はじめてのパターン認識
第9 章　部分空間法

山木翔馬

July 7, 2015

次元の縮約

効率的な学習のためにはデータの次元が少ないほうが良い．
次元を縮約する手法として

- 主成分分析
  共分散行列の固有値問題を解き，大きな固有値を持つ固有ベクトルで部分空間を構成．
- 部分空間法
  クラスごとに主成分分析を行い，それぞれ部分空間を構成してパターン認識を行う．

部分空間法はクラス間の識別関数を構成して識別を行うのではなく，入力データと各クラスの類似度を直接に評価して識別する．

部分空間の定義

d 次元ベクトル空間 $\mathbb{V}$ の部分空間

$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r (r \leq d)$ を $\mathbb{V}$ のベクトルとする．$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ の 1 次結合の集合

$$W = \{a_1 \boldsymbol{x_1} + \cdots + a_r \boldsymbol{x}_r | a_i \in \mathbb{R}, i = 1, \ldots, r\}$$

は $\mathbb{V}$ の部分空間となる．

$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ が 1 次独立であるとき，$\boldsymbol{W}$ は $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ で張られる r 次元の部分空間である．
部分空間の定義から，$\boldsymbol{W}$ が $\mathbb{V}$ の部分空間であるための必要十分条件は，

1. $W \neq \phi$
2. $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{W} \Rightarrow \boldsymbol{x} + \boldsymbol{y} \in \boldsymbol{W}$
3. $\boldsymbol{x} \in \boldsymbol{W}, \lambda \in \mathbb{R} \Rightarrow \lambda \boldsymbol{x} \in \boldsymbol{W}$

また，ベクトル空間 $\mathbb{V}$ は部分空間 W とそれに直交する部分空間 $W^{\perp}$ に分解でき，$\mathbb{V} = W \cup W^{\perp}, W \cap W^{\perp} = \phi$ が成り立つ．

正規直交化

1 次独立であるが直交していないベクトルは，内積計算を用いて分解するためには直交座標系へ変換する必要がある．その手法の一つのグラム-シュミットの正規直交化では，次のようにして部分空間の正規直交基底を求めることができる．

グラム-シュミットの正規直交化

1. $\boldsymbol{n}_1 = \boldsymbol{x}_1/\|\boldsymbol{x}_1\|$ に設定
2. $i > 1$ について，順次以下を繰り返す

$$\tilde{\boldsymbol{n}}_i = \boldsymbol{x}_i - \sum_{j=1}^{i-1}(\boldsymbol{n}_j^T\boldsymbol{x}_i)\boldsymbol{n}_j, \qquad \boldsymbol{n}_i = \frac{\tilde{\boldsymbol{n}}_i}{\|\tilde{\boldsymbol{n}}_i\|}$$

主成分分析

主成分分析は，学習データ $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{id})^T (i = 1, \ldots, N)$ の分散が最大になる方向への線形変換を求める手法.
分散が最大になる方向へ射影することと，データの無相関化は同じこと.
(ただし無相関化では次元の縮約は考慮しない)

主成分分析の流れ

1. 共分散行列 Σ を求める.
2. 分散を最大にする射影ベクトルを求める.

共分散行列

N 個のデータからなるデータ行列を $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)^T$，それらから平均ベクトル $\bar{\boldsymbol{x}} = (\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_d)^T$ を引き算したデータ行列を $\bar{\boldsymbol{X}} = (\boldsymbol{x}_1 - \bar{\boldsymbol{x}}, \ldots, \boldsymbol{x}_N - \bar{\boldsymbol{x}})^T$ とすれば，共分散行列 Σ は

$$\Sigma = Var\{\bar{\boldsymbol{X}}\} = \frac{1}{N}\bar{\boldsymbol{X}}^T\bar{\boldsymbol{X}}$$

で定義される．
N 個のデータ $\boldsymbol{x}_i$ を係数ベクトル $\boldsymbol{a}_j = (a_{j1}, \ldots, a_{jd})^T (j = 1, \ldots, d)$ を用いて線形変換すれば，

$$\boldsymbol{s}_j = (s_{1j}, \ldots, s_{Nj})^T = \bar{\boldsymbol{X}}\boldsymbol{a}_j$$

が得られる．この変換後のデータの分散は

$$Var\{\boldsymbol{s}_j\} = \frac{1}{N}\boldsymbol{s}_j^T\boldsymbol{s}_j = \frac{1}{N}(\bar{\boldsymbol{X}}\boldsymbol{a}_j)^T\bar{\boldsymbol{X}}\boldsymbol{a}_j = \boldsymbol{a}_j^T\frac{1}{N}\bar{\boldsymbol{X}}^T\bar{\boldsymbol{X}}\boldsymbol{a}_j = \boldsymbol{a}_j^T Var\{\bar{\boldsymbol{X}}\}\boldsymbol{a}_j$$

