More Linear Algebra

Edps/Soc 584, Psych 594

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Spring 2017

Overview

- Eigensystems: decomposition of square matrix
- Singular Value Decompositions: decomposition of rectangular matrix
- Maximization:

Reading: Johnson & Wichern pages 60–66, 73–75, 77–81

Eigensystems

Let $\mathbf{A}$ be a $p \times p$ square matrix, then the scalars $\lambda_1, \lambda_2, \ldots, \lambda_p$ that satisfy the polynomial equation

$$|\mathbf{A} - \lambda \mathbf{I}| = 0$$

are called eigenvalues (or "characteristic roots") of matrix $\mathbf{A}$. The equation $|\mathbf{A} - \lambda \mathbf{I}| = 0$ is called the "characteristic equation."

Example:

$$\begin{aligned}
\mathbf{A} &= \begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix} \\
|\mathbf{A} - \lambda \mathbf{I}| &= \begin{vmatrix} (1-\lambda) & -5 \\ -5 & (1-\lambda) \end{vmatrix} = 0 \\
& (1-\lambda)^2 - (-5)(-5)) = 0 \\
& \lambda^2 - 2\lambda - 24 = 0 \\
& (\lambda - 6)(\lambda + 4) = 0 \longrightarrow \lambda_1 = 6 \text{ and } \lambda_2 = -4
\end{aligned}$$

Quadratic Formula: $ax^2 + bx + c = 0 \rightarrow (-b \pm \sqrt{b^2 - 4ac})/(2a)$

Eigenvectors

A square matrix $\mathbf{A}$ is said to have eigenvalues λ with a corresponding eigenvector $\mathbf{x} \neq \mathbf{0}$ if

$$\mathbf{Ax} = \lambda\mathbf{x} \qquad \text{or} \qquad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$$

- We usually normalize $\mathbf{x}$ so that it has length $= 1$.

$$\mathbf{e} = \frac{\mathbf{x}}{L_{\mathbf{x}}} = \frac{\mathbf{x}}{\sqrt{\mathbf{x}'\mathbf{x}}} \qquad \text{so} \qquad \mathbf{e}'\mathbf{e} = 1$$

- $\mathbf{e}$ is also an eigenvector of $\mathbf{A}$ because

$$\begin{aligned} \mathbf{Ae} &= \lambda\mathbf{e} \\ \mathbf{A}(L_{\mathbf{x}}\mathbf{e}) &= \lambda(L_{\mathbf{x}}\mathbf{e}) \\ \mathbf{Ax} &= \lambda\mathbf{x} \end{aligned}$$

- Any multiple of $\mathbf{x}$ is an eigenvector associated with λ.

  All that matters is the direction and not the length of $\mathbf{x}$.

Eigenvectors continued

Example:

$$\mathbf{A} = \begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\begin{aligned} x_1 - 5x_2 &= \lambda x_1 \\ -5x_1 + x_2 &= \lambda x_2 \end{aligned}$$

So we have 2 equations and 3 unknowns (x_1, x_2 and λ).

Set $\lambda = 6$, now there are 2 equations with 2 unknowns:

$$\begin{aligned} x_1 - 5x_2 &= 6x_1 \\ -5x_1 + x_2 &= 6x_2 \longrightarrow \mathbf{x} = \mathbf{e} = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \end{aligned}$$

Symmetric Matrix

Now $\mathbf{A}$ is $(p \times p)$ symmetric

Let $\mathbf{A}_{(p\times p)}$ be a symmetric matrix. Then $\mathbf{A}$ has p pairs of eigenvalues and eigenvectors

$$\lambda_1, \mathbf{e}_1; \quad \lambda_2, \mathbf{e}_2; \quad \cdots \quad ; \quad \lambda_p, \mathbf{e}_p.$$

- The eigenvectors are chosen to have length$= 1$:

$$\mathbf{e}_1'\mathbf{e}_1 = \mathbf{e}_2'\mathbf{e}_2 = \cdots = \mathbf{e}_p'\mathbf{e}_p = 1.$$

- The eigenvectors are also chosen to be mutually orthogonal (perpendicular):

$$\mathbf{e}_i \perp \mathbf{e}_k \qquad \text{that is} \qquad \mathbf{e}_i'\mathbf{e}_k = 0 \text{ for all } i \neq k$$

- The eigenvectors are all unique if no 2 eigenvalues are equal.
- Typically the eigenvalues are ordered from largest to smallest

Little Example continued

$$\mathbf{A} = \begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix}$$

and

$$\begin{array}{rr} \lambda_1 = & 6 \\ \lambda_2 = & -4 \end{array} \qquad \mathbf{e}_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \qquad \mathbf{e}_2 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

Note that $\mathbf{e}_1'\mathbf{e}_2 = 0$ and $L_{\mathbf{e}_1} = L_{\mathbf{e}_2} = 1$.

