Honours Multivariate Analysis

Lecture 2 - Singular Value Decomposition, Eigenvalue Decomposition and Spectral Decomposition revisited

Stefan S. Britz

Department of Statistical Sciences
University of Cape Town

Course Outline

1. Introduction, Examples of Multivariate Data †
2. TOOLS
   1. Visualization and Summary Statistics †
   2. Singular Value Decomposition, Eigenvalue Decomposition and Spectral Decomposition revisited †
   3. The Multivariate Normal Distribution †
   4. Multivariate Maximum Likelihood Estimation †
   5. Multivariate Inference †
3. EXPLORATORY ANALYSIS
   1. Principal Component Analysis §
   2. Factor Analysis §
   3. Correspondence Analysis §
4. CONFIRMATORY ANALYSIS
   1. For grouped Multivariate Data:
      1. Manova †
      2. Discriminant Analysis §
   2. Regression
      1. Multivariate Regression †
      2. Canonical Correlation Analysis §

† Mr Stefan Britz

§ Prof Francesca Little

Decomposition of Matrices

In the Matrix Methods pre-course you will have learnt about

1. Eigendecomposition
2. Spectral decomposition
3. Singular value decomposition

These concepts are interrelated, and will be important going forward (especially SVD).

Recap: Eigenvectors and Eigenvalues

Given a square matrix, $\boldsymbol{A}$, a vector-scalar pair $\boldsymbol{x}$ and λ that satisfy the equation

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$$

are referred to as an eigenvector and corresponding eigenvalue of $\boldsymbol{A}$.

Recap: Eigenvectors and Eigenvalues

Given a square matrix, $\boldsymbol{A}$, a vector-scalar pair $\boldsymbol{x}$ and λ that satisfy the equation

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$$

are referred to as an eigenvector and corresponding eigenvalue of $\boldsymbol{A}$.

This is equivalent to solving $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = 0$, which only has a non-trivial solution if $|\boldsymbol{A} - \lambda\boldsymbol{I}| = 0$. This is referred to as the characteristic equation.

Recap: Eigenvectors and Eigenvalues

Given a square matrix, $\boldsymbol{A}$, a vector-scalar pair $\boldsymbol{x}$ and λ that satisfy the equation

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

are referred to as an eigenvector and corresponding eigenvalue of $\boldsymbol{A}$.

This is equivalent to solving $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = 0$, which only has a non-trivial solution if $|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0$. This is referred to as the characteristic equation.

Geometrically, an eigenvector is one of which the span remains unchanged by the transformation of $\boldsymbol{A}$, whilst the eigenvalue is the factor by which vectors lying on this span are stretched during the transformation.

Note that eigenvectors aren't unique, but are often expressed in their normalised form (length 1).

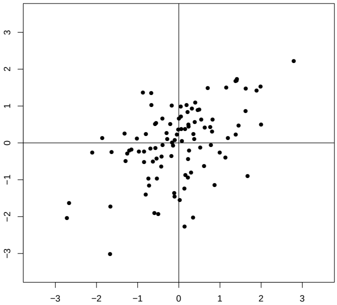
Spatial interpretation

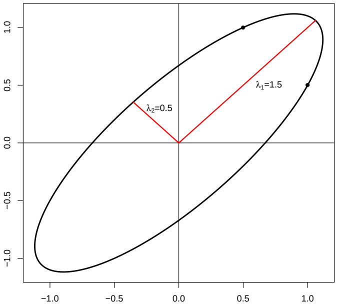
The covariance matrix for the data shown on the next slide is $\boldsymbol{S} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$.

The eigenvalue – (normalised) eigenvector pairs for $\boldsymbol{S}$ are

$$\lambda_1 = 1.5 \quad \boldsymbol{e}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \quad \text{and} \quad \lambda_2 = 0.5 \quad \boldsymbol{e}_2 = \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

Spatial interpretation





The major axis of the data corresponds to the largest eigenvalue and the secondary axis to the smaller eigenvalue. Note that the direction is determined by the eigenvectors, and that the ellipse goes through the points (s_{11}, s_{12}) and (s_{21}, s_{22}).

Eigendecomposition

Let $\boldsymbol{A}$ be a $k \times k$, symmetric, positive definite matrix with $\lambda_i > 0 \;\; \forall \;\; i$ and $\boldsymbol{e}_i, \; i = 1, \ldots, k$ indicating the k eigenvalue-eigenvector pairs.

