Session 7: Spectral Theorem, PCA and SVD

Optimization and Computational Linear Algebra for Data Science

Léo Miolane

Contents

The Spectral Theorem

The spectral theorem

Theorem

Let $A \in \mathbb{R}^{n \times n}$ be a **symmetric** matrix. Then there is a orthonormal basis of $\mathbb{R}^n$ composed of eigenvectors of A.

Theorem (Matrix formulation)

Let $A \in \mathbb{R}^{n \times n}$ be a **symmetric** matrix. Then there exists an orthogonal matrix P and a diagonal matrix D of sizes $n \times n$ such that

$$A = PDP^{\mathsf{T}}.$$

Geometric interpretation

A key result

Proposition

Let A be a $n \times n$ symmetric matrix and let $\lambda_1 \geq \cdots \geq \lambda_n$ be its n eigenvalues and $v_1, \ldots, v_n$ be an associated orthonormal family of eigenvectors. Then

$$\lambda_1 = \max_{\|v\|=1} v^\mathsf{T} A v \qquad \text{and} \qquad v_1 = \arg\max_{\|v\|=1} v^\mathsf{T} A v\,.$$

Moreover, for $k = 2, \ldots, n$:

$$\lambda_k = \max_{\|v\|=1,\, v \perp v_1, \ldots, v_{k-1}} v^\mathsf{T} A v\,, \quad \text{and} \quad v_k = \arg\max_{\|v\|=1,\, v \perp v_1, \ldots, v_{k-1}} v^\mathsf{T} A v.$$

Proof

Proof

Principal Component Analysis

Empirical mean and covariance

We are given a dataset of n points $a_1, \dots, a_n \in \mathbb{R}^d$

$\underline{d = 1}$

- **Mean**

$$\mu = \frac{1}{n} \sum_{i=1}^{n} a_i \qquad \in \mathbb{R}$$

- **Variance**

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (a_i - \mu)^2 \qquad \in \mathbb{R}$$

Empirical mean and covariance

We are given a dataset of n points $a_1, \ldots, a_n \in \mathbb{R}^d$

$\underline{d = 1}$

- **Mean**

$$\mu = \frac{1}{n}\sum_{i=1}^{n} a_i \qquad \in \mathbb{R}$$

- **Variance**

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - \mu)^2 \qquad \in \mathbb{R}$$

$\underline{d \geq 2}$

- **Mean**

$$\mu = \frac{1}{n}\sum_{i=1}^{n} a_i \qquad \in \mathbb{R}^d$$

- **Covariance matrix**

$$S = \frac{1}{n}\sum_{i=1}^{n}(a_i - \mu)(a_i - \mu)^\mathsf{T} \qquad \in \mathbb{R}^{d\times d}$$

$$= \frac{1}{n}\sum_{i=1}^{n} a_i a_i^\mathsf{T} \qquad \text{if } \mu = 0.$$

PCA

- We are given a dataset of n points $a_1, \dots, a_n \in \mathbb{R}^d$, where d is «large».
- **Goal:** represent this dataset in lower dimension, i.e. find $\widetilde{a}_1, \dots, \widetilde{a}_n \in \mathbb{R}^k$ where $k \ll d$.
- Assume that the dataset is centered: $\sum_{i=1}^n a_i = 0$.
- Then, S can be simply written as:

$$S = \sum_{i=1}^{n} a_i a_i^\mathsf{T} = A^\mathsf{T} A.$$

where A is the $n \times d$ "data matrix":

$$A = \begin{pmatrix} - \ a_1 \ - \\ \vdots \\ - \ a_n \ - \end{pmatrix}.$$

Direction of maximal variance

Direction of maximal variance

2nd direction of maximal variance

Which value of k should we take?

Singular Value Decomposition

Singular values/vectors

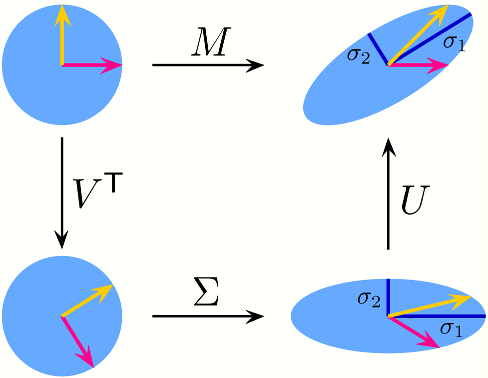
Singular Value decomposition

Theorem

Let $A \in \mathbb{R}^{n \times m}$. Then there exists two orthogonal matrices $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ and a matrix $\Sigma \in \mathbb{R}^{n \times m}$ such that $\Sigma_{1,1} \geq \Sigma_{2,2} \geq \cdots \geq 0$ and $\Sigma_{i,j} = 0$ for $i \neq j$, that verify

$$A = U \Sigma V^{\mathsf{T}}.$$

Geometric interpretation of $U\Sigma V^\top$



$$M = U\Sigma V^\top$$

Questions?

Questions?