2010 年度前期　応用統計学　第７回

主成分分析 (3) ー 多変量の主成分分析とＫＬ変換

今回は，多変量の主成分分析，および「第１主成分の分散が最大」と「主成分間の相関がないこと」が同じ意味であることと，主成分分析をデータ圧縮に用いる Karhunen-Loève 変換について説明します．

多変量の主成分分析

今回は，p 個の変量 $x_1, x_2, \ldots, x_p$ からなるデータ n 個が観測されているとします．２変量の主成分分析の場合と同様に，新しい変量 z を

$$z = a_1 x_1 + a_2 x_2 + \cdots + a_p x_p \tag{1}$$

として，z の分散を最大にすることを考えます．すなわち，軸 $x_1, x_2, \ldots, x_p$ で構成される p 次元空間で，新しい軸 z を (1) 式によって作り，軸上でのデータの分散が最大になるようにします．

z の分散 $V(z)$ は

$$\begin{aligned}
V(z) &= \frac{1}{n}\sum_{i=1}^{n}\left\{(a_1^2(x_{1i}-\bar{x_1}) + (a_2^2(x_{2i}-\bar{x_2}) + \cdots + (a_p^2(x_{pi}-\bar{x_p})\right\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} a_1^2(x_{1i}-\bar{x_1})^2 + \cdots + \frac{1}{n}\sum_{i=1}^{n} a_p^2(x_{pi}-\bar{x_p})^2 \\
&\quad +\frac{1}{n}\sum_{i=1}^{n} a_1 a_2 (x_{1i}-\bar{x_1})(x_{2i}-\bar{x_2}) + \cdots + \frac{1}{n}\sum_{i=1}^{n} a_1 a_p (x_{1i}-\bar{x_1})(x_{pi}-\bar{x_p}) \\
&\quad +\cdots+ \\
&\quad +\frac{1}{n}\sum_{i=1}^{n} a_p a_2 (x_{pi}-\bar{x_p})(x_{2i}-\bar{x_2}) + \cdots + \frac{1}{n}\sum_{i=1}^{n} a_p a_{(p-1)} (x_{pi}-\bar{x_p})(x_{(p-1)i}-\bar{x_{(p-1)}}) \\
&= (s_{11}a_1^2 + \cdots + s_{pp}a_p^2) + (s_{12}a_1 a_2 + \cdots + s_{1p}a_1 a_p) + \cdots + (s_{p1}a_p a_2 + \cdots + s_{p(p-1)}a_p a_{(p-1)}) \\
&= \sum_{j=1}^{p}\sum_{k=1}^{p} s_{jk} a_j a_k
\end{aligned} \tag{2}$$

と表されます．

また，やはり２変量の場合と同様に，$a_1, a_2, \ldots, a_p$ を方向余弦と考えると，$a_1, a_2, \ldots, a_p$ は

$$a_1^2 + \cdots + a_p^2 = \sum_{j=1}^{p} a_j^2 = 1 \tag{3}$$

を満たす必要があります．

したがって，(2) 式の $V(z)$ を (3) 式の制約条件のもとで最大化するように，Lagrange の未定乗数法を適用します．これによると，この最大化問題は未定乗数を λ として

$$F(a_1, \cdots, a_p) = \sum_{j=1}^{p}\sum_{k=1}^{p} s_{jk} a_j a_k - \lambda\left(\sum_{j=1}^{p} a_j^2 - 1\right) \tag{4}$$

を最大化する $a_1, a_2, \dots, a_p$ を求める問題となります．そこで (4) 式を a_l $(l = 1, \dots, p)$ で偏微分して 0 とおくと，

$$
\begin{aligned}
\frac{\partial F}{\partial a_l} = \underbrace{\sum_{j=1}^{p} s_{jl} a_j}_{(k=l)} + \underbrace{\sum_{k=1}^{p} s_{lk} a_k}_{(j=l)} - 2\lambda a_l &= 0 \\
2\sum_{k=1}^{p} s_{lk} a_k - 2\lambda a_l &= 0 \\
\sum_{k=1}^{p} s_{lk} a_k - \lambda a_l &= 0
\end{aligned} \tag{5}
$$

