

Lecture 3.1: The rank

Optimization and Computational Linear Algebra for Data Science

Rank of a family of vectors

Definition

We define the rank of a family x_1, \dots, x_k of vectors of \mathbb{R}^n as the dimension of its span:

$$\text{rank}(x_1, \dots, x_k) \stackrel{\text{def}}{=} \dim(\text{Span}(x_1, \dots, x_k)).$$

Rank of a matrix

Definition

Let $M \in \mathbb{R}^{n \times m}$. Let $c_1, \dots, c_m \in \mathbb{R}^n$ be its columns. We define

$$\text{rank}(M) \stackrel{\text{def}}{=} \text{rank}(c_1, \dots, c_m) = \dim(\text{Im}(M)).$$

Example

« Rank of columns = rank of rows »

Proposition

Let $M \in \mathbb{R}^{n \times m}$. Let $r_1, \dots, r_n \in \mathbb{R}^m$ be the rows of M and $c_1, \dots, c_m \in \mathbb{R}^n$ be its columns. Then we have

$$\text{rank}(r_1, \dots, r_n) = \text{rank}(c_1, \dots, c_m) = \text{rank}(M).$$

The rank in Data Science

Consider a matrix M of size 1000×500 :

$$M = \begin{pmatrix} - & r_1 & - \\ & \vdots & \\ - & r_{1000} & - \end{pmatrix}$$

The rank in Data Science

Assume now that

- ❖ The rows of M corresponds to Netflix's users.
- ❖ The columns of M corresponds to Netflix's movies.
- ❖ The entry $M_{i,j}$ is rating of the movie j by the user i .