Introduction, Multivariate Data, Vectors and Matrices

February 19, 2024

Motto

Everything existing in the universe is the fruit of chance.

Democritus, the 5th Century BC

Let us know each other – Part I: Me

- Who am I? Małgorzata Bogdan
  e-mail: Malgorzata.Bogdan@uwr.edu.pl
- Why teaching this course? The quickly developing field of statistics with wide applications in science, economics, engineering, industry, etc.
  Simultaneous studying many aspects of the data is important to understand underlying mechanisms.
- Why me? This is my main area of research. I have attended high-dimensional and multivariate statistics classes at Stanford University and taught similar courses in Lund (Sweden) and Bologna (Italy) .

Let us know each other – Part II: You

Some honest answers to the following questions:

- Why have you decided to take this course?
- How much do you know about the type of problems that multivariate and high-dimensional statistical analysis is dealing with?
- How well do you feel prepared to take this course?
- Do you feel comfortable with operations on vectors and matrices?
- Do you feel comfortable with basic concepts of statistics?

Communication channels

- The easiest way to communicate is through e-mail: Malgorzata.Bogdan@uwr.edu.pl
- Office hours each Friday 15:00-16:00 or by appointment

Course 0rganization

- Highlights:
  - Two midterm exams: on April 19th and June 14th. Each midterm is worth a maximum of 50 points, totaling up to 100 points.
  - Labs – there will be assignments with both theoretical and computer problems. You will need to submit the report with your solutions for each set of assignment problems.

Projects – working with data

- The computer labs and projects will require using statistical software. We will use mainly the R-package – free and very popular statistical package. It contains a large variety of statistical routines. This choice has been dictated by a large number of supporting materials in R that are available for illustration of multivariate and high-dimensional statistical methods.

Downloading R-package and first steps

- Statistical R-package available for free download here
- Available on any PC platform (Mac, Windows, Linux).
- Worry free downloading (a couple of minutes).

Example of a very simple R session

- Suppose the following data are file `Table2_1.txt`

0.51	0.51	0.51	0.50	0.51	0.49	0.52	0.53	0.50	0.47
0.51	0.52	0.53	0.48	0.49	0.50	0.52	0.49	0.49	0.50
0.49	0.48	0.46	0.49	0.49	0.48	0.49	0.49	0.51	0.47
0.51	0.51	0.51	0.48	0.50	0.47	0.50	0.51	0.49	0.48
0.51	0.50	0.50	0.53	0.52	0.52	0.50	0.50	0.51	0.51

- Then the following code will be reading data and computing its mean and standard deviation:

```
#Getting data in a vector
x=scan("Table2_1.txt")
mean(x)
#[1] 0.4998
sd(x)
#[1] 0.01647385
```

- # at the beginning of the line denotes a commentary (so the following characters are not interpreted by R when the lines are copied to the command line).

Contents of the textbooks (1)

Our first textbook is

APPLIED MULTIVARIATE STATISTICAL ANALYSIS, BY RICHARD A. JOHNSON AND DEAN W. WICHERN; 6TH EDITION; PRENTICE HALL. NOTE: THE 5TH EDITION IS ALSO ACCEPTABLE.

- This is the most popular classic textbook on the multivariate statistics, including most popular unsupervised multivariate techniques like the principal component, discrimination and cluster analyses.
- Our intention is to cover chapters 1-5, 8, 11 and 12.
- However many topics will be covered very briefly due to the time limitation.

Contents of the textbooks (2)

Our second textbook is

HASTIE, T., TIBSHIRANI, R. WAINWRIGHT, M. (2015). STATISTICAL LEARNING WITH SPARSITY THE LASSO AND GENERALIZATIONS. BOCA RATON: CRC PRESS. [ALSO AVAILABLE AS E-BOOK.]