Spectral Decomposition of A

The "Spectral Decomposition" of **A** where $A_{(p\times p)}$ symmetric.

$$\mathbf{A} = \lambda_1 \underbrace{\mathbf{e}_1\mathbf{e}_1'}_{p\times p} + \lambda_2 \underbrace{\mathbf{e}_2\mathbf{e}_2'}_{p\times p} + \cdots + \lambda_p \underbrace{\mathbf{e}_k\mathbf{e}_k'}_{p\times p}$$

If **A** is also "positive definite", then $k = p$.

Matrix **A** is decomposed into p $(p \times p)$ component matrices.
where $\mathbf{e}_i'\mathbf{e}_i = 1$ for all i, and $\mathbf{e}_i'\mathbf{e}_j = 0$ for all $i \neq j$.

$$\mathbf{A} = \begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix} \quad \begin{matrix} \lambda_1 = & 6 \\ \lambda_2 = & -4 \end{matrix} \quad \mathbf{e}_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \quad \mathbf{e}_2 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

$$\begin{aligned} \lambda_1\mathbf{e}_1\mathbf{e}_1' + \lambda_2\mathbf{e}_2\mathbf{e}_2' &= 6\begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix} - 4\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix} = \mathbf{A} \end{aligned}$$

A Bigger Example

$$\mathbf{A} = \begin{pmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{pmatrix} \quad \lambda_1 = \lambda_2 = 9, \lambda_3 = 18$$

$$\mathbf{e}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \quad \mathbf{e}_2 = \begin{pmatrix} \frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \\ -\frac{4}{\sqrt{18}} \end{pmatrix} \quad \mathbf{e}_3 = \begin{pmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{pmatrix}$$

Note that since $\lambda_1 = \lambda_2$ the labeling of $\mathbf{e}_1$ and $\mathbf{e}_2$ is arbitrary.

- The lengths: $\mathbf{e}_1'\mathbf{e}_1 = \mathbf{e}_2'\mathbf{e}_2 = \mathbf{e}_3'\mathbf{e}_3 = 1$.
- Orthogonality: $\mathbf{e}_1'\mathbf{e}_2 = \mathbf{e}_1'\mathbf{e}_3 = \mathbf{e}_2'\mathbf{e}_3 = 0$.
- Decomposition:

$$\mathbf{A} = 9\mathbf{e}_1\mathbf{e}_1' + 9\mathbf{e}_2\mathbf{e}_2' + 18\mathbf{e}_3\mathbf{e}_3'$$

Decomposition of (3×3)

$$\mathbf{A} = 9\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right) + 9\begin{pmatrix} \frac{1}{\sqrt{18}} \\ \frac{-1}{\sqrt{18}} \\ \frac{-4}{\sqrt{18}} \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{18}} & \frac{-1}{\sqrt{18}} & \frac{-4}{\sqrt{18}} \end{pmatrix}$$

$$+18\begin{pmatrix} \frac{2}{3} \\ \frac{-2}{3} \\ \frac{1}{3} \end{pmatrix}\begin{pmatrix} \frac{2}{3} & \frac{-2}{3} & \frac{1}{3} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{9}{2} & \frac{9}{2} & 0 \\ \frac{9}{2} & \frac{9}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \frac{9}{18} & \frac{-9}{18} & \frac{-36}{18} \\ \frac{-9}{18} & \frac{9}{18} & \frac{36}{18} \\ \frac{-36}{18} & \frac{36}{18} & \frac{144}{18} \end{pmatrix} + \begin{pmatrix} \frac{72}{9} & \frac{-72}{9} & \frac{36}{9} \\ \frac{-72}{9} & \frac{72}{9} & \frac{-36}{9} \\ \frac{54}{9} & \frac{36}{9} & \frac{18}{9} \end{pmatrix}$$

$$= \frac{1}{18}\begin{pmatrix} 234 & -72 & 36 \\ -72 & 234 & -36 \\ 36 & -36 & 180 \end{pmatrix} = \begin{pmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{pmatrix}$$

Recall: Quadratic Form is defined as

$$\mathbf{x}'\mathbf{A}\mathbf{x} \quad \text{for} \quad \mathbf{x}_p \quad \text{and } \mathbf{A}_{p\times p} \quad \text{symmetric}$$

