matrix

Authors

Alexander Jung[*], Department of Computer Science, Aalto University, Espoo, Finland
[*]Corresponding author: alex.jung@aalto.fi

Abstract

A matrix is a rectangular array of numbers arranged in rows and columns. The entry $A_{r,j}$ of a matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$ sits in row r and column j. A prominent example in machine learning (ML) is the feature matrix of a dataset, obtained by stacking the feature vectors of m data points row-wise. A matrix represents several distinct mathematical objects: a system of linear equations, such as the normal equations of linear regression; a linear map between two vector spaces, after fixing a basis for each; or a tabular dataset with one row per data point and one column per feature. Matrices are the two-dimensional special case of an array.

Definition

Consider a tabular dataset of m data points, each described by d numerical features. Stacking the feature vectors row-wise produces an $m \times d$ rectangular array of numbers. This two-dimensional numeric array is referred to as the feature matrix $\mathbf{X}$ of the dataset.

More generally, a matrix of size $m \times d$ is a two-dimensional array of numbers denoted by

$$\mathbf{A} = \begin{pmatrix} A_{1,1} & A_{1,2} & \ldots & A_{1,d} \\ A_{2,1} & A_{2,2} & \ldots & A_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m,1} & A_{m,2} & \ldots & A_{m,d} \end{pmatrix} \in \mathbb{R}^{m \times d}.$$

Here, $A_{r,j}$ denotes the matrix entry in the r-th row and the j-th column. Matrices are the two-dimensional special case of the more general array.

Matrices represent several distinct mathematical objects (Strang, 2016), including the following:

- Systems of linear equations: A matrix can represent a system of linear equations

  $$\begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \text{ compactly as } \mathbf{A}\mathbf{w} = \mathbf{y}.$$

  One important example is the normal equations $\mathbf{X}^\top \mathbf{X}\, \mathbf{w} = \mathbf{X}^\top \mathbf{y}$ that give the optimal model parameters in linear regression.

- Linear maps: Consider two vector spaces $\mathcal{U}$ and $\mathcal{V}$, of dimension d and m, respectively. Fixing a basis $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(d)}$ for $\mathcal{U}$ and a basis $\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(m)}$ for $\mathcal{V}$, each matrix $\mathbf{A} \in \mathbb{R}^{m \times d}$ naturally defines a linear map $\alpha : \mathcal{U} \to \mathcal{V}$ (see Fig. 1) such that

$$\mathbf{u}^{(j)} \mapsto \sum_{r=1}^{m} A_{r,j} \mathbf{v}^{(r)}.$$

- Datasets: as in the opening example, a matrix can represent a dataset: each row a single data point, each column a specific feature or label.

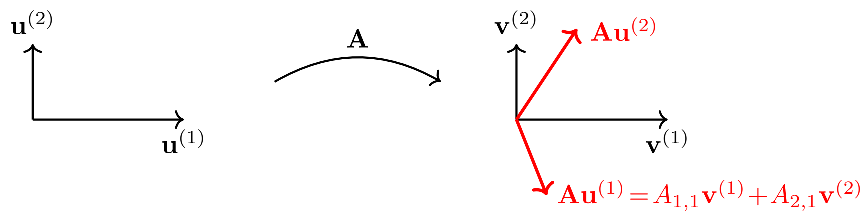

Figure 1: Matrix $\mathbf{A}$ defining a linear map between two vector spaces

Cross References

linear map, dataset, linear model, vector, feature matrix, normal equations, CNN

References

1. Strang (2016). Introduction to Linear Algebra. 5th ed., Wellesley-Cambridge Press, Wellesley, MA, USA.