If we arrange $\boldsymbol{e}_1 \ldots \boldsymbol{e}_k$ into the matrix $\boldsymbol{U}$, and define $\boldsymbol{D} = diag(\lambda_i)$, then

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'$$

Eigendecomposition

Let $\boldsymbol{A}$ be a $k \times k$, symmetric, positive definite matrix with $\lambda_i > 0 \;\; \forall \;\; i$ and $\boldsymbol{e}_i, \; i = 1, \ldots, k$ indicating the k eigenvalue-eigenvector pairs.

If we arrange $\boldsymbol{e}_1 \ldots \boldsymbol{e}_k$ into the matrix $\boldsymbol{U}$, and define $\boldsymbol{D} = diag(\lambda_i)$, then

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'$$

It can be shown that since $\boldsymbol{A}$ is symmetric, $\boldsymbol{U}$ is orthogonal. Therefore, $\boldsymbol{U}\boldsymbol{U}' = \boldsymbol{I}$ and $\boldsymbol{U}' = \boldsymbol{U}^{-1}$. We can then write the above as

$$\boldsymbol{A}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{D}$$

Spectral Decomposition

Let $\boldsymbol{A}$ be a $k \times k$, symmetric, positive definite matrix with $\lambda_i > 0 \;\; \forall \;\; i$ and $\boldsymbol{e}_i, \; i = 1, \ldots, k$ indicating the k eigenvalue-eigenvector pairs.

A different way of writing the eigendecomposition is by expressing $\boldsymbol{A}$ as the sum of k rank 1 matrices derived from the eigenvalue-eigenvector pairs. This is referred to as spectral decomposition.

$$\boldsymbol{A} = \lambda_1 \boldsymbol{e}_1 \boldsymbol{e}_1' + \lambda_2 \boldsymbol{e}_2 \boldsymbol{e}_2' + \ldots + \lambda_k \boldsymbol{e}_k \boldsymbol{e}_k'$$

Singular Value Decomposition (SVD)

Consider $\boldsymbol{A}$, an $m \times k$ matrix of real numbers. There exists an $m \times m$ orthogonal matrix $\boldsymbol{U}$ and a $k \times k$ orthogonal matrix $\boldsymbol{V}$ such that

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}'$$

where

- $\boldsymbol{U}$ contains the eigenvectors of $\boldsymbol{A}\boldsymbol{A}'$
- $\boldsymbol{V}$ contains the eigenvectors of $\boldsymbol{A}'\boldsymbol{A}$
- $\boldsymbol{\Lambda}$ contains $diag(\sqrt{\lambda_i})$, where the λ_i are the descending eigenvalues of $\boldsymbol{A}'\boldsymbol{A}$ (or $\boldsymbol{A}\boldsymbol{A}'$). If $m > k$, then either the square root of the diagonal matrix of eigenvalues of $(\boldsymbol{A}'\boldsymbol{A})_{k \times k}$ needs to be padded with $m - k$ rows of zeros; or the last $m - k$ columns of the square root of the diagonal matrix of eigenvalues of $(\boldsymbol{A}\boldsymbol{A}')_{m \times m}$ must be omitted.

Singular Value Decomposition (SVD)

Consider $\boldsymbol{A}$, an $m \times k$ matrix of real numbers. There exists an $m \times m$ orthogonal matrix $\boldsymbol{U}$ and a $k \times k$ orthogonal matrix $\boldsymbol{V}$ such that

$$\boldsymbol{A} = \boldsymbol{U\Lambda V'}$$

If $\boldsymbol{A}$ has rank r, then there exists r positive constants $\lambda_1, \ldots, \lambda_r$, r orthogonal $m \times 1$ unit vectors $\boldsymbol{u_1, \ldots, u_r}$ and r orthogonal $k \times 1$ unit vectors $\boldsymbol{v_1, \ldots, v_r}$ such that

$$\boldsymbol{A} = \sum_{i=1}^{r} \sqrt{\lambda_i} \boldsymbol{u}_i \boldsymbol{v}_i'$$

similar to the spectral decomposition theorem.

Example

Let $\boldsymbol{A} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$. Find the SVD of $\boldsymbol{A}$.

