クレジット:
UTokyo Online Education 統計データ解析 Ⅱ 2018 小池祐太

東京大学 THE UNIVERSITY OF TOKYO

CC BY NC ND

統計データ解析 (II) 第 10 回

小池祐太

2018 年 6 月 14 日

1. 重回帰分析: 交互作用モデル・変数の非線形変換

2. 主成分分析
   - 目的
   - R での実行
   - 計算法
   - 分析の評価
     - 寄与率
     - バイプロット
   - 補足: 計算の詳細の説明
     - 主成分の計算 ($d = 1$ の場合)
     - 主成分の計算 ($d = 2$ の場合)
     - 寄与率の計算

重回帰分析: 交互作用モデル・変数の非線形変換

- 目的変数を Y, 説明変数を $X_1, \ldots, X_p$ で表す
- 前述のように, 説明変数として $X_j X_k$ (交差項と呼ばれる) や $\log X_j$ といったものを新たに加えることで, 目的変数 Y と説明変数 $X_1, \ldots, X_p$ の非線形な関係をモデル化することができる
- R においてはこのような回帰式の柔軟なモデル化を可能にするためのモデル式の記述法が実装されている
- 実行例 `cross.r`

主成分分析

- **主成分分析 (principal component analysis, PCA)**
  - 多数の変数/データが与えられたとき, 変数/データたちのもつ情報を効率的に縮約して少数の特徴量を構成することで, 変数/データ間の関係を明らかにするための分析法
  - イメージ: 株価指数. ただし, 通常の株価指数とは異なり, 純粋にデータの情報のみを使って指数を作成する
  - 「教師なし学習」の代表的手法の 1 つ
- 主成分分析では, 特徴量は変数/データたちの線形結合として構成
- 幾何学的にいうと, 観測データの含まれる p 次元空間にうまく座標軸を設定することにより, その座標軸上にデータのもつ情報が最大限反映されるようにすることが目的

主成分分析

- 数式で表すと, 与えられた変数を $X_1, \ldots, X_p$ としたとき, d を p 以下の正の整数とし (通常 p より小さくとる), $X_1, \ldots, X_p$ の線形結合として表される d 個の変数

$$Z_k = a_{1k}X_1 + \cdots + a_{pk}X_p \qquad (k = 1, \ldots, d)$$

を, もとの変数のもつ情報を最大限保持しつつ適切に構成することが目的

- ここで, Z_k と Z_k の (0 でない) 定数倍は互いに同じ情報量をもつので, そのような定数倍の任意性をなくすため, ベクトル $\boldsymbol{a}_k := (a_{1k}, \ldots, a_{pk})^\top$ の長さが 1 となるようにする. すなわち,

$$\|\boldsymbol{a}_k\|^2 := \sum_{j=1}^{p} a_{jk}^2 = 1$$

と仮定

主成分分析

- k 番目の特徴量 Z_k を第 k **主成分 (得点) (principal component (score))** と呼ぶ
- Z_k の係数ベクトル $\boldsymbol{a}_k$ を第 k **主成分方向 (principal component direction)** または第 k **主成分負荷量 (principal component loading)** と呼ぶ
- 主成分分析の目的
  - 主成分 $Z_1, \ldots, Z_d$ が変数 $X_1, \ldots, X_p$ の情報を効率よく反映するように, 主成分負荷量 $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ を $X_1, \ldots, X_p$ たちの観測データから「うまく」決定する

主成分分析: R での実行

- R には主成分分析を実行するための関数として, 関数 `prcomp()` および関数 `princomp()` が用意されている
- 前者と後者には計算法に若干の違いがあり, 一般には前者の方が数値計算の観点からみると優れている (後者は S 言語との互換性を重視した実装となっている)
- 従って以下では関数 `prcomp()` を利用する
- 実行例 `pca-simulate.r`

