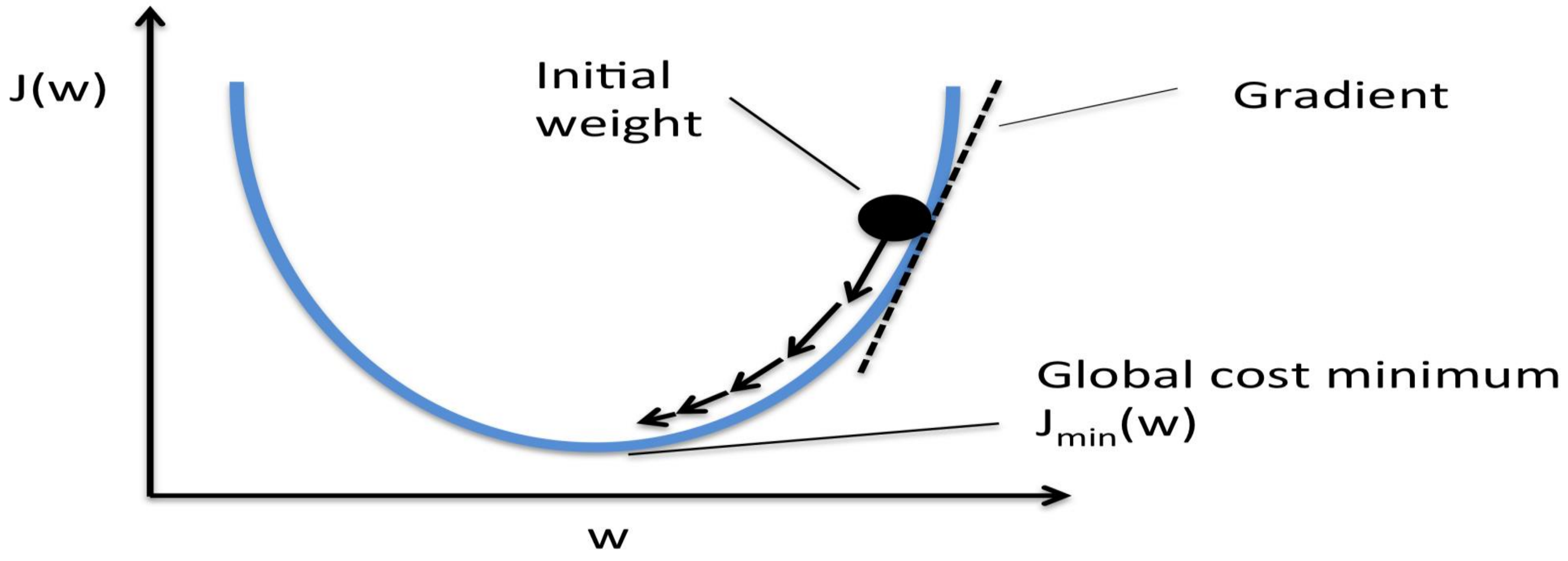
Introduction to Linear Algebra

Francesco Pugliese, PhD

neural1977@gmail.com

WHAT ARE THE PREREQUISITES FOR MACHINE LEARNING?

- **Applied Math** allows us to define **functions** of **many variables**, find the **highest** and **lowest** point of these functions and quantify degrees of **belief**.
- We need to design **cost functions** that measures how well those **beliefs** correspond with **reality** and using **a training algorithm** to **minimize** the cost function.



LINEAR ALGEBRA

- **Linear Algebra** is a branch of **continuous** rather than **discrete maths**, dealing with **linear equations** like:

$$a_1 x_1 + \cdots + a_n x_n = b,$$

- Where the **linear functions** is:

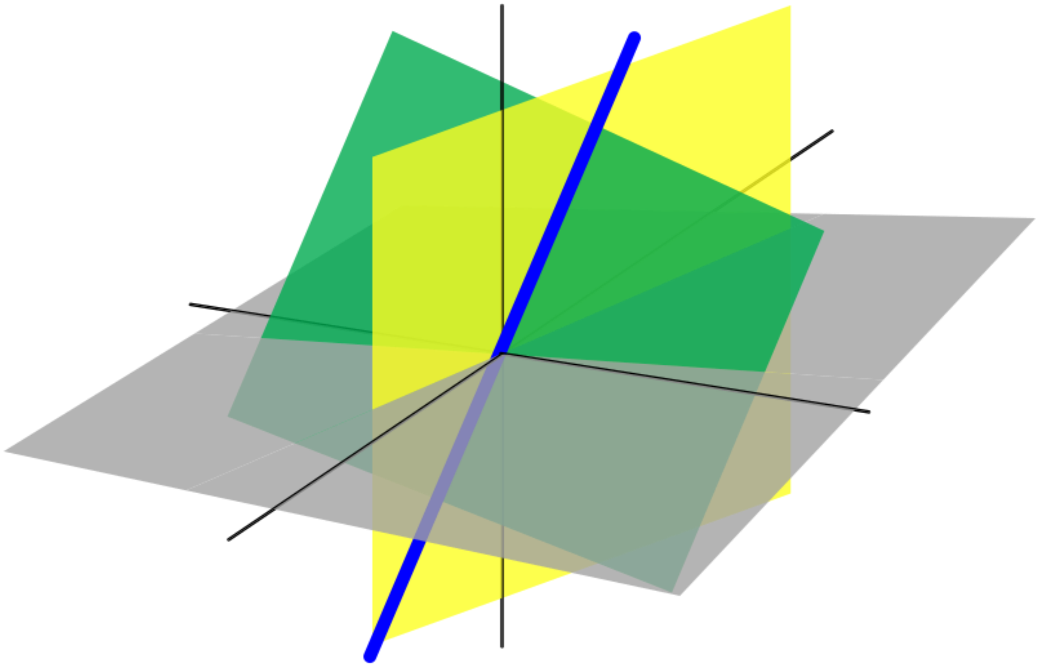
$$(x_1, \ldots, x_n) \mapsto a_1 x_1 + \ldots + a_n x_n,$$

- **Linear Algebra** regards **representation** of linear equations and linear functions into **vector spaces** and **matrices**.

LINEAR ALGEBRA

- A good **understanding** of Linear Algebra is **essential** for understanding and **working** with **many machine learnig algorithms** especially **deep learning**.

- In the three-dimensional **Euclidean Space**, three **planes** represent solutions of linear equations and their **intersection represents** the set of common **solutions**: in this case, a **unique point**. The blue line is the common solution of a **pair** of **linear equations**.



LINEAR ALGEBRA: Objects

- The study of Linea Algebra involves some **mathematical objects**:

1) **Scalars:** It is just a **single number**, we write them in italics, scalar names are usually lower-case and when we introduce them we specify which kind of scalar they are: **Real-valued** numbers, **Natural** numbers, etc.

2) **Vectors:** A vector is an array of numbers which are **arranged** in an **order**. We can **identify** each individual number by its **index** in that **ordering**. Vectors' names are **lower-case** typically and are written in **italics bold typeface** like ***x*.**

LINEAR ALGEBRA : Objects

- A typical **vector** is as follows: $$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- The elements of a vector are scalars, if they are **Real** and they are n elements then we say that belongs to R^n.

- We can think of a **vector** as a **point of space**, with each element giving the **coordinate along** a **different axis**.

LINEAR ALGEBRA: Objetcs

- **3) Matrices:** A **matrix** is a **2-D array** of numbers, so each element is identified by **two indices** instead of just one. We usually denote matrices with **upper-case variable names** with italics bold typeface, su as ***A.*** If a real-valued matrix ***A*** has an **height** of *m* and a **width** of *n* then we say that $\boldsymbol{A} \in R^{mxn}$**.** We usally denote an element at **row** *i* and **column** *j* as $\boldsymbol{A}_{i,j}$**.**

- ***4)*** **Transposition:** One **important operation** with matrices is the **transpose**. The transpose of a matrix can be seen as a **mirror image** across the **main diagonal**

$$\boldsymbol{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \Rightarrow \boldsymbol{A}^{\top} = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

LINEAR ALGEBRA: Objects

- Formally: $$(\boldsymbol{A}^{\top})_{i,j} = A_{j,i}.$$

- **Vectors** can be thought of as **matrices** that contain only **one column**. So the transpose of a vector is a **matrix** with only **one** row:

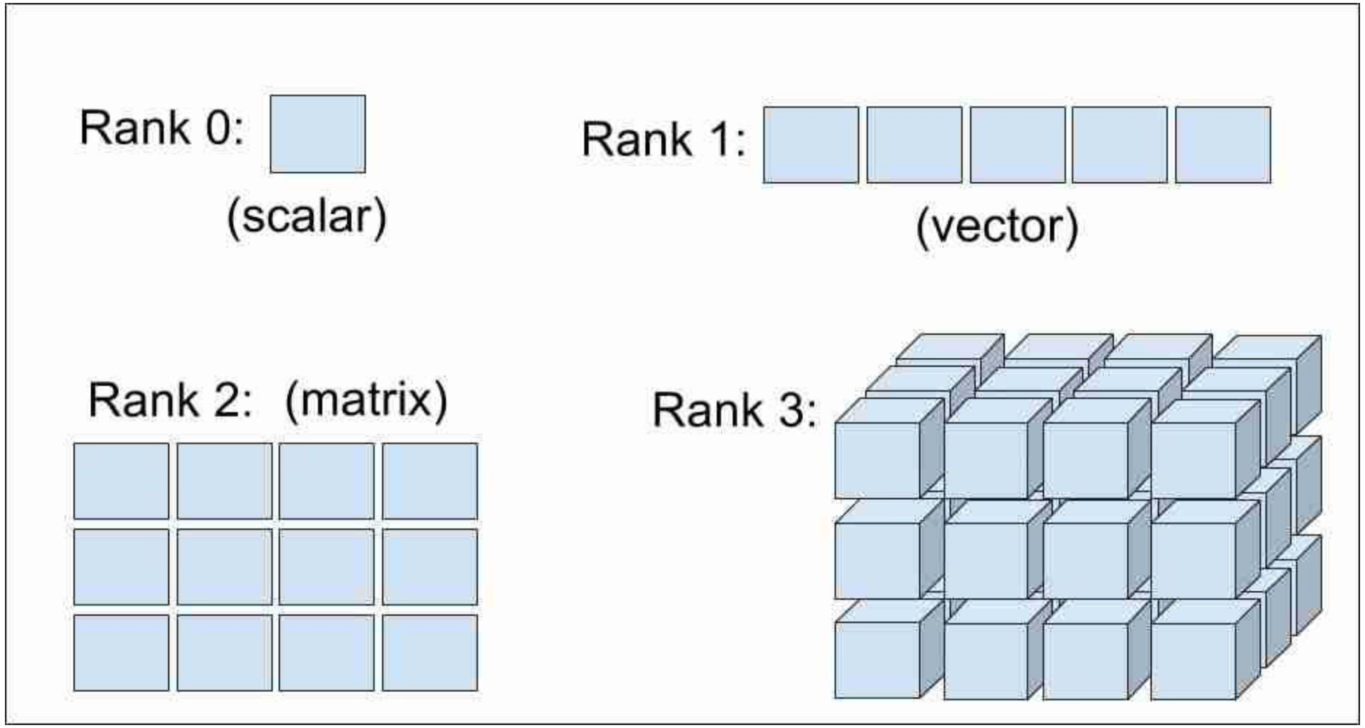
$$\boldsymbol{x} = [x_1, x_2, x_3]^{\top}$$

2	4	-1
-10	5	11
18	-7	6

2	-10	-18
4	5	-7
-1	11	6

LINEAR ALGEBRA: Objects

- **Tensors:** In some cases we will need an array with **more than two axis**. An array with numbers on a **regular grid** with a variable number of axis is known as a **tensor**. We denote a tensor with a non italics bold typeface **A.** Similarly to matrices **we depict elements** of a tensor **A** by writing $\boldsymbol{A}_{i,j,k}$.