Example

Let $\boldsymbol{A} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$. Find the SVD of $\boldsymbol{A}$.

$$\boldsymbol{AA}' = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$$

To find the eigenvalues, we solve the characteristic equation:

$$\begin{aligned} |\boldsymbol{AA}' - \lambda \boldsymbol{I}| &= \begin{vmatrix} 11-\lambda & 1 \\ 1 & 11-\lambda \end{vmatrix} \\ &= (11-\lambda)^2 - 1 \\ &= \lambda^2 - 22\lambda + 120 \\ &= (\lambda - 12)(\lambda - 10) = 0 \end{aligned}$$

$\therefore \lambda_1 = 12$ and $\lambda_2 = 10$

Example

For $\lambda_1 = 12$:

$$\boldsymbol{A}\boldsymbol{A}'\boldsymbol{x} = \lambda_1 \boldsymbol{x}$$

$$\begin{bmatrix} 11x_1 + x_2 \\ x_1 + 11x_2 \end{bmatrix} = \begin{bmatrix} 12x_1 \\ 12x_2 \end{bmatrix}$$

$\therefore x_1 = x_2$ such that $\boldsymbol{u}_1' = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ is a normalised eigenvector of $\lambda_1 = 12$.

Example

For $\lambda_1 = 12$:

$$\boldsymbol{A}\boldsymbol{A}'\boldsymbol{x} = \lambda_1 \boldsymbol{x}$$

$$\begin{bmatrix} 11x_1 + x_2 \\ x_1 + 11x_2 \end{bmatrix} = \begin{bmatrix} 12x_1 \\ 12x_2 \end{bmatrix}$$

$\therefore x_1 = x_2$ such that $\boldsymbol{u}_1' = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ is a normalised eigenvector of $\lambda_1 = 12$.

We can show likewise that $\boldsymbol{u}_2' = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$ is a normalised eigenvector of $\lambda_2 = 10$.

Example

To find the columns of $\boldsymbol{V}$, consider

$$\boldsymbol{A}'\boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$

such that

$$\begin{aligned} |\boldsymbol{A}'\boldsymbol{A} - \lambda \boldsymbol{I}| &= (10-\lambda)\left[(10-\lambda)(2-\lambda) - 16\right] + 2\left[-2(10-\lambda)\right] \\ &= -\lambda^3 + 22\lambda^2 - 120\lambda \\ &= -\lambda(\lambda - 12)(\lambda - 10) = 0 \end{aligned}$$

yielding eigenvalues $\lambda_1 = 12$, $\lambda_2 = 10$ and $\lambda_3 = 0$.

Example

To find the columns of $\boldsymbol{V}$, consider

$$\boldsymbol{A}'\boldsymbol{A} = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$

such that

$$\begin{aligned} |\boldsymbol{A}'\boldsymbol{A} - \lambda \boldsymbol{I}| &= (10-\lambda)\left[(10-\lambda)(2-\lambda) - 16\right] + 2\left[-2(10-\lambda)\right] \\ &= -\lambda^3 + 22\lambda^2 - 120\lambda \\ &= -\lambda(\lambda - 12)(\lambda - 10) = 0 \end{aligned}$$

yielding eigenvalues $\lambda_1 = 12$, $\lambda_2 = 10$ and $\lambda_3 = 0$.

Solving the sets of linear equations given by $\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$ for each λ, we can determine the eigenvectors $\boldsymbol{v}_1' = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$, $\boldsymbol{v}_2' = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0 \end{bmatrix}$ and $\boldsymbol{v}_3' = \begin{bmatrix} \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & -\frac{5}{\sqrt{30}} \end{bmatrix}$.

Example

The SVD of $\boldsymbol{A}$ is therefore given by

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & -\frac{5}{\sqrt{30}} \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

Example

The SVD of $\boldsymbol{A}$ is therefore given by

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & -\frac{5}{\sqrt{30}} \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

which can also be expressed as

$$\boldsymbol{A} = \sqrt{12} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} + \sqrt{10} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

i.e. the sum of $r = 2$ matrices, both of rank 1.

Homework exercise 2.1

Johnson & Wichern exercise 2.22

Using the matrix

$$\boldsymbol{A} = \begin{bmatrix} 4 & 8 & 8 \\ 3 & 6 & -9 \end{bmatrix}$$

a) Calculate $\boldsymbol{A}'\boldsymbol{A}$ and obtain its eigenvalues and eigenvectors.

b) Calculate $\boldsymbol{A}\boldsymbol{A}'$ and obtain its eigenvalues and eigenvectors. Check that the non-zero eigenvalues are the same as those in part a).

c) Obtain the singular value decomposition of $\boldsymbol{A}$.