主成分分析: 計算法

- まず $d = 1$ の場合を考える
- 組 $(X_1, \ldots, X_p)$ に対する n 個の観測データ $\{(x_{i1}, \ldots, x_{ip})\}_{i=1}^n$ が与えられているとする
- i 番目の観測データに対応する p 次元ベクトルを $\boldsymbol{x}_i := (x_{i1}, \ldots, x_{ip})^\top$ とする
- 目的 長さ 1 の p 次元ベクトル $\boldsymbol{a} = (a_1, \ldots, a_p)^\top$ を「うまく」選んで, 観測データ $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ のもつ情報を最大限保持するように 1 変量データ $\boldsymbol{a} \cdot \boldsymbol{x}_1, \ldots, \boldsymbol{a} \cdot \boldsymbol{x}_n$ を構成する
  - 各 $\boldsymbol{x}_i$ は p 次元空間内の点だとみなせるが, このとき $(\boldsymbol{a} \cdot \boldsymbol{x}_i)\boldsymbol{a}$ はベクトル $\boldsymbol{a}$ で張られる部分空間 (直線) への点 $\boldsymbol{x}_i$ の直交射影に一致する
  - すなわち, $\boldsymbol{a} \cdot \boldsymbol{x}_i$ はベクトル $\boldsymbol{x}_i$ の $\boldsymbol{a}$-方向成分であると解釈できる

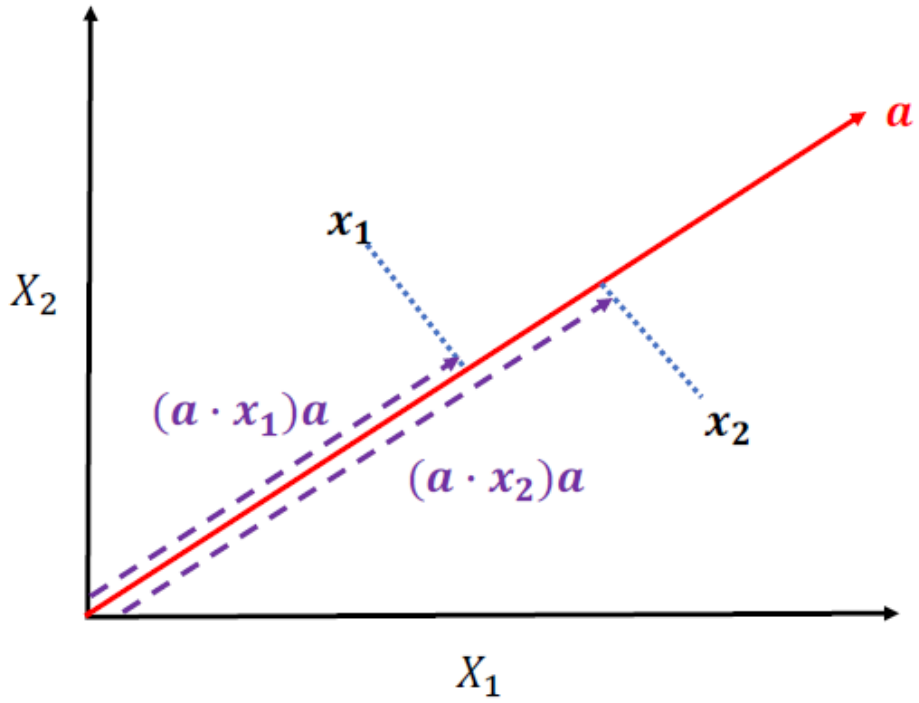

図 1: ベクトル $\boldsymbol{a}$ への観測データの直交射影 ($p = 2, n = 2$ の場合)

- このような幾何学的考察から, ベクトル $\boldsymbol{a}$ の適切な選び方の指針としては以下の考え方が自然:
  - 構成した特徴量がもとのデータのばらつきを最大限反映するように, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ たちのばらつきが最も大きい方向 $\boldsymbol{a}$ を選ぶ. すなわち,

$$\sum_{i=1}^{n} (\boldsymbol{a} \cdot \boldsymbol{x}_i - \boldsymbol{a} \cdot \bar{\boldsymbol{x}})^2$$

を最大化するように $\boldsymbol{a}$ を選ぶ. ここに,

$$\bar{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i$$

であり, 従って $\boldsymbol{a} \cdot \bar{\boldsymbol{x}}$ は $\boldsymbol{a} \cdot \boldsymbol{x}_1, \ldots, \boldsymbol{a} \cdot \boldsymbol{x}_n$ の平均に対応する.