- This is the popular textbook on the high dimensional inference.
- It discusses the application of the multivariate techniques when the model dimension is larger than the sample size.
- Its main focus are the regularization techniques which lead to the dimensionality reduction.
- Our intention is to cover chapters 2,4, 7-9.
- However many topics will be covered very briefly due to the time limitation.

Other Reading Materials

- Lectures of Prof. E. J. Candès from Stanford University on the Theory of Statistics, will be available on MSTeams.
- Prof. Candès is a world leading expert in high-dimensional inference and regularization methods.
- We will cover the lectures related to the estimation of the multivariate vector of means and the training and prediction error in multiple regression.

Course outline – Highlights

Multivariate distributions
Estimation of the multivariate vector of means
High dimensional multiple regression
Principal component analysis
Discriminant analysis
Cluster analysis
Sparse multivariate methods
Sparse Graphical Models

What is multivariate statistics?

Multivariate statistics is analysis of data where you observe *several* variables for each object or item. You study dependence or interaction between variables.

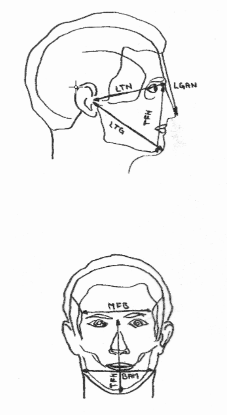
Example One – Scores on tests

In the following table grades in four subjects, Swedish language, English language, Mathematics and Civics, for six students are given

	Swedish	English	Mathematics	Civics
Anna	81	89	68	69
Bertil	73	72	75	80
Carina	53	50	59	63
David	90	88	83	82
Erika	73	59	83	92
Fredrik	73	60	79	90

Example – Skull measurements

Six head measurements from 34 individuals.



MFB	BAM	TFH	LGAN	LTN	LTG
113.2	111.7	119.6	53.9	127.4	143.6
117.6	117.3	121.2	47.7	124.7	143.9
112.3	124.7	131.6	56.7	123.4	149.3
116.2	110.5	114.2	57.9	121.6	140.9
112.9	111.3	114.3	51.5	119.9	133.5
104.2	114.3	116.5	49.9	122.9	136.7
110.7	116.9	128.5	56.8	118.1	134.7
105.0	119.2	121.1	52.2	117.3	131.4
115.9	118.5	120.4	60.2	123.0	146.8
96.8	108.4	109.5	51.9	120.1	132.2

MFB	BAM	TFH	LGAN	LTN	LTG
111.5	111.1	127.1	57.9	115.8	135.1
115.7	117.3	123.0	50.8	122.2	143.1
112.2	120.6	119.6	61.3	126.7	141.1
118.7	122.9	126.7	59.8	125.7	138.3
118.9	118.4	127.7	64.6	125.6	144.3
114.2	109.4	119.3	58.7	121.1	136.2
116.2	110.5	114.2	57.9	121.6	140.9
112.9	111.3	114.3	51.5	119.9	133.5
104.2	114.3	116.5	49.9	122.9	136.7
110.7	116.9	128.5	56.8	118.1	134.7
105.0	119.2	121.1	52.2	117.3	131.4
115.9	118.5	120.4	60.2	123.0	146.8
96.8	108.4	109.5	51.9	120.1	132.2

MFB	BAM	TFH	LGAN	LTN	LTG
113.8	113.6	135.8	54.3	119.5	130.9
122.4	117.2	122.2	56.4	123.3	142.9
110.4	110.8	122.1	51.2	115.6	132.7
114.9	108.6	122.9	56.3	122.7	140.3
108.4	118.7	117.8	50.0	113.7	131.0
105.3	107.2	116.0	52.5	117.4	133.2

Things you might want to do

- “Understand” the variables, based on samples. What are the connections between them?
  - Univariate and bivariate plots
  - 3D plots
  - Finding variables that behave in a related manner
  - Detect errors in the data

  (Feature extraction)
- Find or construct important or informative variables, and hereby reducing the amount of data. (Principal component analysis)

More things you might want to do

- Explore the data and generate new hypotheses about the underlying phenomenon. (Exploratory analysis)
- Divide the sample into two or more groups (Clustering)
- Test hypotheses: Are two variables correlated? Is the mean of a variables equal to some given values? Are the means of two variables equal?
- Conversely, we can look at the confidence regions for the parameters.

