Machine Learning

- Linear Algebra Crash Course -

Matteo Matteucci, PhD (matteo.matteucci@polimi.it)
Artificial Intelligence and Robotics Laboratory
Politecnico di Milano

Linear Algebra … why bother?

Because linear algebra is everywhere!

- Optimization and Operational Research
- Statistics and Machine Learning
- (Digital) Signal Processing
- Control, System theory, Engineering, …
- …

Where could you find out more about it …

- www.youtube.com/playlist?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab
- https://en.wikipedia.org/wiki/Linear_algebra
- http://www.ekof.bg.ac.rs/wp-content/uploads/2016/09/Ponavljanje-matematike-Wayne-Winston-Operations-Research-Applications-and-Algorithms-4-edition.pdf

Vectors and Matrices

A vector is a list of numbers:

- It can be a column vector or row vector
- Numbers are called elements
- Dimensions are *"rows x columns"* (e.g., 5x1)
- Transpose operator exchanges rows and columns

$$\mathrm{v} = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 9.3 \\ -.01 \end{bmatrix}$$

$$\mathrm{v}^{\mathrm{T}} = \begin{bmatrix} 1 & -2 & 0 & 9.3 & -0.1 \end{bmatrix}$$

A Matrix is a bidimensional arrangement of numbers

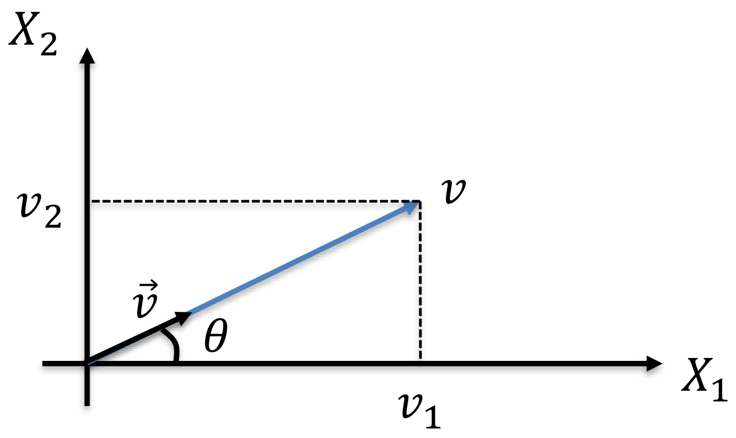
$$A = [a_{ij}]_{m \times n} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{vmatrix}$$

Vector Interpretation

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

$$\text{magnitude} = \|v\| = \sqrt{v_1^2 + v_2^2}$$

$$\text{orientation} = \theta = \tan^{-1}\left(\frac{v_2}{v_1}\right)$$



If $\|v\| = 1$ then v is a <u>unit</u> vector

- Sometimes named <u>versor</u>
- Defines the <u>direction</u> pointed by v

$$\vec{v} = \frac{v}{\|v\|} = \begin{bmatrix} v_1/\|v\| \\ v_2/\|v\| \end{bmatrix}$$

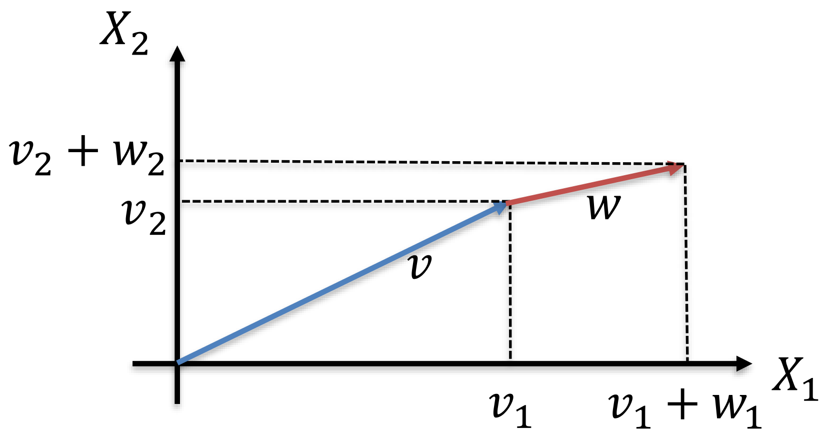
Basic Vector Operations

Sums and difference between vectors

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \qquad w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

