はじめてのパターン認識

第９章　部分空間法

11t4062a　広原裕亮

部分空間（１）

- ｄ次元のベクトル空間Ｖ

  －$x_1,\ldots,x_r (r \leqq d)$
- Ｖの部分空間

  →$x_1,\ldots,x_r$の一次結合全体の集合

  $W=\{a_1x_1+\ldots+ a_rx_r \mid a_i \in R, i=1,\ldots,r\}$
- すなわち$x_1,\ldots,x_r$は一次独立であるということ

部分空間（２）

・ベクトル空間Vは、部分空間Sとそれと直交する部分空間$S^{\perp}$に分解でき

　－$V=S\cup S^{\perp}$、$S\cap S^{\perp}=\emptyset$が成り立つ。

・任意のベクトルxは$x=x_S+x_S{}^{\perp}$のように分解できる

・次元の縮約という観点も含め正規直交基底を求める手法として主成分分析がある

主成分分析

- 学習データの$x_i=(x_{i1},\ldots x_{id})^T$（i=1,・・・N）の分散が最大になる方向への線形変換を求める手法である。

→共分散行列を求めることが出発点

N個のデータからなるデータ行列を$X=(x_1,\ldots x_N)^T$

平均ベクトル　$\bar{x}=(\bar{x}1,\ldots,\bar{x}d)T$、$\bar{X}= X－\bar{x}$

共分散行列 $\Sigma＝Var\{\bar{X}\}=1/N\ \bar{X}^T\ \bar{X}$

主成分分析

- 係数ベクトル$a_j=(a_{j1},\ldots,a_{jd})^T$ (j=1,…d)を用いて線形変換

$$\rightarrow s_j=(s_{1j},\ldots,a_{Nj})^T=\bar{X}a_j$$

このデータの分散は

$$\mathrm{Var}\{s_j\}=a_j^T\mathrm{Var}\{\bar{X}\}a_j$$

この分散が最大となる射影ベクトルは、a_jのノルムを1に制約したラグランジェ関数

$E(a_j)=a_j^T\mathrm{Var}\{\bar{X}\}a_j-\lambda(a_j^T a_j-1)$

を最大にするa_jを見つける。

主成分分析

・A_jで微分して0とおくと

$$Var\{\bar{X}\}a_j = \lambda a_j$$

　→固有値問題を解く事により分散最大となる射影ベクトルa_jが得られることを示している。

・固有値を$\lambda_1 \geqq \ldots \geqq \lambda_d$とし、対応する固有ベクトルを$a_1, \ldots, a_d$とする。共分散行列が実対称行列であることから

$$a_i^T a_j = \{ 1(i=j) \quad 0(i \neq j) \}$$

　が成り立つ

主成分分析

- 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の分散は

$$\mathrm{Var}\{ s_1\}=\lambda_1$$

より最大固有値に一致する。

K番目の固有値に対応する固有ベクトルで変換された特徴量を第k主成分という

特異値分解

- 任意のn×p行列Xは

$X=UAV^T$

$$=(u_1\ u_2\ \dots\ u_p)\begin{pmatrix}\sqrt{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p\end{pmatrix}\begin{pmatrix} v_1{}^{\mathrm{T}} \\ \vdots \\ v_{\mathrm{p}}{}^{\mathrm{T}}\end{pmatrix}$$

３つの行列の積に分解することができる。これを特異値分解という。

特異値分解

- Uは、XX^Tの非ゼロ固有値に対応する固有ベクトルからなる、n×p列正規直交行列
- Vは、X^TXの非ゼロ固有値に対応する固有ベクトルからなる、p×p列正規直交行列
- Λは、 XX^TまたはX^TXの非ゼロ固有値の平方根を$\lambda_1 \geqq \ldots \geqq \lambda_p$の順に並べて対角要素としたp×p対角行列

特異値分解

- ｎ×ｐ行列Xをp個の属性を持つn個のデータと考えれば、X^TXは共分散行列
- 固有値と固有ベクトルがλ_iとv_iとなっている

→これらから特異値分解と主成分分析の関係は明らか。

特異値分解

- また$X=U\Lambda V^T$より$XV＝U\Lambda$が成り立つので

$(X\ v_1\ \ldots\ X\ v_p)=(\sqrt{\lambda_1}u_1\ \ldots\sqrt{\lambda_p}u_p)$ となる

- データ行列v_1で線形変換したベクトルが$\sqrt{\lambda_1}u_1$となるので分散は

$$Var\{X\ v_1\}=\lambda_1$$

となり、第一主成分の最大固有値に一致し最大となる

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
- 相関行列を使う方法はCLAFIC法という

CLAFIC法

K個のクラス　$C_1,\ldots,C_K$

クラスの部分空間　$S_1,\ldots,S_K$

クラスiの部分空間S_iを張る基底ベクトル$\{ u_{i1},\ldots, u_{idi}\}$

データxがi番目のクラスに所属しているとき、部分空間S_iへ正射影した長さの期待値

$E\{\sum_{j=1}^{d_i}(xTu_{ij})^2|x \in Ci\} = E\{x^T P_i x \mid x \in C_i \}$

が最大になるように$||u_{ij}||=1$の制約の下で選択する。

CLAFIC法

- 基準を満たすi番目の部分空間の基底ベクトルは次のラグランジェ関数を微分して0とおいて求める

$$L_i(u_{ij}) = \ldots = \sum_{j=1}^{d_i} u_{ij}^T Q_i u_{ij} - \sum_{j=1}^{d_i} \lambda_{ij}(u_{ij}^T u_{ij} - 1)$$

$$\partial L_i(u_{ij})/\partial u_{ij}=0$$

$$Q_i u_{ij}=\lambda_{ij} u_{ij}$$

よりi番目のクラスの相関行列に関する固有値問題となる

CLAFIC法

- Q_iの固有値 $\lambda_{i1} \geqq \ldots \geqq \lambda_i d_i$
- 対応する固有ベクトル u_{ij}
- i番目のクラスのデータxの、j番目の基底への射影の長さの2乗の期待値は

$$u_{ij}^T Q_i u_{ij} = \lambda_{ij}$$

よりj番目の固有値の大きさに等しくなる

CLAFIC法

- Q_iから求めたi番目のクラスの固有ベクトルは正規直交基底なので、個々の基底の射影の和で部分空間への射影の長さの２乗の期待値が表せる。

$$\sum_{j=1}^{d_i} u_{ij}{}^{T} Q_i u_{ij} = \sum_{j=1}^{d_i} \lambda_{ij}$$

よって上記のようになる。

CLAFIC法

- 部分空間の次元により長さが異なる。
- 射影した長さで所属クラスを決める

→次元の大きさでク所属クラスが変わってしまう

そこで固有値の累積値を用いる

$$a(d_i) = \sum_{j=1}^{d_i} \lambda_{ij}$$

CLAFIC法

- 共通なパラメータkを導入し

$$a(d_i-1) \leqq k < a(d_i)$$

を満たすd_iをi番目のクラスの部分空間の次元とする。

Kは忠実度とよばれる。