Session 7: Spectral Theorem, PCA and SVD

Optimization and Computational Linear Algebra for Data Science

Léo Miolane

Contents

1. The Spectral Theorem
2. Principal Component Analysis
3. Singular Value Decomposition

Midterm

- The Midterm exam is in 2 weeks.
- **Scope:** everything that we have seen so far (this week's video included).
- **Knowing is not enough!** You need to practice: review problems available on the course's webpage.
- Past years midterms also available, with solutions.
- **Important:** when working on a problem, take **at least** 10min on it before looking at the solution (in case you are stuck).
- The midterm is open books/notes, but **if you think that you need them for the exam, this probably means that you are not prepared enough**.

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

The Theorem behind PCA

Theorem

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

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**

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

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

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 $\tilde{a}_1, \dots, \tilde{a}_n \in \mathbb{R}^k$ 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”:

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

Direction of maximal variance

Direction of maximal variance

Direction of maximal variance

Good news: $S = A^\mathsf{T} A$ is symmetric.
Spectral Theorem: let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of S and $(v_1, \ldots, v_n)$ an associated orthonormal basis of eigenvectors.

2nd direction of maximal variance

j^{th} direction of maximal variance

- The « j^{th} direction of maximal variance » is v_j since v_j is solution of

$$\text{maximize } v^\top S v, \qquad \text{subject to } \|v\| = 1,\ v \perp v_1, v \perp v_2, \ldots, v \perp v_{j-1}.$$

- The dimensionally reduced dataset is then

$$\begin{pmatrix} \langle v_1, a_1 \rangle \\ \langle v_2, a_1 \rangle \\ \vdots \\ \langle v_k, a_1 \rangle \end{pmatrix}, \begin{pmatrix} \langle v_1, a_2 \rangle \\ \langle v_2, a_2 \rangle \\ \vdots \\ \langle v_k, a_2 \rangle \end{pmatrix}, \begin{pmatrix} \langle v_1, a_3 \rangle \\ \langle v_2, a_3 \rangle \\ \vdots \\ \langle v_k, a_3 \rangle \end{pmatrix} \cdots \begin{pmatrix} \langle v_1, a_n \rangle \\ \langle v_2, a_n \rangle \\ \vdots \\ \langle v_k, a_n \rangle \end{pmatrix}.$$

Recap

Which value of k should we take?

Which value of k should we take?

Singular Value Decomposition

Singular values/vectors

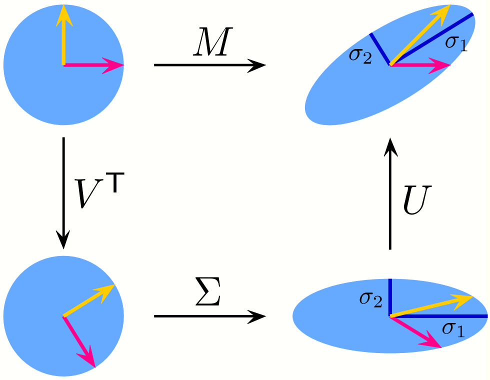
Singular Value decomposition

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?