LINEAR ALGEBRA: Operations

- **Addition:** We can **add matrices** to each other, as long as thet have the **same shape**, just by **adding** the **corresponding elements**:

$$C = A + B \longrightarrow C_{i,j} = A_{i,j} + B_{i,j}$$

$$\begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 3 \\ 5 & -1 \end{bmatrix} = \begin{bmatrix} 1+4 & 2+3 \\ -3+5 & 4+(-1) \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 5 \\ 2 & 3 \end{bmatrix}$$

LINEAR ALGEBRA: Operations

- **Addition and Multiplication scalar-matrix:** We can also **add** a **scalar** to a **matrix** or **multiply** a **matrix** by a **scalar**, just by performing that operation on **each element** of a matrix:

$$D = a \cdot B + c \longrightarrow D_{i,j} = a \cdot B_{i,j} + c$$

$$2 \;.\; \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 2.1 & 2.2 & 2.3 \\ 2.4 & 2.5 & 2.6 \\ 2.7 & 2.8 & 2.9 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{bmatrix}$$

LINEAR ALGEBRA: Operations

- **Multiplying Matrices and Vectors:** One of the most important operations involving matrices is multiplication of two matrices. The **matrix product** of matrices ***A*** and ***B*** is a third matrix ***C***. ***A*** must have the same number of columns as ***B*** has rows. In other words, if ***A*** is of shape ***m x n*** and ***B*** is of shape ***n x p***, the ***C*** is of shape ***m x p***. We can write the **matrix product** as follows:

$$C = AB \longrightarrow C_{i,j} = \sum_k A_{i,k} B_{k,j}.$$

LINEAR ALGEBRA : Operations

- **Element-wise product or Hadamard product:** it is the **product element by element** of two matrices of same size. Is is denoted as follows:

$$A \odot B$$

- **The dot produt between two vectors x and y:** is the matrix product

$$x^\top y$$

LINEAR ALGEBRA : Operations

- **We can think** of the matrix product ***C = AB*** as **computing** $C_{i,j}$ as the **dot product** between row *i* of ***A*** and column *j* of ***B.***
- **Matrix product properties:** it has many useful properties that make mathematical analysis of matrices convenient.

Distributive property $$A(B+C) = AB + AC$$

Associative property $$A(BC) = (AB)C$$

Matrix multiplication is *not* commutative.

- However, the **dot product** between two vectors is **commutative**. $$x^\top y = y^\top x.$$

LINEAR ALGEBRA: Operations

Matrix Product

$$\mathbf{m}_1^T \cdot \mathbf{m}_2 = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

$$= \begin{bmatrix} (a_{11}b_{11} + a_{21}b_{21}) & (a_{11}b_{12} + a_{21}b_{22}) & (a_{11}b_{13} + a_{21}b_{23}) \\ (a_{12}b_{11} + a_{22}b_{21}) & (a_{12}b_{12} + a_{22}b_{22}) & (a_{12}b_{13} + a_{22}b_{23}) \\ (a_{13}b_{11} + a_{23}b_{21}) & (a_{13}b_{12} + a_{23}b_{22}) & (a_{13}b_{13} + a_{23}b_{23}) \end{bmatrix}$$

Dot Product

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = a_1 b_1 + a_2 b_2 + a_3 b_3$$

LINEAR ALGEBRA : Equations

- A **system** of **linear equations** can be denoted as follows: ***Ax = b****.* Where $\boldsymbol{A} \in R^{mxn}$ is a **known matrix** and $\boldsymbol{b} \in R^{m}$. $\boldsymbol{x} \in R^{n}$ is a vector of **unknown** we would like to **solve for**.
- We can **rewrite** it as follows:

$$\boldsymbol{A}_{1,:}x = b_1$$

$$\boldsymbol{A}_{2,:}x = b_2$$

$$\dots$$

$$\boldsymbol{A}_{m,:}x = b_m$$

LINEAR ALGEBRA : Equations

- Or more explicitly:

$$\begin{aligned} \boldsymbol{A}_{1,1}x_1 + \boldsymbol{A}_{1,2}x_2 + \cdots + \boldsymbol{A}_{1,n}x_n &= b_1 \\ \boldsymbol{A}_{2,1}x_1 + \boldsymbol{A}_{2,2}x_2 + \cdots + \boldsymbol{A}_{2,n}x_n &= b_2 \\ &\cdots \\ \boldsymbol{A}_{m,1}x_1 + \boldsymbol{A}_{m,2}x_2 + \cdots + \boldsymbol{A}_{m,n}x_n &= b_m \end{aligned}$$

- Linear Algebra offers a powerful tool called **matrix inversion.** This tool allows us to analytically solve a system of linear equations for many values of ***A.***

LINEAR ALGEBRA : Equations

- To better describe **matrix inversion,** we first have to introduce the concept of an **Identity Matrix**. An Identity Matrix is a matrix that does not change any vector when we multiply that vector by that matrix. We denote the **Identity Matrix** that preserves n-dimensional vectors as $\boldsymbol{I}_n$.

