

# Math Basics

To apply machine learning, you should be familiar with basic linear algebra concepts. The most important terms that will come up are listed here, with some further examples in the numpy/pandas notebook.

## Notation

Symbol	Space	Example	Description
$x$	$\mathbb{R}$	42	a scalar, i.e., a regular number
$\mathbf{x}$	$\mathbb{R}^d$	$\begin{pmatrix} 3 \\ 6 \\ -1 \end{pmatrix} \in \mathbb{R}^3$	a $d$ -dimensional vector, i.e., a list of $d$ numbers
$X$	$\mathbb{R}^{n \times d}$	$\begin{pmatrix} 5 & 2 \\ -6 & 8 \\ 2.5 & 7 \end{pmatrix} \in \mathbb{R}^{3 \times 2}$	a $n \times d$ matrix, i.e., a collection of numbers in a grid with $n$ rows and $d$ columns
$X^\top$	$\mathbb{R}^{d \times n}$	$\begin{pmatrix} 5 & -6 & 2.5 \\ 2 & 8 & 7 \end{pmatrix} \in \mathbb{R}^{2 \times 3}$	the transpose of the matrix $X$ above; the numbers from $X$ were mirrored at the diagonal and the shape is now $d$ rows and $n$ columns

## Operations

### L2-norm of a vector:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{k=1}^d x_k^2} = \sqrt{x_1^2 + x_2^2 + \dots + x_d^2} = p \in \mathbb{R}$$

The L2-norm (or Euclidean norm or magnitude or length) of a vector is, amongst other purposes, needed to compute the cosine similarity of two vectors.

### Scalar product between two vectors:

$$\langle \mathbf{x}, \mathbf{w} \rangle = \sum_{k=1}^d x_k \cdot w_k = x_1 w_1 + x_2 w_2 + \dots + x_d w_d = y$$

Each entry  $x_k$  in the  $d$ -dimensional vector  $\mathbf{x}$  is multiplied with the corresponding entry from the (also  $d$ -dimensional) vector  $\mathbf{w}$  and then all these products are summed up, resulting in a single number (i.e. a scalar), in this case denoted with  $y$ .

### Matrix multiplication:

$$XW = Y$$

The matrix  $X \in \mathbb{R}^{n \times d}$  multiplied by the matrix  $W \in \mathbb{R}^{d \times m}$  produce the matrix  $Y \in \mathbb{R}^{n \times m}$ , where each entry in  $Y$ ,  $Y_{ij}$  (in the  $i$ th row and  $j$ th column), is computed as the scalar product of the  $i$ th row of  $X$  and the  $j$ th column of  $W$ . Please note that  $X$  and  $W$  need to agree on the inner dimension  $d$ .

The matrix multiplication is the generalization of the scalar product. You can also multiply matrices and vectors, as vectors are just  $d \times 1$  or  $1 \times d$  matrices.

### Functions:

Any predefined computation mapping from some input to an output. For example:

$$f(\mathbf{x}) = 25 \cdot \langle \mathbf{x}, \mathbf{x} \rangle = y$$

Here the function  $f$  is a mapping from some vector  $\mathbf{x}$  to the scalar  $y$ , where  $y$  is computed as 25 times the scalar product of  $\mathbf{x}$  with itself.