Even more things ...

- Prediction: We can fit a linear relationsship betwen the varibles. Then observing some of the variables how to predict the rest?
- Classification: Based on a grouping of some objects, how to place a new object into one of the groups.

Description of data

The observations of a sample can be arranged into an $n \times p$ matrix (data matrix) $\mathbf{X}$, where n is the number of survey or experimental units (the size of the sample) and p is the number of variables studied.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \dots & x_{jk} & \dots & x_{jp} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} & \dots & x_{np} \end{bmatrix}$$

Example

In the following table the grades in four subjects, Swedish language, English language, Mathematics and Civics, for six students are given

	Swedish	English	Mathematics	Civics
Anna	81	89	68	69
Bertil	73	72	75	80
Carina	53	50	59	63
David	90	88	83	82
Erika	73	59	83	92
Fredrik	73	60	79	90

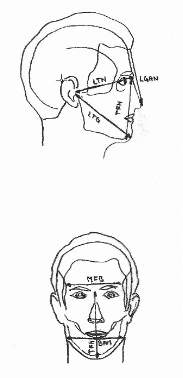
..

Data can be expressed as a matrix

$$\mathbf{X} = \begin{bmatrix} 81 & 89 & 68 & 69 \\ 73 & 72 & 75 & 80 \\ 53 & 50 & 59 & 63 \\ 90 & 88 & 83 & 82 \\ 73 & 59 & 83 & 92 \\ 73 & 60 & 79 & 90 \end{bmatrix}$$

with $n = 6$ rows (individuals, items) and $p = 4$ columns (variables).

Measuring Heads



Measuring Heads

Six head measurements from 34 individuals.

MFB	BAM	TFH	LGAN	LTN	LTG
113.2	111.7	119.6	53.9	127.4	143.6
117.6	117.3	121.2	47.7	124.7	143.9
112.3	124.7	131.6	56.7	123.4	149.3
116.2	110.5	114.2	57.9	121.6	140.9
112.9	111.3	114.3	51.5	119.9	133.5
104.2	114.3	116.5	49.9	122.9	136.7
110.7	116.9	128.5	56.8	118.1	134.7
105.0	119.2	121.1	52.2	117.3	131.4
115.9	118.5	120.4	60.2	123.0	146.8
96.8	108.4	109.5	51.9	120.1	132.2

MFB	BAM	TFH	LGAN	LTN	LTG
110.7	117.5	115.4	55.2	125.0	140.6
108.4	113.7	122.2	56.2	124.5	146.3
104.1	116.0	124.3	49.8	121.8	138.1
107.9	115.2	129.4	62.2	121.6	137.9
106.4	109.0	114.9	56.8	120.1	129.5
112.7	118.0	117.4	53.0	128.3	141.6
109.9	105.2	122.2	56.6	122.2	137.8
116.6	119.5	130.6	53.0	124.0	135.3
109.9	113.5	125.7	62.8	122.7	139.5
107.1	110.7	121.7	52.1	118.6	141.6

MFB	BAM	TFH	LGAN	LTN	LTG
113.3	117.8	120.7	53.5	121.6	138.6
108.1	116.3	123.9	55.5	125.4	146.1
111.5	111.1	127.1	57.9	115.8	135.1
115.7	117.3	123.0	50.8	122.2	143.1
112.2	120.6	119.6	61.3	126.7	141.1
118.7	122.9	126.7	59.8	125.7	138.3
118.9	118.4	127.7	64.6	125.6	144.3
114.2	109.4	119.3	58.7	121.1	136.2