主成分分析: 計算法

- 以上をまとめると, 関数

$$f(\boldsymbol{a}) = \sum_{i=1}^{n} (\boldsymbol{a} \cdot \boldsymbol{x}_i - \boldsymbol{a} \cdot \bar{\boldsymbol{x}})^2$$

を制約条件 $\|\boldsymbol{a}\| = 1$ の下で最大化するように, ベクトル $\boldsymbol{a}$ を選ぶのが方針となる

主成分分析: 計算法

- そのようなベクトル $\boldsymbol{a}$ は存在して次の性質をもつ (詳細な導出は配布資料参照. 時間があれば後で説明):
  - $f(\boldsymbol{a})$ は行列 $\boldsymbol{X}^\top \boldsymbol{X}$ の固有値であり, $\boldsymbol{a}$ はこの固有値に対する固有ベクトルである.

  ただし, $n \times p$ 行列 $\boldsymbol{X}$ を

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1^\top - \bar{\boldsymbol{x}}^\top \\ \vdots \\ \boldsymbol{x}_n^\top - \bar{\boldsymbol{x}}^\top \end{pmatrix} = \begin{pmatrix} x_{11} - \bar{x}_1 & \cdots & x_{1p} - \bar{x}_p \\ \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & \cdots & x_{np} - \bar{x}_p \end{pmatrix}$$

  で定義する ($\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$)

主成分分析: 計算法

- 従って, 求めるべき $\boldsymbol{a}$ は行列 $\boldsymbol{X}^\top \boldsymbol{X}$ の最大固有値に対する固有ベクトルで長さ 1 のものであり, このとき $f(\boldsymbol{a})$ はその最大固有値に一致する
- このようにして求まったベクトル $\boldsymbol{a}$ が**第 1 主成分負荷量**となる. また, 観測データから計算される**第 1 主成分**たちは

$$z_{i1} = a_1 x_{i1} + \cdots + a_p x_{ip} \qquad (i = 1, \ldots, n)$$

  となる
- 実行例 `pca-simulate.r`

主成分分析: 計算法

- 次に $d \geq 2$ の場合を考える
- まず記号を準備する
  - $\boldsymbol{X}^\top \boldsymbol{X}$ は非負定値対称行列だから, その固有値はすべて 0 以上の実数である
  - そこで, $\boldsymbol{X}^\top \boldsymbol{X}$ の固有値を (重複を許して) 降順に並べて $\lambda_1 \geq \cdots \geq \lambda_p (\geq 0)$ と書くことにする
  - さらに, 各 $j = 1, \ldots, p$ について $\boldsymbol{a}_j$ を λ_j に対する固有ベクトルとする
  - このとき, $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$ はそれぞれ長さ 1 かつ互いに直交するようにとることができる
  - すなわち,

$$\|\boldsymbol{a}_j\| = 1 \ (j = 1, \ldots, p), \qquad j \neq k \Rightarrow \boldsymbol{a}_j \cdot \boldsymbol{a}_k = 0 \tag{1}$$

  が成り立つと仮定してよい

主成分分析: 計算法

- 1 つめの特徴量は前節で構成した第 1 主成分を用いる
- 前節の議論より, ベクトル $\boldsymbol{a}_1$ は第 1 主成分方向に対応する
- 第 1 主成分方向に関してデータが有する情報はベクトル $(\boldsymbol{a}_1 \cdot \boldsymbol{x}_i)\boldsymbol{a}_1$ $(i = 1, \ldots, n)$ にすべて縮約されているので, 第 1 主成分 $\boldsymbol{a}_1 \cdot \boldsymbol{x}_i$ $(i = 1, \ldots, n)$ にすべて含まれている
- 従って, 2 つめの特徴量を構成する指針として, 観測データから第 1 主成分方向の成分を取り除いたデータ

$$\tilde{\boldsymbol{x}}_i := \boldsymbol{x}_i - (\boldsymbol{a}_1 \cdot \boldsymbol{x}_i)\boldsymbol{a}_1 \qquad (i = 1, \ldots, n)$$

に対して, 前節と同様の考え方でこれらのデータたちのばらつきが最も大きい方向 $\boldsymbol{a}$ を求めて, 特徴量 $\boldsymbol{a} \cdot \boldsymbol{x}_i$ $(i = 1, \ldots, n)$ を構成するのが自然である

主成分分析: 計算法

- すなわち,

$$\sum_{i=1}^{n}(\boldsymbol{a}\cdot\tilde{\boldsymbol{x}}_i-\boldsymbol{a}\cdot\bar{\tilde{\boldsymbol{x}}})^2 \qquad \text{ただし} \quad \bar{\tilde{\boldsymbol{x}}}:=\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_i$$