が得られます．この式はすなわち

$$
s_{l1}a_1 + s_{l2}a_2 + \cdots + s_{lp}a_p - \lambda a_l = 0 \tag{6}
$$

で，これが $l = 1, \dots, p$ でなりたつわけですから

$$
\begin{aligned}
s_{11}a_1 + s_{12}a_2 + \cdots + s_{1p}a_p &= \lambda a_1 \\
s_{21}a_1 + s_{22}a_2 + \cdots + s_{2p}a_p &= \lambda a_2 \\
&\vdots \\
s_{p1}a_1 + s_{p2}a_2 + \cdots + s_{pp}a_p &= \lambda a_p
\end{aligned} \tag{7}
$$

となり，これを行列で表現すると

$$
\begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{12} & s_{22} & \cdots & s_{2p} \\ \vdots & & \ddots & \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \lambda \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \tag{8}
$$

という分散共分散行列の固有値問題になります．

ここで，z の分散と固有値との関係を見てみましょう．(6) 式の両辺に a_l をかけると

$$
s_{l1}a_1a_l + s_{l2}a_2a_l + \cdots + s_{lp}a_pa_l = \lambda a_l^2 \tag{9}
$$

が得られます．これが $l = 1, \dots, p$ でなりたちますから，それらをすべて足すと

$$
\begin{aligned}
\sum_{l=1}^{p} \left(s_{l1}a_1a_l + s_{l2}a_2a_l + \cdots + s_{lp}a_pa_l \right) &= \lambda \sum_{l=1}^{p} a_l^2 \\
\sum_{k=1}^{p} \sum_{l=1}^{p} s_{lk}a_ka_l = \lambda &
\end{aligned} \tag{10}
$$

となります（(3) 式より，$\sum_{l=1}^{p} a_l^2 = 1$）．左辺は (2) 式から $V(z)$ ですから，

$$
V(z) = \lambda \tag{11}
$$

が得られます．すなわち，固有値は z 上の分散となります．そこで，最大の固有値を $\lambda_{(1)}$ とするとき，それに対応する固有ベクトル $(a_{1(1)}, a_{2(1)}, \dots, a_{p(1)})'$ を用いた $z_{(1)}$ をつくると[1]，この $z_{(1)}$ 上で分散が最大

[1] 記号「$'$」は，ベクトル・行列の転置を表します．数学の本では上付きの ^{T} を使うことが多いですが，統計の本では「$'$」を使うことが多いようです．

になります．これが第 1 主成分です．以下，各固有値に対応する固有ベクトルを用いた各 z を，大きな固有値に対応するものから順に $z_{(2)}, z_{(3)}, \ldots, z_{(p)}$ と表し，第 2 主成分，第 3 主成分，…，第 p 主成分とよびます．

分散共分散行列は対称行列ですから，異なる固有値に対応する固有ベクトルは互いに直交します（付録 1 を参照）．すなわち，p 変量の主成分分析は，もとの変量 $x_1, x_2, \ldots, x_p$ を p 次元空間と考えたとき，この座標軸を互いに直交したまま回転して $z_{(1)}, z_{(2)}, \ldots, z_{(p)}$ という別の直交座標に変換し，軸上での分散が $z_{(1)}, z_{(2)}, \ldots, z_{(p)}$ の順に集中するようにしたものということができます．

行列の対角化の意味

ここまでで，「第 1 主成分の分散が最大」であることと「第 1 主成分は固有値最大の固有ベクトル」であることが同じ意味であることを説明しました．しかし，「第 1 主成分の分散が最大」であることと「各主成分の間の相関がない，すなわち分散共分散行列が対角行列である」ことが同じ意味であることは，直観的にしか説明していませんでした．そこで，「分散共分散行列の対角化」の意味について説明します．

固有値 λ_k に対応する固有ベクトルを $(a_{1(k)}, a_{2(k)}, \ldots, a_{p(k)})'$ とすると，第 k 主成分 z_k は (1) 式から

$$
\begin{aligned}
z_k &= a_{1(k)}x_1 + a_{2(k)}x_2 + \cdots + a_{p(k)}x_p \\
&= \begin{pmatrix} a_{1(k)} & a_{2(k)} & \cdots & a_{p(k)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}
\end{aligned} \tag{12}
$$