The terms of $\mathbf{x}'\mathbf{A}\mathbf{x}$ are squares of x_i (i.e., x_i^2) and cross-products of x_i and x_k (i.e., $x_i x_k$):

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{p}\sum_{k=1}^{p} a_{ik}x_i x_k$$

e.g.,

$$\begin{aligned}
&(x_1, x_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= ((a_{11}x_1 + a_{21}x_2), (a_{12}x_1 + a_{22}x_2)) \times \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= a_{11}x_1^2 + a_{21}x_1x_2 + a_{12}x_1x_2 + a_{22}x_2^2 = \sum_{i=1}^{2}\sum_{k=1}^{2} a_{ik}x_i x_k
\end{aligned}$$

Eigenvalues and Definiteness

- IF $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x}$, matrix $\mathbf{A}$ is positive definite.
- IF $\mathbf{x}'\mathbf{A}'\mathbf{x} \geq 0$ for all $\mathbf{x}$, matrix $\mathbf{A}$ is non-negative definite.

Important:
All eigenvalues of $\mathbf{A} > 0 \Leftrightarrow \mathbf{A}$ is positive definite.
All eigenvalues of $\mathbf{A} \geq 0 \Leftrightarrow \mathbf{A}$ is non-negative definite

Implication: If $\mathbf{A}$ is positive definite, then the diagonal elements of $\mathbf{A}$ must be positive.

$$\text{If } \mathbf{x} = (0, \ldots, \underbrace{1}_{i^{th} position}, \ldots 0) \qquad \text{then } \mathbf{x}'\mathbf{A}\mathbf{x} = a_{ii}x_i^2 > 0$$

More on Spectral Decomposition

When $\mathbf{A}_{p\times p}$ symmetric and positive definite,
(i.e., diagonals of $\mathbf{A}$ are all > 0, and $\lambda_i > 0$ for all i).

We can write the spectral decomposition of $\mathbf{A}$ as the sum of the weighted vector products,

$$\mathbf{A}_{p\times p} = \sum_{i=1}^{p} \lambda_i \mathbf{e}_i \mathbf{e}_i'$$

In matrix form this is $\boxed{\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'}$ where

$$\boldsymbol{\Lambda}_{p\times p} = \text{diag}(\lambda_i) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

and $\quad \mathbf{P}_{p\times p} = (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_p).$

Showing that $A = P\Lambda P'$

$$
\begin{aligned}
A_{p\times p} &= \mathbf{P}_{p\times p}\mathbf{\Lambda}_{p\times p}\mathbf{P}'_{p\times p} \\
&= (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_p)\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}\begin{pmatrix} \mathbf{e}'_1 \\ \mathbf{e}'_2 \\ \vdots \\ \mathbf{e}'_p \end{pmatrix} \\
&= (\lambda_1\mathbf{e}_1, \lambda_2\mathbf{e}_2, \cdots, \lambda_p\mathbf{e}_p)\begin{pmatrix} \mathbf{e}'_1 \\ \mathbf{e}'_2 \\ \vdots \\ \mathbf{e}_p \end{pmatrix} \\
&= \sum_{i=1}^{p} \lambda_i\mathbf{e}_i\mathbf{e}'_i
\end{aligned}
$$

More about P

Since The lengths of $\mathbf{e}_i$ equal 1 (i.e., $\mathbf{e}_i'\mathbf{e}_i = 1$), and $\mathbf{e}_i$ and $\mathbf{e}_k$ are orthogonal for all $i \neq k$ (i.e., $\mathbf{e}_i'\mathbf{e}_k = 0$).

$$\begin{aligned}\mathbf{P}'\mathbf{P} &= \begin{pmatrix} \mathbf{e}_1' \\ \mathbf{e}_2' \\ \vdots \\ \mathbf{e}_p' \end{pmatrix} (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_p) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \\ &= \mathbf{I} = \mathbf{P}\mathbf{P}'\end{aligned}$$

P is an orthogonal matrix.

Rank r decompositions

If $\mathbf{A}$ is non-negative definite (semi-definite):

$$\begin{array}{lll} \lambda_i > 0 & \text{for} & i = 1, \ldots, r < p \\ \lambda_i = 0 & \text{for} & i = r+1, \ldots, p \end{array}$$

So

$$\mathbf{A}_{p\times p} = \mathbf{P}_{p\times r}\mathbf{\Lambda}_{r\times r}\mathbf{P}'_{r\times p}.$$

If $\mathbf{A}$ is positive or positive semi-definite, we sometimes want to approximate $\mathbf{A}$ by a rank r decomposition, where $r <$ Rank of $\mathbf{A}$,

$$\mathbf{B} = \lambda_1\mathbf{e}_1\mathbf{e}_1' + \ldots + \lambda_r\mathbf{e}_r\mathbf{e}_r'$$