  を制約条件 $\|\boldsymbol{a}\|=1$ の下で最大化するような $\boldsymbol{a}$ を選べばよい
- $\Rightarrow$ $\boldsymbol{X}^\top\boldsymbol{X}$ の 2 番目に大きい固有値 λ_2 に対応する固有ベクトル $\boldsymbol{a}_2$ がそのようなベクトル $\boldsymbol{a}$ を与えることを示すことができる
  - 詳細な計算は配布資料参照. 時間があれば後で説明
- 同様の議論を繰り返すことによって, 第 k 主成分負荷量は $\boldsymbol{X}^\top\boldsymbol{X}$ の k 番目に大きい固有値 λ_k に対応する固有ベクトル $\boldsymbol{a}_k$ に取ればよいことがわかる
- 実行例 `pca.r`

分析の評価: 寄与率

- 構成した主成分が元のデータがもっていた情報をどの程度保持しているかを評価する方法の 1 つとして, 回帰分析の場合と同様に, その主成分のばらつき (分散) がもとのデータのばらつき (分散) をどの程度説明できているか評価する方法がある
- 第 k 主成分に対しては, これは

$$\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{a}_k \cdot \boldsymbol{x}_i - \boldsymbol{a}_k \cdot \bar{\boldsymbol{x}})^2}{\frac{1}{n-1}\sum_{i=1}^{n}\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|^2}$$

を計算することに相当する

分析の評価: 寄与率

- この量を第 k 主成分の**寄与率 (proportion of variance)** と呼ぶ
- 計算すると,

$$\sum_{i=1}^{n}(\boldsymbol{a}_k \cdot \boldsymbol{x}_i - \boldsymbol{a}_k \cdot \bar{\boldsymbol{x}})^2 = \lambda_k, \quad \sum_{i=1}^{n}\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|^2 = \sum_{l=1}^{p}\lambda_l$$

  が成り立つことがわかる (詳細な計算は配布資料参照. 時間があれば後で説明)
- 以上より, 第 k 主成分の寄与率は,

$$\frac{\lambda_k}{\sum_{l=1}^{p}\lambda_l}$$

  で計算される

分析の評価: 寄与率

- 第 1 主成分から第 k 主成分までを特徴量として用いた際に説明できるデータのばらつきの割合は第 k 主成分までの**累積寄与率 (cumulative proportion)** と呼ばれ, 第 1 主成分の寄与率から第 k 主成分の寄与率までの総和として計算される:

$$\frac{\sum_{l=1}^{k} \lambda_l}{\sum_{l=1}^{p} \lambda_l}.$$

- 累積寄与率はいくつの主成分を用いるべきかの基準として用いられる. 一般に, 累積寄与率が 80%程度の主成分を使って分析を行うことが多い
- R では, 寄与率および累積寄与率は, 関数 `prcomp()` のアウトプットに関数 `summary()` を適用することで計算できる
- 実行例 `pca-summary.r`

分析の評価: バイプロット

- 関連がある 2 枚の散布図を 1 つの画面に表示する散布図を**バイプロット (biplot)** という
- 主成分分析では, 得られた主成分の意味を解釈するために, 主成分方向の散布図と主成分の散布図を対応づけて分析を進める場合が多い
- より具体的には, 2 つの主成分方向 $\boldsymbol{a}_k = (a_{1k}, \ldots, a_{pk})^\top$ と $\boldsymbol{a}_l = (a_{1l}, \ldots, a_{pl})^\top$ に着目する
  - 主成分方向の散布図とは点 $\{(a_{jk}, a_{jl})\}_{j=1}^p$ の散布図であり, 各変数が着目した主成分方向にどれだけの変動成分をもつかを図示している (主成分分析ではこれらの点を対応するベクトルで描画することが多い)
  - 一方で, 主成分の散布図とは点 $\{(\boldsymbol{a}_k \cdot (\boldsymbol{x}_i - \bar{\boldsymbol{x}}, \boldsymbol{a}_l \cdot (\boldsymbol{x}_i - \bar{\boldsymbol{x}})\}_{i=1}^n$ の散布図であり, 各サンプルが着目した主成分方向にどれだけの変動成分をもつかを図示している

分析の評価: バイプロット

- R では, 関数 `prcomp()` のアウトプットに関数 `biplot()` を適用することでバイプロットを描画できる
- 実行例 `biplot2.r`

主成分の計算 ($d = 1$ の場合)