MFB	BAM	TFH	LGAN	LTN	LTG
113.8	113.6	135.8	54.3	119.5	130.9
122.4	117.2	122.2	56.4	123.3	142.9
110.4	110.8	122.1	51.2	115.6	132.7
114.9	108.6	122.9	56.3	122.7	140.3
108.4	118.7	117.8	50.0	113.7	131.0
105.3	107.2	116.0	52.5	117.4	133.2

... can be expressed as a matrix

$$\mathbf{X} = \begin{bmatrix} 113.2 & 111.7 & 119.6 & 53.9 & 127.4 & 143.6 \\ 117.6 & 117.3 & 121.2 & 47.7 & 124.7 & 143.9 \\ 112.3 & 124.7 & 131.6 & 56.7 & 123.4 & 149.3 \\ . & . & . & . & . & . \\ 108.4 & 118.7 & 117.8 & 50.0 & 113.7 & 131.0 \\ 105.3 & 107.2 & 116.0 & 52.5 & 117.4 & 133.2 \end{bmatrix}$$

with $n = 34$ rows (individuals, items) and $p = 6$ columns (variables).

Mathematical description of data and models

For obvious reasons the multivariate data and corresponding models require efficient notation of linear algebra, i.e. algebra of vectors and matrices.

While for practical purposes most of the operations on matrices and vectors can be easily made by means of computers, we still need to review basic formal concepts, while trying to minimize mathematical jargon to minimum.

The following concepts have to be reviewed:

- Matrix and vector algebra
- Determinant and rank
- Unit matrix, and Inverse matrix
- Eigenvalues and Eigenvectors

Whatever we will not cover in today's lecture will be continued during Exercise 1 when we will work on Assignment 1!

Vectors

(column) vector **x** is an ordered n-tuple of numbers:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{bmatrix}$$

with element x_i. (The numbers x_i will be real numbers subsequently.)
Transposed vector (or row vector)

$$\mathbf{x}^T = [x_1 \; x_2 \; \ldots \; x_n]$$

Scalar multiplication

Multiplication of vector **x** with a scalar (constant) c:

$$c\mathbf{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ . \\ . \\ cx_n \end{bmatrix}$$

Addition of two vectors

Addition of two vectors **x** and **y**:

$$\mathbf{x}+\mathbf{y}=\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{bmatrix}+\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix}=\begin{bmatrix} x_1+y_1 \\ x_2+y_2 \\ . \\ . \\ x_n+y_n \end{bmatrix}$$

The length of a vector x:

$$\|\mathbf{x}\|=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}$$

inner product (or scalar product)

$$\mathbf{x}^T\mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \mathbf{y}^T\mathbf{x}$$

The length of **x** can be written

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}}$$

Alternative notation is $L_{\mathbf{x}} = \|x\|_2$.
If $\mathbf{x}^T\mathbf{y} = 0$ then **x** and **y** are orthogonal. We write $\mathbf{x} \perp \mathbf{y}$.
The angle θ between two vectors **x** and **y** is given by

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}$$

Matrices

A matrix is a rectangular scheme of (often real) numbers arranged in n rows and p columns

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix}$$

$\mathbf{A}$ is said to be of size $n \times p$ (n times p). The element a_{ij} is the number in row i and column j.

Addition and subtraction

If $\mathbf{A}$ and $\mathbf{B}$ have the same size then their sum ($\mathbf{A}+\mathbf{B}$) and their difference ($\mathbf{A}-\mathbf{B}$) are defined by elementwise addition and subtraction respectively.