とあらわされます．さらに，その固有ベクトルを，対応する固有値の大きさの順に並べた行列を，P とします．すなわち，

$$
P = \begin{pmatrix} a_{1(1)} & a_{1(2)} & \cdots & a_{1(p)} \\ a_{2(1)} & a_{2(2)} & \cdots & a_{2(p)} \\ \vdots & & \ddots & \\ a_{p(1)} & a_{p(2)} & \cdots & a_{p(p)} \end{pmatrix} \tag{13}
$$

です．このとき，分散共分散行列を S とすると，(8) 式から

$$
S \begin{pmatrix} a_{1(k)} \\ a_{2(k)} \\ \vdots \\ a_{p(k)} \end{pmatrix} = \lambda_k \begin{pmatrix} a_{1(k)} \\ a_{2(k)} \\ \vdots \\ a_{p(k)} \end{pmatrix} \quad (k = 1, 2, \ldots, p) \tag{14}
$$

がなりたちますから，$k = 1, 2, \ldots, p$ を合わせると

$$\begin{aligned} & S \begin{pmatrix} a_{1(1)} & a_{1(2)} & \cdots & a_{1(p)} \\ a_{2(1)} & a_{2(2)} & \cdots & a_{2(p)} \\ \vdots & & \ddots & \\ a_{p(1)} & a_{p(2)} & \cdots & a_{p(p)} \end{pmatrix} \\ = & \begin{pmatrix} a_{1(1)} & a_{1(2)} & \cdots & a_{1(p)} \\ a_{2(1)} & a_{2(2)} & \cdots & a_{2(p)} \\ \vdots & & \ddots & \\ a_{p(1)} & a_{p(2)} & \cdots & a_{p(p)} \end{pmatrix} \begin{pmatrix} \lambda_{(1)} & & & 0 \\ & \lambda_{(2)} & & \\ & & \ddots & \\ 0 & & & \lambda_{(p)} \end{pmatrix} \end{aligned} \tag{15}$$

で，これは

$$\Lambda = \begin{pmatrix} \lambda_{(1)} & & & 0 \\ & \lambda_{(2)} & & \\ & & \ddots & \\ 0 & & & \lambda_{(p)} \end{pmatrix} \tag{16}$$

とおくと

$$SP = P\Lambda, \text{ すなわち } P^{-1}SP = \Lambda \tag{17}$$

と表されます．(3) 式のように各固有ベクトルは正規化されており，また分散共分散行列 S は対称行列なので，前節で述べたように各固有ベクトルは直交しています．すなわち，各固有ベクトルは正規直交基底をなしています．したがって P は正規直交行列です．P が直交行列のとき $P^{-1} = P'$ が成り立つので (付録 2 を参照)，

$$P'SP = \Lambda, \text{ または } S = P\Lambda P' \tag{18}$$

となります．これを，対称行列 S の**対角化** (diagonalization) とよんでいます．

一方，(12) 式を $k = 1, 2, \ldots, p$ についてまとめると，

$$\begin{aligned} \begin{pmatrix} z_{(1)} \\ z_{(2)} \\ \vdots \\ z_{(p)} \end{pmatrix} &= \begin{pmatrix} a_{1(1)} & a_{2(1)} & \cdots & a_{p(1)} \\ a_{1(2)} & a_{2(2)} & \cdots & a_{p(2)} \\ \vdots & & \ddots & \\ a_{1(p)} & a_{2(p)} & \cdots & a_{p(p)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \\ &= P' \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \end{aligned} \tag{19}$$

となりますから，行列 P' はもとの変量 $x_1, x_2, \ldots, x_p$ を新たな変量（主成分）$z_{(1)}, z_{(2)}, \ldots, z_{(p)}$ に変換する行列となります（この形式による変換を**直交変換** (orthogonal transformation) とよんでいます)．

したがって，(18) 式による対角化は，「いったん主成分 $z_{(1)}, z_{(2)}, \ldots, z_{(p)}$ に P' で変換すると」「主成分を座標とする空間では，分散共分散行列は固有値をならべた対角行列 Λ であり」「$(P')^{-1} = P$ でもう一度変換して，もとの変量 $x_1, x_2, \ldots, x_p$ にもどると，もとの変量 $x_1, x_2, \ldots, x_p$ での分散共分散行列が得られる」ということを示しています．主成分 $z_{(1)}, z_{(2)}, \ldots, z_{(p)}$ による座標では，分散共分散行列が対角行列 Λ