- 第 1 主成分負荷量は, 関数

$$f(\boldsymbol{a}) = \sum_{i=1}^{n} (\boldsymbol{a} \cdot \boldsymbol{x}_i - \boldsymbol{a} \cdot \bar{\boldsymbol{x}})^2$$

  を制約条件 $\|\boldsymbol{a}\| = 1$ の下で最大化するようなベクトル $\boldsymbol{a}$ として選ぶのであった
- $f(\boldsymbol{a})$ は連続関数であり, また集合 $\{\boldsymbol{a} \in \mathbb{R}^p : \|\boldsymbol{a}\| = 1\}$ はコンパクト (有界閉集合) であるから, この最大化問題は解をもつ

- 更に, Lagrange の乗数法から, 求めるべき解は, Lagrange 関数

$$L(\boldsymbol{a}, \lambda) = f(\boldsymbol{a}) + \lambda(1 - \|\boldsymbol{a}\|^2)$$

の勾配を 0 にするベクトルである

- いま,

$$f(\boldsymbol{a}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{p} a_j (x_{ij} - \bar{x}_j) \right)^2$$

と書けるから, 各 $j = 1, \ldots, p$ について

$$\begin{aligned} \frac{\partial L}{\partial a_j}(\boldsymbol{a}, \lambda) &= 2 \sum_{i=1}^{n} \left(\sum_{k=1}^{p} a_k (x_{ik} - \bar{x}_k) \right) (x_{ij} - \bar{x}_j) - 2\lambda a_j \\ &= 2 \sum_{k=1}^{p} \left(\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \right) a_k - 2\lambda a_j \qquad (2) \end{aligned}$$

が成り立つ

- 従って, (2) の右辺第 1 項は p 次元ベクトル $\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{a}$ の第 j 成分に等しいことがわかる
- 以上より, 求めるべき $\boldsymbol{a}$ は, 方程式

$$\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{a} = \lambda \boldsymbol{a} \tag{3}$$

の解となることがわかる
- 特に, λ は p 次正方行列 $\boldsymbol{X}^\top \boldsymbol{X}$ の固有値であり, $\boldsymbol{a}$ は λ に対する固有ベクトルとなる
- また, (3) の両辺に左から $\boldsymbol{a}^\top$ をかけると,

$$\boldsymbol{a}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{a} = \lambda \boldsymbol{a}^\top \boldsymbol{a} = \lambda \|\boldsymbol{a}\|^2 = \lambda$$

を得る

主成分の計算 ($d = 1$ の場合)

- ここで,

$$\boldsymbol{X}\boldsymbol{a} = ((\boldsymbol{x}_1 - \bar{\boldsymbol{x}}) \cdot \boldsymbol{a}, \ldots, (\boldsymbol{x}_n - \bar{\boldsymbol{x}}) \cdot \boldsymbol{a})^\top$$

  であることに注意すれば,

$$\boldsymbol{a}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{a} = (\boldsymbol{X}\boldsymbol{a})^\top \boldsymbol{X}\boldsymbol{a} = \|\boldsymbol{X}\boldsymbol{a}\|^2 = f(\boldsymbol{a})$$

  を得る
- このことから, $f(\boldsymbol{a})$ は行列 $\boldsymbol{X}^\top \boldsymbol{X}$ の固有値となる
- 要約すると, 次のようになる:
  - $f(\boldsymbol{a})$ は行列 $\boldsymbol{X}^\top \boldsymbol{X}$ の固有値であり, $\boldsymbol{a}$ はこの固有値に対する固有ベクトルである.

主成分の計算 ($d = 2$ の場合)

- 第 2 主成分負荷量は, 関数

$$\sum_{i=1}^{n} |\tilde{\boldsymbol{x}}_i - \bar{\boldsymbol{x}} - (\boldsymbol{a}_1 \cdot \bar{\boldsymbol{x}})\boldsymbol{a}_1|^2$$

を制約条件 $\|\boldsymbol{a}\| = 1$ の下で最大化するようなベクトル $\boldsymbol{a}$ として選ぶのであった

- 前と同様に, 行列 $\boldsymbol{X}_{(-1)}$ を

$$\boldsymbol{X}_{(-1)} = \begin{pmatrix} \tilde{\boldsymbol{x}}_1^\top - \bar{\boldsymbol{x}}^\top - (\boldsymbol{a}_1 \cdot \bar{\boldsymbol{x}})\boldsymbol{a}_1^\top \\ \vdots \\ \tilde{\boldsymbol{x}}_n^\top - \bar{\boldsymbol{x}}^\top - (\boldsymbol{a}_1 \cdot \bar{\boldsymbol{x}})\boldsymbol{a}_1^\top \end{pmatrix}$$