となる．

分散を最大にする射影ベクトル

$Var\{\boldsymbol{s}_j\}$ が最大となる射影ベクトルは，係数ベクトル $\boldsymbol{a}_j$ のノルムを 1 に制約したラグランジュ関数

$$E(\boldsymbol{a}_j) = \boldsymbol{a}\ T_j Var\{\bar{\boldsymbol{X}}\}\boldsymbol{a}_j - \lambda(\boldsymbol{a}_j^T\boldsymbol{a}_j - 1)$$

を最大にする $\boldsymbol{a}_j$ を見つければよい.（ λ はラグランジュ未定乗数.） $\boldsymbol{a}_j$ で微分して 0 とおけば

$$\frac{\partial E(\boldsymbol{a}_j)}{\partial \boldsymbol{a}_j} = 2Var\{\bar{\boldsymbol{X}}\}\boldsymbol{a}_j - 2\lambda\boldsymbol{a}_j = 0 \quad \text{より} \quad Var\{\bar{\boldsymbol{X}}\}\boldsymbol{a}_j = \lambda\boldsymbol{a}_j$$

が得られ，この固有値問題を解くことで分散最大となる射影ベクトル $\boldsymbol{a}_j$ が得られる.
固有値を $\lambda_1 \geq \ldots \geq \lambda_d$，対応する固有ベクトルを $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ とする.
共分散行列が実対称行列であることから固有ベクトルは相互に直交し，

$$\boldsymbol{a}_j^T\boldsymbol{a}_j = \delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

が成り立つ．δ_{ij} はクロネッカーのデルタである．

全分散量と寄与率

元のデータの共分散行列は $d \times d$ の行列なので，得られる非ゼロ固有値の数は共分散行列のランクで決まり，最大 d である．最大固有値に対応する固有ベクトルで線形変換した特徴量の分散は，

$$Var\{\boldsymbol{s}_1\} = \boldsymbol{a}_1^T Var\{\bar{\boldsymbol{X}}\}\boldsymbol{a}_1 = \lambda_1 \boldsymbol{a}_1^T \boldsymbol{a}_1 = \lambda_1$$

となり，最大固有値に一致する．
k 番目の大きさの固有値に対応する固有ベクトルで線形変換された特徴量を第 k 主成分という．
変換された特徴量の全分散は

$$V_{total} = \sum_{i=1}^{d} Var\{\boldsymbol{s}_i\} = \sum_{i=1}^{d} \lambda_i$$

となり，これは元データの持つ全分散量とも一致する．
第 k 成分の寄与率 c_k，第 k 成分までの累積寄与率 r_k は

$$c_k = \frac{\lambda_k}{V_{total}}, \qquad r_k = \frac{\sum_{i=1}^{k} \lambda_i}{V_{total}}$$

で定義される．

特異値分解

$n \times p$ 行列 $\boldsymbol{X}$ の特異値分解

$$
\begin{aligned}
\boldsymbol{X} &= \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T \\
&= \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_p \end{pmatrix}
\begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_p} \end{pmatrix}
\begin{pmatrix} \boldsymbol{v}_1^T \\ \boldsymbol{v}_2^T \\ \vdots \\ \boldsymbol{v}_p^T \end{pmatrix}
\end{aligned}
$$

- $\boldsymbol{U}$ は，$\boldsymbol{X}\boldsymbol{X}^T$ の非ゼロ固有値に対応する固有ベクトルからなる，$n \times p$ 列正規直交行列
- $\boldsymbol{V}$ は，$\boldsymbol{X}^T\boldsymbol{X}$ の非ゼロ固有値に対応する固有ベクトルからなる，$p \times p$ 列正規直交行列
- $\boldsymbol{\Lambda}$ は，$\boldsymbol{X}\boldsymbol{X}^T$ または $\boldsymbol{X}^T\boldsymbol{X}$ の非ゼロ固有値の平方根（特異値という）を，$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ の順に並べて対角要素とした $p \times p$ 対角行列

特異値分解 2

特異値分解の定義から

$$\boldsymbol{X}\boldsymbol{X}^T = (\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T)(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^T)^T = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\boldsymbol{T}}\boldsymbol{U}^{\boldsymbol{T}}$$

$\boldsymbol{U}$ をかければ，$\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\boldsymbol{T}}$ となる．これを列ベクトルに分解すると，

$$\begin{pmatrix} \boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{u_1} & \boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{u_2} & \cdots & \boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{u_p} \end{pmatrix} = \begin{pmatrix} \lambda_1\boldsymbol{u}_1 & \lambda_2\boldsymbol{u}_2 & \cdots & \lambda_p\boldsymbol{u}_p \end{pmatrix}$$

特異値分解と主成分分析の関係は，$n \times p$ 行列 $\boldsymbol{X}$ を p 個の属性を持つ n 個のデータ（あらかじめ平均ベクトルが引かれている）と考えれば，$\boldsymbol{X}^T\boldsymbol{X}$ は共分散行列であり，その固有値と固有ベクトルが λ_i と $\boldsymbol{u}_i$ になっている．