で表される，ということは，この座標では共分散が全て０，すなわち各主成分 $z_{(1)}, z_{(2)}, \ldots, z_{(p)}$ が，互いに無相関であることを意味しているわけです．

Karhunen-Loève 変換

「第 k 主成分の分散の，分散の合計に対する割合」，すなわち λ_k の $(\lambda_1 + \lambda_2 + \cdots + \lambda_p)$ に対する割合が，第 k 主成分の**寄与率**です．つまり，寄与率は，前回述べたように「データの価値のうち，第 k 主成分で表されているものの割合」を表しています．ここで，ある k 以降の主成分の寄与率が 0，あるいはほぼ 0 とみなせる場合を考えてみましょう．例えば，２変量の主成分分析で，第２主成分の寄与率をほぼ 0 とみなすことができるとしましょう．すると，第２主成分 $z_{(2)}$ の分散がほぼ 0 ということになりますから，$z_{(2)}$ は不要で，$z_{(2)}$ の平均に置き換えてしまってもよく，もとのデータは第 1 主成分 $z_{(1)}$ という 1 つの変量だけでほとんど表現できることになります．

主成分分析では，なるべく第１，２，といった番号の若い主成分が分散がなるべく大きくなるように変換しています．いいかえれば，最後のほうの番号の主成分は寄与率がなるべく小さくなるようにしているわけで，最後のほうの番号の主成分を捨てて変換後の変量の数を減らしたとしても，元のデータとの誤差は最小になります．主成分分析には，このようにして，もとのデータのもつ情報をなるべく損なわずにデータ量を減らすことができるという側面があります．

そこで，いま p 変量のデータが多数あるとします．これを伝達するのに，1 度に $(p/2)$ 変量しか使えないとしましょう．このときに情報をなるべく損なわずに伝達するには，どうすればよいでしょうか？それは，ここまで述べたように，データを表す変量を主成分に変換し，第 1 から第 $(p/2)$ までの $p/2$ 個の主成分だけを各データについて伝達して，それ以外の主成分については，全データに共通である各主成分の平均値を 1 回だけ伝達しておけばよいことがわかります．もとの変量を各主成分に変換する (19) 式の直交変換を，この意味で用いるとき **Karhunen-Loève 変換（KL 変換）**とよんでいます．

この主成分を受け取った側では，主成分への変換の逆変換（z から x への変換）を行って，元のデータに戻します．この逆変換は

$$
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \simeq (P')^{-1} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{p/2} \\ \overline{z_{p/2+1}} \\ \vdots \\ \overline{z_p} \end{pmatrix} = P \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{p/2} \\ \overline{z_{p/2+1}} \\ \vdots \\ \overline{z_p} \end{pmatrix} \tag{20}
$$

となります．こうすると，変量の数を 1/2 にしたときに情報の損失が最小になります．

しかし，この方法を用いるには，取り扱う全てのデータを観察して分散共分散行列を求める必要がありますが，一般にはそれは不可能です．つまり，どんなデータが入ってくるかわからなければ，KL 変換はできませんが，そんなことはわかるはずもありません．そこで，主成分のかわりに適当な基底ベクトルをあらかじめ決めておき，情報の損失が最小ではなくともなるべく少なくする方法が用いられています．

この方法がもっとも広く用いられているのが，デジタル画像の伝達です．デジタル画像は，各画素を

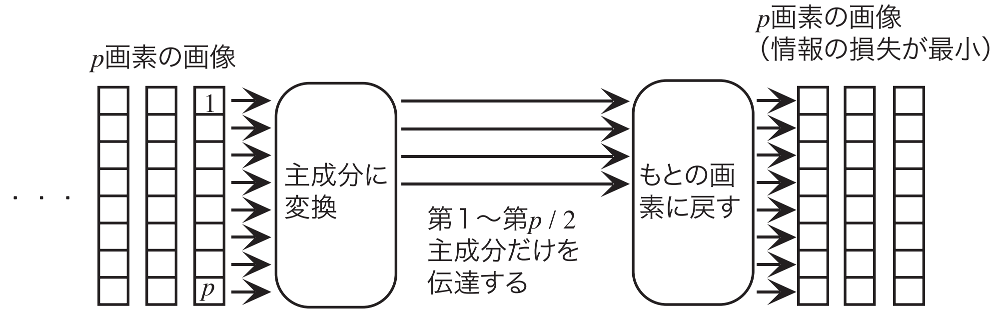

図 1: KL 変換によるデジタル画像の圧縮

変量とする多変量データと考えられます．デジタル画像を伝達する際，見た目がほとんど変わらないようにデータ量を減らすために，KL 変換の考え方にもとづいて情報の損失を最小にしています．このとき，どんな画像を伝達するかがわからなければ分散共分散行列を求めることはできないため，あらかじめ基底ベクトルを決めています．JPEG 方式の基礎となっている「離散コサイン変換」はその代表的なものです．