Ex: If $\mathbf{A} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 5 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 5 & 2 & -6 \\ 2 & -3 & 0 \end{bmatrix}$ then

$$\mathbf{A}+\mathbf{B} = \begin{bmatrix} 8 & 1 & -4 \\ 3 & 2 & 4 \end{bmatrix}$$

$$\mathbf{A}-\mathbf{B} = \begin{bmatrix} -2 & -3 & -8 \\ -1 & 8 & 4 \end{bmatrix}$$

Multiplication

If **A** is $n \times k$ and **B** is $k \times p$ then the product $(\mathbf{A} \cdot \mathbf{B})$ of the matrices is the $n \times p$ matrix **C** whose (i, j) element c_{ij} is given by

$$c_{ij} = \sum_{\ell=1}^{k} a_{i\ell} b_{\ell j}$$

Ex: If $\mathbf{A} = \begin{bmatrix} 3 & 5 \\ 4 & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$ then $\mathbf{A} \cdot \mathbf{B} = \begin{bmatrix} 26 & 39 \\ 12 & 18 \end{bmatrix}$.
Subsequently we use juxtaposition to denote matrix multiplication and write **AB** instead of $\mathbf{A} \cdot \mathbf{B}$.

Transpose

The matrix $\mathbf{A}^T$ (or $\mathbf{A}'$ in other notation) (*read:* $\mathbf{A}$ *transpose*) of the $m \times n$ matrix $\mathbf{A}$ is the $n \times m$ matrix whose (i, j)th element is equal to the (j, i)th element of $\mathbf{A}$. The rows of $\mathbf{A}$ equal the columns of $\mathbf{A}^T$ and vice versa.

Ex: If $\mathbf{A} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 5 & 4 \end{bmatrix}$ then $\mathbf{A}^T = \begin{bmatrix} 3 & 1 \\ -1 & 5 \\ 2 & 4 \end{bmatrix}$.

We have

$$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$$

and

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$$

Special matrices

- A matrix with $n = m$, i.e as many columns as rows is called a *square matrix*.
- Ex: The matrix $\mathbf{A} = \begin{bmatrix} 3 & 5 \\ 4 & 1 \end{bmatrix}$ is square.
- *Symmetric matrix:* A square matrix with element (i, j) equal to element (j, i), i.e. $\mathbf{A}^T = \mathbf{A}$.
- Ex: $\mathbf{A} = \begin{bmatrix} 3 & 5 \\ 5 & 1 \end{bmatrix}$ is symmetric.

Rank

- Vectors of the same dimension are called linearly independent if none one of them can be expressed as linear combination of the others (linear combination of vectors is a sum of vectors multiplied by some scalars).
- The rank of a matrix is the number of linearly independent columns.
- The rank is at most the minimum from the number of columns and the number of rows.

Determinant

- Definition: The determinant of a square matrix **A**, $(k \times k)$ is denoted $|\mathbf{A}|$, or $\det(\mathbf{A})$, and formally is a number such that

$$|\mathbf{A}| = a_{11} \quad \text{if } k = 1$$

$$|\mathbf{A}| = \sum_{j=1}^{k} a_{1j}(-1)^{1+j}|\mathbf{A}_{1j}|$$

  where $\mathbf{A}_{1j}$ is the matrix obtained if row 1 and column j are deleted.
- The definition is not that important (typically determinant is computed by means of computers), but properties are.
- If $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $\det(\mathbf{A}) = a_{11}a_{22} - a_{21}a_{12}$.

Properties

Some properties of the determinant:

1. $|\mathbf{A}^T| = |\mathbf{A}|$
2. If a row (or a column) in $\mathbf{A}$ is zero, then $|\mathbf{A}| = 0$
3. If any two rows (or columns) are identical then $|\mathbf{A}| = 0$
4. If $\mathbf{A}$ has rank equal to its size (non-singular matrix) then $|\mathbf{A}^{-1}| = 1/|\mathbf{A}|$.
5. $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$
6. $|c\mathbf{A}| = c^k|\mathbf{A}|$

Trace

The trace of a square matrix $\mathbf{A}$ is denoted $tr(\mathbf{A})$ is defined as the sum of the elements in the main diagonal of $\mathbf{A}$.