特異値分解 3

$\boldsymbol{X} = \boldsymbol{U\Lambda V}^T$ より $\boldsymbol{XV} = \boldsymbol{U\Lambda}$.

$$\begin{pmatrix} \boldsymbol{Xv_1} & \boldsymbol{Xv_2} & \cdots & \boldsymbol{Xv_p} \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1}\boldsymbol{u_1} & \sqrt{\lambda_2}\boldsymbol{u_2} & \cdots & \sqrt{\lambda_p}\boldsymbol{u_p} \end{pmatrix}$$

$$Var\{\boldsymbol{Xv_1}\} = (\boldsymbol{Xv_1})^T(\boldsymbol{Xv_1}) = \lambda_1$$

分散は第 1 主成分の最大固有値に一致し最大となる.
第 1 主成分から第 $q(\leq p)$ 主成分までの v_i で構成された部分空間

$$\tilde{\boldsymbol{V}} = \begin{pmatrix} \boldsymbol{v_1} & \cdots & \boldsymbol{v_q} \end{pmatrix}$$

へのデータ $\boldsymbol{X}$ の射影 $\boldsymbol{X\tilde{V}}$ は, 共分散行列が

$$Var\{\boldsymbol{X\tilde{V}}\} = \boldsymbol{\tilde{V}^T V \Lambda^2 V^T \tilde{V}}$$

となる.

特異値分解 4

$\tilde{\boldsymbol{V}}^{\boldsymbol{T}}\boldsymbol{V}$ は,

$$\tilde{\boldsymbol{V}}\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V_1^T} \\ \vdots \\ \boldsymbol{V_q^T} \end{pmatrix} \begin{pmatrix} \boldsymbol{v_1} & \cdots & \boldsymbol{v_q} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}$$

となるので，共分散行列は,

$$Var\{\boldsymbol{X}\tilde{\boldsymbol{V}}\} = \boldsymbol{\Lambda_q^2}$$

$\boldsymbol{\Lambda_q}$ は特異値行列の最初の q 個の特異値以外を 0 にした対角行列．
これを用いた特異値分解

$$\tilde{\boldsymbol{X}} = \boldsymbol{U}\boldsymbol{\Lambda_q}\boldsymbol{V^T} = \sum_{i=1}^{q} \sqrt{\lambda_i}\boldsymbol{u_i}\boldsymbol{v_i^T}$$

は，$\boldsymbol{X}$ のランク q の誤差最小という意味での最良近似となっている．

部分空間法

- クラスごとに部分空間を構成する正規直交基底を学習データから求め，入力データを各クラスの部分空間に射影して識別する手法.
- 相関行列を使う方法と，共分散行列を使う方法がある
  - 相関行列を使う方法に CLAFIC 法
- 相関行列の場合はデータそのもののばらつきが評価される
- 共分散行列の場合は平均ベクトルを中心としたばらつきが評価される
- どちらの方法も基本的な認識原理は同じ

CLAFIC 法

データ $\boldsymbol{x} \in \mathbb{R}^d$ は、K 個のクラス $C_1, \ldots, C_K$ のどれかに属しているものとする．クラスごとの部分空間を $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_K$ とし，クラス i の部分空間 $\boldsymbol{S}_i$ を張る基底ベクトルを，

$$\boldsymbol{u_{i1}}, \ldots, \boldsymbol{u_{id_i}}$$

で表す．(d_i は部分空間 $\boldsymbol{S}_i$ の次元)
部分空間 $\boldsymbol{S_i}$ へ正射影した長さの期待値

$$\begin{aligned} E\left\{\sum_{j=1}^{d_i}(\boldsymbol{x}^T\boldsymbol{u_{ij}})^2|\boldsymbol{x} \in C_i\right\} &= E\left\{\boldsymbol{x}^T\left(\sum_{j=1}^{d_i}\boldsymbol{u_{ij}}\boldsymbol{u_{ij}^T}\right)\boldsymbol{x}|\boldsymbol{x} \in C_i\right\} \\ &= E\left\{\boldsymbol{x^T}\boldsymbol{P_i}\boldsymbol{x}|\boldsymbol{x} \in C_i\right\} \end{aligned}$$

が最大となるように $\|\boldsymbol{u_{ij}}\| = 1$ の制約の下で選択する．
$\boldsymbol{P_i} = \sum_{j=1}^{d_i}\boldsymbol{u_{ij}}\boldsymbol{u_{ij}^T}$ をクラス i の射影行列とよび，これを用いた部分空間法の識別規則は，

すべての $j \neq i$ について,$\boldsymbol{x^T}\boldsymbol{P_j}\boldsymbol{x} < \boldsymbol{x^T}\boldsymbol{P_i}\boldsymbol{x}$ であれば,$\boldsymbol{x} \in C_i$

となる．