This decomposition minimized the loss function

$$\sum_{i=1}^{p}\sum_{k=1}^{p}(a_{ik} - b_{ik})^2 = \lambda_{r+1}^2 + \lambda_{r+2}^2 + \cdots \lambda_p^2$$

Inverse of A

If **A** is positive definite, the inverse of **A** equals

$$\mathbf{A}^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}'$$

where

$$\text{diag}\left(\frac{1}{\lambda_i}\right) = \begin{pmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\lambda_p \end{pmatrix}$$

Why:

$$\mathbf{A}\mathbf{A}^{-1} = (\mathbf{P}\mathbf{\Lambda}\ \underbrace{\mathbf{P}')\,(\mathbf{P}}_{\mathbf{I}}\ \mathbf{\Lambda}^{-1}\mathbf{P}'\ ) = \mathbf{P}\,\underbrace{\mathbf{\Lambda}\mathbf{\Lambda}^{-1}}_{\mathbf{I}}\mathbf{P}' = \mathbf{P}\mathbf{P}' = \mathbf{I}$$

What does $\mathbf{A}^{-1}\mathbf{A}$ equal?

Square Root Matrix

If $\mathbf{A}$ is symmetric, the Square Root Matrix of $\mathbf{A}$ is

$$\mathbf{A}^{1/2} = \sum_{i=1}^{p} \sqrt{\lambda_i}\; \mathbf{e}_i \mathbf{e}_i' = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}'$$

Common mistake: $\mathbf{A}^{1/2} = \{\sqrt{a_{ij}}\}$.

Properties of $\mathbf{A}^{1/2}$:

- $(\mathbf{A}^{1/2})' = \mathbf{A}^{1/2}$ …since $\mathbf{A}^{1/2}$ is symmetric.
- $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$
- $(\mathbf{A}^{1/2})^{-1} = \sum_{i=1}^{p}(1/\sqrt{\lambda_i})\mathbf{e}_i\mathbf{e}_i' = \mathbf{P}\boldsymbol{\Lambda}^{-1/2}\mathbf{P}' = \mathbf{A}^{-1/2}$
- $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{I}$
- $\mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1}$

Determinant, Trace and Eigenvalues

$$|\mathbf{A}| = \prod_{i=1}^{p} \lambda_i = \lambda_1 \lambda_2 \cdots \lambda_p.$$

Implication: A positive definite matrix has $|\mathbf{A}| > 0$, because $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$

$$\sum_{i=1}^{p} a_{ii} = \text{trace}(\mathbf{A}) = \sum_{i=1}^{p} \lambda_i$$

Now let's consider what's true for $\boldsymbol{\Sigma}$ and $\mathbf{S}$.

Numerical Example

We'll use the psychological test data from Rencher (2002) who got it from Beall (1945) to illustrate these properties

32 males and 32 females had measures on four psychological tests.

The tests were

x_1 = pictorial inconsistencies x_2 = paper form board
x_3 = tool recognition x_4 = vocabulary

$$\mathbf{S} = \begin{pmatrix} 10.387897 & 7.7926587 & 15.298115 & 5.3740079 \\ 7.7926587 & 16.657738 & 13.706845 & 6.1755952 \\ 15.298115 & 13.706845 & 57.057292 & 15.932044 \\ 5.3740079 & 6.1755952 & 15.932044 & 22.133929 \end{pmatrix}$$

Note that the total sample variance = trace($\mathbf{S}$) = 106.23686 and that the generalize sample variance = det($\mathbf{S}$) = 65980.199

Numerical Example continued

Eigenvalue of **S** are

$$\mathbf{\Lambda} = \begin{pmatrix} 72.717 & 0 & 0 & 0 \\ 0 & 16.111 & 0 & 0 \\ 0 & 0 & 13.114 & 0 \\ 0 & 0 & 0 & 4.295 \end{pmatrix}$$

and the eigenvectors are

$$\begin{aligned} \mathbf{P} &= \begin{pmatrix} 0.274 & -0.002 & 0.327 & 0.904 \\ 0.284 & 0.185 & 0.854 & -0.394 \\ 0.856 & -0.409 & -0.271 & -0.163 \\ 0.333 & 0.8936 & -0.300 & 0.009 \end{pmatrix} \\ &= (\mathbf{e}_1, \quad \mathbf{e}_2, \quad \mathbf{e}_3, \quad \mathbf{e}_4) \end{aligned}$$

Note that (for example)