付録 1：対称行列の固有ベクトル

列ベクトル $\boldsymbol{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ の内積 $\boldsymbol{a} \cdot \boldsymbol{b}$ は

$$\begin{aligned} \boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}'\boldsymbol{b} &= (a_1, \ldots, a_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \\ &= a_1 b_1 + \cdots + a_n b_n \end{aligned} \tag{A1}$$

で定義されます．ここで，行列 A を n 次（実）正方行列とすると，

$$(A\boldsymbol{a}) \cdot \boldsymbol{b} = (A\boldsymbol{a})'\boldsymbol{b} = (\boldsymbol{a}'A')\boldsymbol{b} = \boldsymbol{a}'(A'\boldsymbol{b}) = \boldsymbol{a} \cdot (A'\boldsymbol{b}) \tag{A2}$$

がなりたちますから，行列 A が（実）対称行列であれば $(A\boldsymbol{a}) \cdot \boldsymbol{b} = \boldsymbol{a} \cdot (A\boldsymbol{b})$ がなりたちます．

さて，行列 A が（実）対称行列で，λ_1, λ_2 をその相異なる固有値とし，$\boldsymbol{a}_1, \boldsymbol{a}_2$ をそれぞれに対応する固有ベクトルとします．このとき，$\lambda_1 \boldsymbol{a}_1 \cdot \boldsymbol{a}_2 = (\lambda_1 \boldsymbol{a}_1) \cdot \boldsymbol{a}_2 = (A\boldsymbol{a}_1) \cdot \boldsymbol{a}_2$ で，A が（実）対称行列であることから，上で示した関係より $(A\boldsymbol{a}_1) \cdot \boldsymbol{a}_2 = \boldsymbol{a}_1 \cdot (A\boldsymbol{a}_2)$ となり，さらに $\boldsymbol{a}_1 \cdot (A\boldsymbol{a}_2) = \boldsymbol{a}_1 \cdot (\lambda_2 \boldsymbol{a}_2) = \lambda_2 \boldsymbol{a}_1 \cdot \boldsymbol{a}_2$ となります．つまり $\lambda_1 \boldsymbol{a}_1 \cdot \boldsymbol{a}_2 = \lambda_2 \boldsymbol{a}_1 \cdot \boldsymbol{a}_2$ ですが，ここで $\lambda_1 \neq \lambda_2$ ですから，$\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 = 0$ となります．すなわち，$\boldsymbol{a}_1, \boldsymbol{a}_2$ は直交することがわかります．

付録２：正規直交行列と逆行列

n 次正規直交行列 A を，列ベクトル $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ を使って $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ と表します．このとき，A' は行ベクトル $\boldsymbol{a}'_1, \ldots, \boldsymbol{a}'_n$ を使って $A' = \begin{pmatrix} \boldsymbol{a}'_1 \\ \vdots \\ \boldsymbol{a}'_n \end{pmatrix}$ と表されますから，

$$\begin{aligned} A'A &= \begin{pmatrix} \boldsymbol{a}'_1 \\ \vdots \\ \boldsymbol{a}'_n \end{pmatrix} (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) \\ &= \begin{pmatrix} \boldsymbol{a}_1 \cdot \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_1 \cdot \boldsymbol{a}_n \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}_n \cdot \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \cdot \boldsymbol{a}_n \end{pmatrix} \end{aligned} \tag{A3}$$

という，A を構成する列ベクトルどうしの内積を要素とする行列となります．A は正規直交行列ですから，内積 $\boldsymbol{a}_i \cdot \boldsymbol{a}_j$ は $i = j$ のとき 1，$i \neq j$ のとき 0 です．したがって，(A3) 式の行列は単位行列となるので，$A' = A^{-1}$ となります．