What's the use of SVD?

The main use is that SVD can be used to compute optimal low-rank approximations of arbitrary matrices. Apart from allowing us to visualise high dimensional data in 2 or 3 dimensions, it also forms part of a very wide range of applications:

- Face recognition – represent facial images as eigenfaces and compute distance between the query face image in the principal component space
- Latent Semantic Indexing for document extraction
- Image compression – Karhunen Loeve (KL) transform performs the best image compression. In MPEG, Discrete Cosine Transform (DCT) has the closest approximation to the KL transform in PSNR

Using SVD for lower-dimensional approximation

If the SVD of $\boldsymbol{A}_{m\times k}$ is $\boldsymbol{U\Lambda V'}$ and $s < k = rank(\boldsymbol{A})$, then $\boldsymbol{B} = \sum_{i=1}^{s} \sqrt{\lambda_i}\boldsymbol{u_i}\boldsymbol{v_i}'$ is the rank s least squares approximation to $\boldsymbol{A}$ that results in the best approximation among all matrices of rank $\leq s$.

Using SVD for lower-dimensional approximation

PROOF:
We use $\boldsymbol{U}\boldsymbol{U}' = \boldsymbol{I}_m$ and $\boldsymbol{V}\boldsymbol{V}' = \boldsymbol{I}_k$ to write the sum of squares as

$$\begin{aligned}
tr[(\boldsymbol{A}-\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})'] &= tr[\boldsymbol{U}\boldsymbol{U}'(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{V}\boldsymbol{V}'(\boldsymbol{A}-\boldsymbol{B})'] \\
&= tr[\boldsymbol{U}'(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{V}\boldsymbol{V}'(\boldsymbol{A}-\boldsymbol{B})'\boldsymbol{U}] \\
&= tr[\boldsymbol{U}'(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{V}(\boldsymbol{U}'(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{V})'] \\
&= tr[(\underbrace{\boldsymbol{U}'\boldsymbol{A}\boldsymbol{V}}_{\boldsymbol{\Lambda}} - \underbrace{\boldsymbol{U}'\boldsymbol{B}\boldsymbol{V}}_{\boldsymbol{C}})(\boldsymbol{U}'\boldsymbol{A}\boldsymbol{V} - \boldsymbol{U}'\boldsymbol{B}\boldsymbol{V})'] \\
&= tr[(\boldsymbol{\Lambda}-\boldsymbol{C})(\boldsymbol{\Lambda}-\boldsymbol{C})'] \\
&= \sum_{i=1}^{m}\sum_{j=1}^{k}(\sqrt{\lambda_{ij}} - c_{ij})^2 \\
&= \sum_{i=1}^{m}(\sqrt{\lambda_i} - c_{ii})^2 + \underbrace{\sum\sum}_{i\neq j} c_{ij}^2
\end{aligned}$$

which will be a minimum when $c_{ij} = 0$ for all $i \neq j$ and $c_{ii} = \sqrt{\lambda_i}$ for the s largest singular values with the other $c_{ii} = 0$.

Using SVD for lower-dimensional approximation

PROOF:

$$tr[(\boldsymbol{A}-\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})'] = tr[(\underbrace{\boldsymbol{U'AV}}_{\boldsymbol{\Lambda}} - \underbrace{\boldsymbol{U'BV}}_{\boldsymbol{C}})(\boldsymbol{U'AV}-\boldsymbol{U'BV})']$$

$$= \sum_{i=1}^{m}(\sqrt{\lambda_i}-c_{ii})^2 + \underbrace{\sum\sum}_{i\neq j} c_{ij}^2$$

which will be a minimum when $c_{ij}=0$ for all $i \neq j$ and $c_{ii}=\sqrt{\lambda_i}$ for the s largest singular values with the other $c_{ii}=0$.

$$\therefore \boldsymbol{UBV'} = \boldsymbol{\Lambda}_s \quad \text{or} \quad \boldsymbol{B} = \sum_{i=1}^{s}\sqrt{\lambda_i}\boldsymbol{u}_i\boldsymbol{v}_i'$$

We (you) will return to this decomposition when discussing Principal Components Analysis (PCA).