$\mathbf{e}_1'\mathbf{e}_1 = (.274^2 + .284^2 + .856^2 + .333^2) = 1 = L^2_{\mathbf{e}_1} = L_{\mathbf{e}_1}.$

$\mathbf{e}_1'\mathbf{e}_2 = (.274(-.002) + .284(.185) + .856(-.409) + .333(.894)) = 0.$

Example: eigenvalues of S

Sum of eigenvalues:

$$\begin{aligned}\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 &= 72.717 + 16.111 + 13.114 + 4.295 \\ &= 106.237 \\ &= \text{trace}(\mathbf{S}) \\ &= \text{Total sample variance}\end{aligned}$$

Product of the eigenvalues:

$$\begin{aligned}\prod_{i=1}^{4} \lambda_i &= 72.717 \times 16.111 \times 13.114 \times 4.295 \\ &= 65986.76 \\ &= \det(\mathbf{S}) \\ &= GSV\end{aligned}$$

Properties of Covariance Matrices

$\boldsymbol{\Sigma}_{p\times p}$ & $\mathbf{S}_{p\times p}$ symmetric population and sample covariance matrices, respectively. Most of following holds true for both.

Eigenvalues and eigenvectors: $\mathbf{S}$ has p pairs of eigenvalues and eigenvectors

$$\lambda_1, \mathbf{e}_1; \quad \lambda_2, \mathbf{e}_2; \quad \cdots ; \quad \lambda_p, \mathbf{e}_p$$

- The λ_i's are the roots of the characteristic equation

$$|\mathbf{S} - \lambda \mathbf{I}| = 0$$

- Eigenvectors are the solutions of the equation

$$\mathbf{S}\mathbf{e}_i = \lambda_i \mathbf{e}_i$$

Properties of Covariance Matrices (continued)

- Since any multiple of $\mathbf{e}_i$ will solve the above equation, we (usually) set the length of $\mathbf{e}_i = 1$ (i.e., $L^2_{\mathbf{e}_i} = L_{\mathbf{e}_i} = \mathbf{e}_i'\mathbf{e}_i = 1$).
- Eigenvectors are orthogonal: $\mathbf{e}_i'\mathbf{e}_k = 0$ for all $i \neq k$.
- Convention to order eigenvalues: $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p$.
- Since $\mathbf{S}$ (& $\mathbf{\Sigma}$) are symmetric, eigenvalues are Real numbers.

More about Covariance Matrices

- Spectral Decomposition:

$$\mathbf{S} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p' = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$$

  - $\mathbf{P}_{p\times p} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p)$
  - $\boldsymbol{\Lambda}_{p\times p} = \text{diag}(\lambda_i)$.
  - $\mathbf{P}'\mathbf{P} = \{\mathbf{e}_i'\mathbf{e}_k\} = \mathbf{P}\mathbf{P}' = \mathbf{I}$, which implies that $\mathbf{P}' = \mathbf{P}^{-1}$.
- Implications for quadratic forms:
  - If $\mathbf{x}'\mathbf{S}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$, then $\mathbf{S}$ is positive definite and $\lambda_i > 0$ for all i.
  - If $\mathbf{x}'\mathbf{S}\mathbf{x} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$, then $\mathbf{S}$ is non-negative or positive semi-definite and $\lambda_i \geq 0$ for all i.
- The inverse of $\mathbf{S}$ (if $\mathbf{S}$ is non-singular, i.e., $\lambda_i > 0$ for all i) is

$$\mathbf{S}^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}' = \mathbf{P}\{diag\{1/\lambda_i\}\mathbf{P}'$$

Numerical Example & Spectral Decomposition

$$
\begin{aligned}
\mathbf{S} &= \mathbf{P\Lambda P'} \\
&= \begin{pmatrix} 0.274 & -0.002 & 0.327 & 0.904 \\ 0.284 & 0.185 & 0.854 & -0.394 \\ 0.856 & -0.409 & -0.271 & -0.163 \\ 0.333 & 0.8936 & -0.300 & 0.009 \end{pmatrix} \begin{pmatrix} 72.717 & 0 & 0 \\ 0 & 16.111 & 0 \\ 0 & 0 & 13.114 \\ 0 & 0 & 0 & 4.29 \end{pmatrix} \\
&\quad \times \begin{pmatrix} 0.274 & 0.284 & 0.856 & 0.333 \\ -0.002 & 0.185 & -0.409 & 0.8936 \\ 0.327 & 0.854 & -0.271 & -0.300 \\ 0.904 & -0.394 & -0.163 & 0.009 \end{pmatrix}
\end{aligned}
$$