$$tr(\mathbf{A}) = \sum_{i=1}^{k} a_{ii}$$

It holds that

$$tr(\mathbf{AB}) = tr(\mathbf{BA})$$

Cyclic property of the trace: for any matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, such that $\mathbf{ABC}$ is the square matrix

$$tr(\mathbf{ABC}) = tr(\mathbf{CAB})$$

Diagonal matrix

- A square matrix whose elements outside the main diagonal are zero is called a *diagonal matrix*.
- Ex:

$$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$$

- The identity matrix is a diagonal matrix with ones in the main diagonal. It is denoted $\mathbf{I}$.
- For an arbitrary matrix $\mathbf{A}$ (for which multiplication with $\mathbf{I}$ is well defined) it holds that

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

- Ex:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Matrix inverses

- Consider square $n \times n$ matrices. The matrix denoted by $\mathbf{A}^{-1}$ is called the inverse of $\mathbf{A}$ if $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.
- If $\mathbf{A}$ is a diagonal with non-zero terms on the diagonal, then $\mathbf{A}^{-1}$ exists.
- A matrix $\mathbf{A}$ has an inverse if and only if $\text{rank}(\mathbf{A}) = n$.
- It holds that

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

  and

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$$

- If $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is non-singular ($ad - bc \neq 0$), then $\mathbf{A}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Inverses are nasty creatures!

- Consider the matrices

$$\mathbf{A} = \begin{bmatrix} 4 & 4.001 \\ 4.001 & 4.002 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 4 & 4.001 \\ 4.001 & 4.002001 \end{bmatrix}$$

- These matrices are identical except for a small difference in the (2,2) position.
- Applying the inverse of a 2 by 2 matrix we have

$$\mathbf{A}^{-1} = -1000000 \begin{bmatrix} 4.002 & -4.001 \\ -4.001 & 4 \end{bmatrix}$$

$$\mathbf{B}^{-1} = -333333 \begin{bmatrix} 4.002001 & -4.001 \\ -4.001 & 4 \end{bmatrix}$$

So $\mathbf{A}^{-1} \approx -3\mathbf{B}^{-1}$.

- Small changes – perhaps caused by rounding – can give subtantially different inverses.

Eigenvectors and eigenvalues

- A square matrix $\mathbf{A}$ has the eigenvalue λ and corresponding eigenvector $\mathbf{e}$ if
$$\mathbf{A}\mathbf{e} = \lambda \mathbf{e}$$
- We will consider normalised eigenvectors: $\mathbf{e}^T\mathbf{e} = 1$.
- When looking for eigenvalues we search for solutions in λ
$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0$$
- The equation is a polynomial in λ of degree k and is called *the characteristic equation*.

Spectral decomposition

- If $\mathbf{A}$ is a symmetric $k \times k$ matrix, then the eigenvectors correponding to different eigenvalues are orthogonal: $\mathbf{e}_j^T \mathbf{e}_i = 0$.
- A symmetric $k \times k$ matrix can be written

$$\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i^T = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^T$$

where the columns of $\mathbf{P}$ are made of eigenvectors of $\mathbf{A}$ and $\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$, and

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & \lambda_k \end{bmatrix}$$

is diagonal.

Positive definite matrices and their square roots

- If all eigenvalues of a symmetric matrix are positive, the matrix is called positive definite.
- A positive definite matrix $\mathbf{A}$ satisfies: $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x}$
- For such a matrix we define

$$\mathbf{A}^{1/2} = \mathbf{P}\sqrt{\mathbf{\Lambda}}\mathbf{P}^T = \sum_{i=1}^{k} \mathbf{e}_i \sqrt{\lambda_i}\mathbf{e}_i^T$$

Here

$$\mathbf{\Lambda}^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & \sqrt{\lambda_k} \end{bmatrix}$$

- Then it holds that

$$\mathbf{A}^{1/2} \cdot \mathbf{A}^{1/2} = \mathbf{A}$$