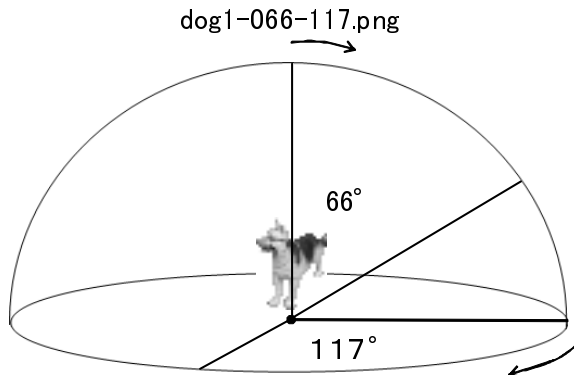
dog1-090-112.png dog1-090-135.png dog1-090-158.png dog1-090-180.png dog1-090-202.png dog1-090-225.png



dog1-090-248.png dog1-090-270.png dog1-090-292.png dog1-090-315.png dog1-090-338.png



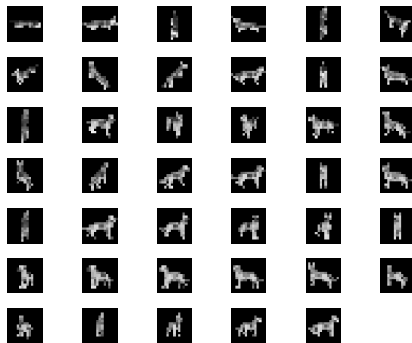
ファイル名の意味



ファイル名は物体名と番号 + 頂点からの角度 + 右水平方向からの角度

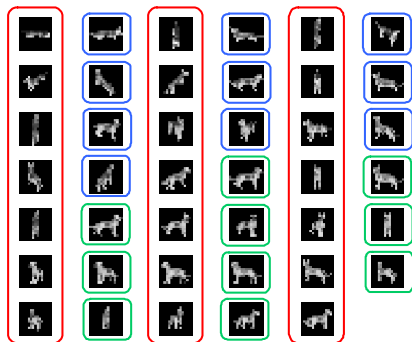
画像処理

make_data.m で施される画像処理



- 全ての画像を $Y = 0.2989R + 0.587G + 0.114B$ によりモノクロ化
- マスクを使用して背景を除去
- 最近傍法を用いて画像を 15×15 ピクセルに縮小

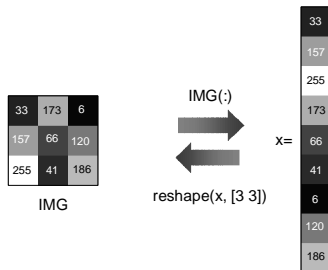
未知・訓練データの作成方法



- 奇数番目の 21 枚の画像を訓練パターン集合とする (赤枠)
- 偶数番目の画像のうち、連続する 10 枚を一つの未知パターン集合とする (青枠)．これを 22 番目の画像を先頭に持つ集合 (緑枠) まですべて作成 (1 物体あたり 11 種類の未知パターン集合)

画像のベクトル化

画像をベクトル化してファイル (doubutsu.mat) に保存



- `IMG(:)`: 行列をベクトル化
- `reshape`: ベクトルを行列化

変数名	意味
ri	未知データの部分空間の次元数
rd	訓練データの部分空間の次元数
nclass	クラス数
ntraí	訓練データ数
ntest	未知データ数
traí(ii).X	第 ii 訓練データ (225×21 の行列)
test(ii).X	第 ii 未知データ (225×21 の行列)
EVD	ソート済み固有値と固有ベクトルを出力する関数
traí(ii).U	第 ii 訓練データの部分空間を張る固有ベクトル
test(ii).U	第 ii 未知データの部分空間を張る固有ベクトル
traí(ii).label	第 ii 訓練データのクラスラベル
test(ii).label	第 ii 未知データのクラスラベル
CONF	混同行列

実験結果

CPU 2.8GHz , メモリ 3.5GB , 32bit の Windows , MATLAB (R14) を使用

認識時間は一つの未知パターン集合を分類するのに必要な平均時間

- WSC: 未知パターンを一つずつ部分空間法で分類した場合
- MSM: 未知パターン集合を相互部分空間法で分類した場合
- CSM: 未知パターン集合を複合部分空間法で分類した場合

識別法	認識率 (%)	認識時間 (s)
WSC ($r_d = 20$)	42.4	0.0007
MSM ($r_i = 5, r_d = 6$)	71.8	0.002
CSM ($r_d = 20$)	80.9	0.002

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)
- 学習により認識率を向上できる
- 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 $d = 2$) に認識率が低下
- 複雑な決定境界を持つパターン分布では認識率が低下
 - 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認識率が低下
 - カーネル非線形部分空間法

部分空間法研究会のホームページ

<http://www.cvlab.cs.tsukuba.ac.jp/~subspace/>

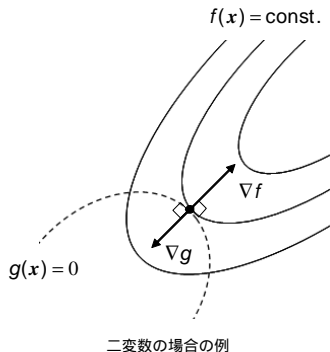
参考文献 I

- [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, Aug. 1998.
- [2] S. Watanabe, P.F. Lambert, C.A. Kulikowski, J.L. Buxton, and R. Walker, “Evaluation and selection of variables in pattern recognition,” Comp. & Info. Sciences, vol. 2, (Julius Tou, ed.). New York: Academic Press, pp. 91–122, 1967.
- [3] T. Iijima, H. Genchi, and K. Mori, “A theory of character recognition by pattern matching method,” Proc. of 1st Int’l J. Conf. on Pattern Recognition, pp. 50–56, 1973.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.

参考文献 II

- [8] 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.
- [9] 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.
- [10] 堀田政二, “線型部分空間識別器と最大事後確率則の関係,” 信学技報, PRMU2010, vol. 110, no. 296, pp. 13–18, Nov. 2010 .
- [11] 堀田政二, “複合決定問題から導かれる複合部分空間法,” 信学技報, PRMU2010, vol. 110, no. 330, pp. 45–50, Dec. 2010.
- [12] J.F. Hannan and H. Robbins, “Asumptotic solutions of the compound decision problem for two completely specified distributions,” Annals of Mathematical Statistics, vol. 26, no. 1, pp. 37–51, 1955.
- [13] 前田賢一, 渡辺貞一, “局所的構造を導入したパターン・マッチング法”, 信学論 (D), vol. J68-D, no. 3, pp. 345–352, 1985.
- [14] B. Leibe and B. Schiele, “Analyzing appearance and contour based methods for object categorization,” Proc. of CVPR, pp. 409–415, 2003.

付録 1: ラグランジュ未定乗数法の直感的な理解

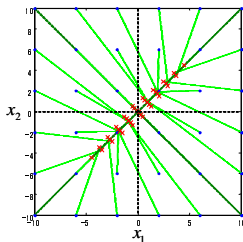


- 制約条件 $g(x) = 0$ と $f(x)$ の等高線の法線ベクトルが極値で平行

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

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