Do SAS/IML Demonstration of this and $\mathbf{S}^{-1} = \mathbf{P\Lambda}^{-1}\mathbf{P}'$.

and Even More about Covariance Matrices

- If $\{\lambda_i, \mathbf{e}_i; i = 1, \ldots, p\}$ for $\mathbf{\Sigma}$ and $\mathbf{\Sigma}$ is non-singular, then $\{1/\lambda_i, \mathbf{e}_i; i = 1, \ldots p\}$ for $\mathbf{\Sigma}^{-1}$

  That is, $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^{-1}$ have the same eigenvectors and their eigenvalues are the inverses of each other.
- $|\mathbf{S}| = \lambda_1\lambda_2 \cdots \lambda_p = \prod_{i=1}^p \lambda_i$.
  This is the generalized sample variance (GSV).
- $\sum_{i=1}^p s_{ii} = \text{trace}(\mathbf{S}) = \text{tr}(\mathbf{S}) = \sum_{i=1}^p \lambda_i$.
  This is the Total Sample Variance.
- If λ_p, the smallest eigenvalue, is greater than 0, then $|\mathbf{S}| > 0$.
- If $\mathbf{S}$ is singular, then there is at least 1 or more eigenvalues equal to 0.

The Rank of S (and Σ)

Definition of rank:

$$\text{The Rank of } \mathbf{S} = \text{the number of linearly independent rows (columns)}$$
$$= \text{the number of non-zero eigenvalues}$$

If $\mathbf{S}_{p\times p}$ is of Full Rank (i.e., rank $= p$), then

- $\lambda_p > 0$
- $\mathbf{S}$ is positive definite
- $|\mathbf{S}| > 0$
- $\mathbf{S}^{-1}$ exists
- $\mathbf{S}$ is non-singular
- definition: p linearly independent rows/columns

Singular Value Decomposition

Given matrix $\mathbf{A}_{n\times p}$, the Singular Value Decomposition (SVD) of $\mathbf{A}$ is

$$\mathbf{A}_{n\times p} = \mathbf{P}_{n\times r}\boldsymbol{\Delta}_{r\times r}\mathbf{Q}'_{r\times p}$$

where

- The r columns of $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots \mathbf{p}_r)$ are orthogonal: $\mathbf{p}_i'\mathbf{p}_i = 1$ and $\mathbf{p}_i'\mathbf{p}_k = 0$ for $i \neq k$; that is, $\mathbf{P}'\mathbf{P} = \mathbf{I}_r$.
- The r columns of $\mathbf{Q} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_r)$ are orthogonal: $\mathbf{q}_i'\mathbf{q}_i = 1$ and $\mathbf{q}_i'\mathbf{q}_k = 0$ for $i \neq k$; that is, $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_r$.
- $\boldsymbol{\Delta}$ is a diagonal matrix with ordered positive values

$$\delta_1 \geq \delta_2 \geq \cdots \geq \delta_r$$

- r is the rank of $\mathbf{A}$, which must be $r \leq \min(n, p)$.

Singular Value Decomposition (continued)

$$\mathbf{A}_{n\times p} = \mathbf{P}_{n\times r}\mathbf{\Delta}_{r\times r}\mathbf{Q}'_{r\times p}$$

Terminology:

- $\mathbf{P}$ are the "left singular vectors"
- $\mathbf{Q}$ are the "right singular vectors"
- The elements of $\mathbf{\Delta}$ are the "singular values"

Relationship between Eigensystems and SVD

To show this let $\mathbf{X}_{n\times p}$ which has rank p, and

$$\mathbf{X}_{n\times p} = \mathbf{P}_{n\times p}\mathbf{\Delta}_{p\times p}\mathbf{Q}'_{p\times p}.$$

The product $\mathbf{X}'_{p\times n}\mathbf{X}_{n\times p}$ is a square and symmetric matrix.

$$\begin{aligned}
\mathbf{X}'_{p\times n}\mathbf{X}_{n\times p} &= (\mathbf{P}_{n\times p}\mathbf{\Delta}_{p\times p}\mathbf{Q}'_{p\times p})'(\mathbf{P}_{n\times p}\mathbf{\Delta}_{p\times p}\mathbf{Q}'_{p\times p}) \\
&= (\,\mathbf{Q}_{p\times p}\mathbf{\Delta}_{p\times p}\underbrace{\mathbf{P}'_{p\times n}\,)(\,\mathbf{P}_{n\times p}}_{\mathbf{I}}\mathbf{\Delta}_{p\times p}\mathbf{Q}'_{p\times p}\,) \\
&= \mathbf{Q}_{p\times p}\mathbf{\Delta}_{p\times p}\mathbf{\Delta}_{p\times p}\mathbf{Q}'_{p\times p} \\
&= \underbrace{\mathbf{Q}_{p\times p}}_{\text{vectors}}\underbrace{\mathbf{\Delta}^2_{p\times p}}_{\text{values}}\underbrace{\mathbf{Q}'_{p\times p}}_{\text{vectors}}
\end{aligned}$$

If $\mathbf{A}$ (e.g., $\mathbf{X}'_{p\times n}\mathbf{X}_{n\times p}$) is square and symmetric, then SVD gives the same as eigenvector/value decomposition.