で定義すれば, $\boldsymbol{a}$ は行列 $\boldsymbol{X}_{(-1)}^\top \boldsymbol{X}_{(-1)}$ の最大固有値に対する固有ベクトルとなることがわかる

主成分の計算 ($d = 2$ の場合)

- ここで, 行列 $\boldsymbol{X}_{(-1)}$ は以下のように書けることに注意する (E_p は p 次単位行列):

$$\boldsymbol{X}_{(-1)} = \boldsymbol{X} - \boldsymbol{X}\boldsymbol{a}_1\boldsymbol{a}_1^\top = \boldsymbol{X}(E_p - \boldsymbol{a}_1\boldsymbol{a}_1^\top).$$

- 従って,

$$\boldsymbol{X}_{(-1)}^\top \boldsymbol{X}_{(-1)} = (E_p - \boldsymbol{a}_1\boldsymbol{a}_1^\top)\boldsymbol{X}^\top\boldsymbol{X}(E_p - \boldsymbol{a}_1\boldsymbol{a}_1^\top)$$

であるから, 条件 (1) より

$$\boldsymbol{X}_{(-1)}^\top \boldsymbol{X}_{(-1)}\boldsymbol{a}_1 = \boldsymbol{0}, \qquad \boldsymbol{X}_{(-1)}^\top \boldsymbol{X}_{(-1)}\boldsymbol{a}_j = \lambda_j \boldsymbol{a}_j \ (j = 2, \ldots, p)$$

が成り立つ

主成分の計算 ($d = 2$ の場合)

- これは $0, \lambda_2, \ldots, \lambda_p$ が $\boldsymbol{X}_{(-1)}^\top \boldsymbol{X}_{(-1)}$ の固有値であり, $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$ がそれぞれに対する固有ベクトルであることを意味する
- 従って, $\boldsymbol{X}_{(-1)}^\top \boldsymbol{X}_{(-1)}$ の最大固有値は λ_2 であり, 求めるべきベクトルは $\boldsymbol{a}_2$ である

寄与率の計算

- $\boldsymbol{a}_k \cdot \boldsymbol{x}_i - \boldsymbol{a}_k \cdot \bar{\boldsymbol{x}}$ が (列) ベクトル $\boldsymbol{X}\boldsymbol{a}$ の第 i 成分に対応することと, 列ベクトル $\boldsymbol{v}$ に対して $\|\boldsymbol{v}\|^2 = \boldsymbol{v}^\top \boldsymbol{v}$ が成り立つことに注意すれば,

$$\sum_{i=1}^{n} (\boldsymbol{a}_k \cdot \boldsymbol{x}_i - \boldsymbol{a}_k \cdot \bar{\boldsymbol{x}})^2 = \|\boldsymbol{X}\boldsymbol{a}_k\|^2 = \boldsymbol{a}_k^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{a}_k$$

$$= \lambda_k \boldsymbol{a}_k^\top \boldsymbol{a}_k = \lambda_k \|\boldsymbol{a}_k\|^2 = \lambda_k$$

を得る

寄与率の計算

- さらに, $\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|^2$ が行列 $\boldsymbol{X}\boldsymbol{X}^\top$ の第 i 対角成分であることと, 直交行列 A を $A = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p)$ で定義すれば,

$$A^\top \boldsymbol{X}\boldsymbol{X} A = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix}$$

と対角化されることに注意すれば, 以下が成り立つ:

寄与率の計算

$$\sum_{i=1}^{n} \|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|^2 = \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^\top) = \mathrm{tr}(\boldsymbol{X}^\top\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{X}^\top\boldsymbol{X}E_p)$$

$$= \mathrm{tr}(\boldsymbol{X}^\top\boldsymbol{X}AA^\top) = \mathrm{tr}(A^\top\boldsymbol{X}^\top\boldsymbol{X}A) = \sum_{l=1}^{p} \lambda_l$$

- ここで, 積 BC および CB が定義されるような行列 B, C に対して $\mathrm{tr}(BC) = \mathrm{tr}(CB)$ が成り立つことを用いた