- Formally, an **Identity Matrix,** $\boldsymbol{I}_n \in R^{nxn}$, and:

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x}$$

- The structure of the identity matrix is made of all entries along the main diagonal are 1, while all of the other entries are zero.

LINEAR ALGEBRA : Equations

- The **matrix inverse** of $\boldsymbol{A}$ is denoted as $\boldsymbol{A}^{-1}$ and it is defined as the matrix such that:

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$$

- Example of ***identity matrix*** and solution of a system of linear equations by ***matrix inversion***:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

$$\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

$$\boldsymbol{I}_n\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

LINEAR ALGEBRA : Equations

- When $\boldsymbol{A}^{-1}$ exists, several different algorithms for finding it in closed form. In theory, the same inverse matrix can then be used to solve the equations many times for different values of ***b***. In order for $\boldsymbol{A}^{-1}$ to exist, a system of linear equations must have exactly one solution for every value of ***b***.
- In order to analyze how many solutions the system of equations has, we can think of the columns of ***A*** as specyfing different directions we can travel from the origin (the point specified by the vector of all zeros), and determine how many ways there are of reaching ***b***.
- In this view, each element of ***x*** specifies how far we should travel in each of these directions, with $\boldsymbol{x}_i$ specifying how far to move in the directions of column *i*:

$$\boldsymbol{A}x = \sum_i x_i \boldsymbol{A}_{:,i}.$$

LINEAR ALGEBRA : Linear Dependence and Span

- **Linear Combination:** This kind of operation is called **Linear Combination.** Fomally, a linear combination of some set of vectors $\{v^{(1)}, v^{(2)}, \dots, v^{(n)}\}$ is given by multiplying each vector $v^{(i)}$ by a

  corresponding scalar coefficient and adding the results:

$$\sum_i c_i \boldsymbol{v}^{(i)}$$

- **Span:** the span of a set of vectors is the set of all points obtainable by linear combination of the original vectors. Determining whether **Ax = b** has a solution thus amounts to testing whether ***b*** is in the span of the columns of ***A.*** This particular span is known as the column space or the range of **A.**

LINEAR ALGEBRA : Norm

- **Norm:** Sometimes we need to measure the size of a vector. In machine learning, we usually measure the size of vectors using a function called **norm**. Formally, the L^p norm is given by:

$$||\boldsymbol{x}||_p = \sum_i |x_i|^p \Bigg)^{\frac{1}{p}}$$

  for $p \in \mathbb{R}, p \geq 1$.

- Norms, including the L^p norm, are functions mapping vectors to non-negative values. On an intuitive level, the norm of a vector ***x*** measures the distance from the origin to the point **x.**

LINEAR ALGEBRA : Norm

- The L^2 norm, with p = 2, is known as the **Euclidean Norm**. It is the Euclidean Distance from the origin to the point identified by **x.**

$$||\boldsymbol{x}||_1 = \sum_i |x_i|$$

- The L^2 norm is used so frequently in machine learning that it is often denoted simply by a $||\boldsymbol{x}||$.

LINEAR ALGEBRA : Type of Matrices

- ***Diagonal matrices:*** consist mostly of zeros and have non-zeros entries only along the main diagonal. Identity matrix is an example. We have *diag(v)* to denote a square diagonal matrix whose diagonal entries are given by entries of the vector *v.*

- ***Symmetric matrix:*** it is any matrix that is equal to its own transpose:

$$\boldsymbol{A} = \boldsymbol{A}^\top.$$

- ***Unit vector:*** it is a vector with **unit norm**:

$$||\boldsymbol{x}||_2 = 1$$

LINEAR ALGEBRA : Type of Matrices

- **Orthogonal matrix:** it is a square matrix whose rows are mutually orthonormal and whose columns are mutually orthonormal:

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}$$

$$\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$$

Francesco **Pugliese**