Lower Rank SVD

Sometimes we want to summarize or approximate the basic structure of a matrix.

In particular, let $\mathbf{A}_{n\times p} = \mathbf{P}_{n\times r}\mathbf{\Delta}_{r\times r}\mathbf{Q}'_{r\times p}$, then

$$\mathbf{B}_{n\times p} = \mathbf{P}_{n\times r^*}\mathbf{\Delta}_{r^*\times r^*}\mathbf{Q}'_{r^*\times p}$$

where $r^* < r$ (note: $r =$rank of matrix $\mathbf{A}$).

This Lower Rank Decomposition minimizes the loss function

$$\sum_{j=1}^{n}\sum_{i=1}^{p}(a_{ji} - b_{ji})^2 = \delta^2_{r^*+1} + \cdots + \delta^2_{r}$$

This result of the least squared approximation of one matrix by another of lower rank is known as the Eckart-Young theorem. See Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 1*, 211–218.

So What can I do with SVD?

- Biplot: Lower rank representation of a data matrix.
- Correspondence Analysis: Lower rank representation of the relationship between two categorical variables.
- Multiple Correspondence Analysis: Lower rank representations of the relationship between multiple categorical variables.
- Multidimensional Scaling
- Reduce the number of parameters in a complex model.
- and Many other scaling and data analytic methods.

We'll examine what a Biplot can give us. . .

Consider the psychological test data: The rank of the data matrix is 4, so

$$\mathbf{X}_c = (\mathbf{X} - \bar{\mathbf{x}}) = \mathbf{P}_{64\times 4}\boldsymbol{\Delta}_{4\times 4}\mathbf{Q}'_{4\times 4} = \underbrace{(\mathbf{P}\boldsymbol{\Delta})}_{cases}\ \underbrace{\mathbf{Q}'}_{variables}$$

Biplot Example: Singular Values

				Cumulative	
i	δ_i	δ_i^2	percent	sum	percent
1	67.685	4581.197	68.45	4581.197	68.45
2	31.859	1014.964	15.16	5896.161	83.61
3	28.744	826.204	12.35	6722.365	95.96
4	16.449	270.557	4.04	6692.922	100.00

where percent $= (\delta_i^2/6692.922) \times 100\%$, sum $= \sum_{k=1}^{i} \delta_i^2$, and cumulative percent $= (\sum_{k=1}^{i} \delta_k^2/6602.922) \times 100\%$.

If we take a rank 2 decomposition,

$\mathbf{B} = \sum_{l=1}^{2} \delta_l \mathbf{p}_l \mathbf{q}_l' = \{\delta_1 p_{j1} q_{i1} + \delta_2 p_{j2} q_{i2}\} = \{b_{ji}\}$
and the value of the loss function is

$$\text{loss} = \sum_{j=1}^{n} \sum_{i=1}^{4} (x_{c,ji} - b_{ji})^2 = 826.204 + 270.557 = 1096.761$$

Only losing $(1096/6692) \times 100\% = 16.39\%$ of the information in the data matrix (loosely speaking).

Biplot Example: Singular Vectors

Left Singular Vectors: $\mathbf{P}_{64\times 4}$

$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$	$\mathbf{p}_4$
−0.002	−0.248	0.139	−0.029
0.157	−0.026	−0.098	0.056
0.092	−0.077	−0.091	−0.001
−0.198	−0.041	0.079	0.120
0.111	0.118	0.031	0.233
0.073	−0.054	0.166	−0.140
0.045	−0.073	−0.081	0.051
−0.046	−0.068	−0.304	0.173
0.042	−0.299	−0.257	0.098
etc.			

Right Singular Vectors: $\mathbf{Q}_{4\times 4}$

$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_3$	$\mathbf{q}_4$
0.274	−0.001	0.326	0.904
0.284	0.184	0.854	−0.394
0.856	−0.408	−0.271	−0.162
0.333	0.893	−0.300	0.009

Biplot: Representing Cases

First let's look at the rank 2 solution/approximation

$$\underbrace{\tilde{\mathbf{X}}_c}_{(64\times 4)} = \underbrace{\mathbf{P}}_{(64\times 2)}\underbrace{\mathbf{\Delta}}_{(2\times 2)}\underbrace{\mathbf{Q}'}_{(2\times 4)}$$

For our rank 2 solution, to represent subjects or cases, we'll plot the rows of the product $\mathbf{P}_{64\times 2}\mathbf{\Delta}_{2\times 2}$ as points in a 2-dimensional space.

Let q_{il} = the value in the i^{th} row of $\mathbf{q}_l$, so post-multiplying both side by $\mathbf{Q}$ gives

$$\begin{aligned}\mathbf{P\Delta} &= \mathbf{X}_{c,(64\times 4)}\mathbf{Q}_{(4\times 4)} \\ &= \begin{pmatrix} \sum_{i=1}^4 q_{i1}x_{c,1i} & \sum_{i=1}^4 q_{i2}x_{c,1i} & \sum_{i=1}^4 q_{i3}x_{c,1i} & \sum_{i=1}^4 q_{i4}x_{c,1i} \\ \sum_{i=1}^4 q_{i1}x_{c,2i} & \sum_{i=1}^4 q_{i2}x_{c,2i} & \sum_{i=1}^4 q_{i3}x_{c,2i} & \sum_{i=1}^4 q_{i4}x_{c,2i} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^4 q_{i1}x_{c,64i} & \sum_{i=1}^4 q_{i2}x_{c,64i} & \sum_{i=1}^4 q_{i3}x_{c,64i} & \sum_{i=1}^4 q_{i4}x_{c,64i} \end{pmatrix}\end{aligned}$$

Biplot: Representing Cases & Variables

For cases, what we are plotting are linear combination of the data (mean centered) matrix.

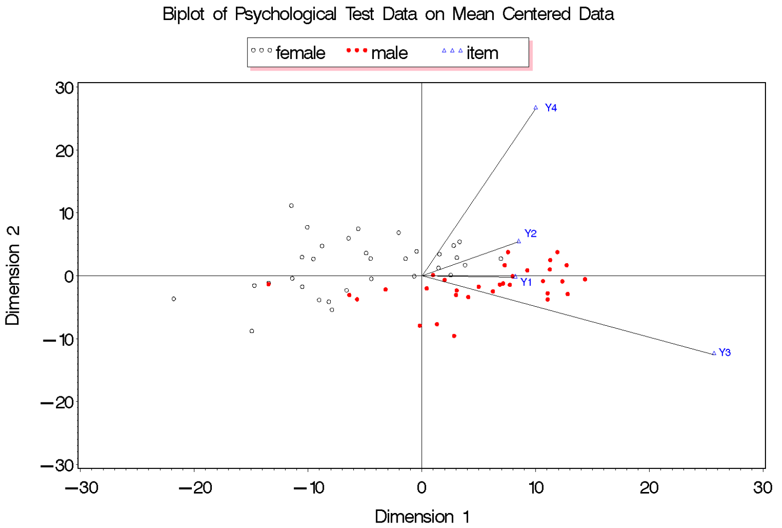
For example, for subject one, we plot the point

$$(p_{j1}\delta_1, p_{j2}\delta_2) = ((-0.002)(67.685), (-0.248)(31.859)) = (-0.135, -7.901).$$

To represent variables, we'll plot the rows of $\mathbf{Q}_{4\times 2}$ as vectors in the 2-dimensional space.

For example, for variable one, we'll plot $(0.274, -0.001)$.

For the plot, I actually plotted variable vectors multiplied by 30 for cosmetic purposes—it doesn't effect the interpretation.


Biplot of Psychological Test Data on Mean Centered Data
female
male
item
Y4
Y2
Y1
Y3
30
20
10
0
−10
−20
−30
Dimension 2
−30 −20 −10 0 10 20 30
Dimension 1

Maximization of Quadratic Forms

for Points on the Unit Sphere

In multivariate analyses, we have different goals and purposes $\longrightarrow$ different criteria to maximize (or minimize).

Let $\mathbf{B}_{p\times p}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ and eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$.

Maximization: $$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1 \quad \text{is obtained when } \mathbf{x} = \mathbf{e}_1$$

Minimization: $$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_p \quad \text{is obtained when } \mathbf{x} = \mathbf{e}_p$$

Maximization under an orthogonality constraint:

$$\max_{\mathbf{x}\perp\mathbf{e}_1,\ldots,\mathbf{e}_k} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_{k+1} \quad \text{is obtained when } \mathbf{x} = \mathbf{e}_{k+1}$$

Overview of the Rest of the Semester

See pages on web-site...