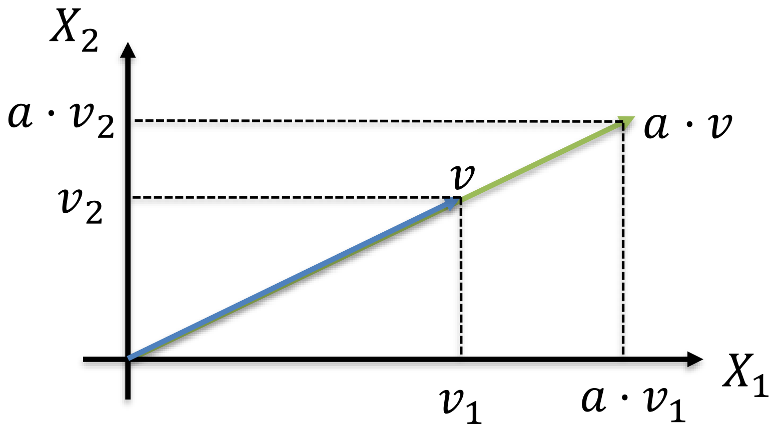
$$v + w = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \end{bmatrix}$$

$$v - w = \begin{bmatrix} v_1 - w_1 \\ v_2 - w_2 \end{bmatrix}$$



Scaling of a vector

$$a \cdot v = a \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} a \cdot v_1 \\ a \cdot v_2 \end{bmatrix}$$

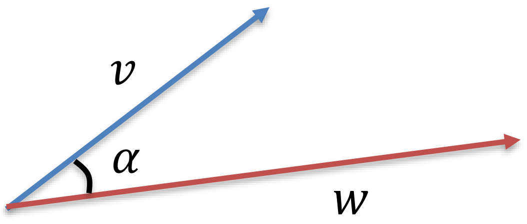


Vectors dot product (a.k.a., inner product or scalar product)

The dot/inner product is a <u>scalar</u>

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \qquad w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$v \cdot w = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = v_1 w_1 + v_2 w_2 = \|v\| \cdot \|w\| \cos \alpha$$



It is zero if the two vectors are orthogonal

$$if\ v \perp w, \qquad v \cdot w = \|v\| \cdot \|w\| \cos \alpha = 0$$

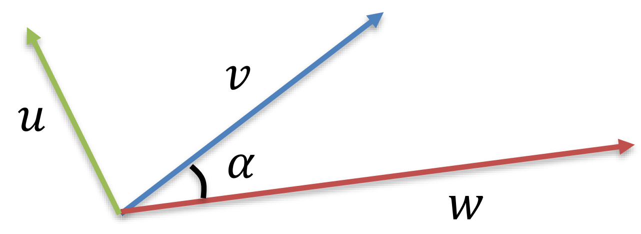
Vectors cross product

The cross product is a <u>vector</u>

$$u = v \times w$$

$$Magnitude: \|u\| = \|v \times w\| = \|v\|\|w\| \sin\alpha$$

$$Orientation: \begin{array}{l} u \perp v \Longrightarrow u \cdot v = (v \times w) \cdot v = 0 \\ u \perp w \Longrightarrow u \cdot w = (v \times w) \cdot w = 0 \end{array}$$



It is zero if the two vectors are parallel

$$if\ v \parallel w, \qquad v \times w = \|v\| \cdot \|w\| \sin\alpha = 0$$

Matrix Basics

Given two matrices with the same shape

$$A = [a_{ij}]_{m \times n} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{vmatrix} \qquad B = [b_{ij}]_{m \times n} = \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{vmatrix}$$

Sum is defined element wise

$$C = A + B = [c_{ij}]_{m \times n} = \begin{vmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn}{\_}b_{mn} \end{vmatrix}$$

Matrix Multiplication

Given two matrixes

$$A = [a_{ij}]_{m\times p} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mp} \end{vmatrix} \qquad B = [b_{ij}]_{p\times n} = \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{p1} & b_{p2} & \dots & b_{pn} \end{vmatrix}$$

Matrix multiplication is defined as

$$C = AB = [a_{ij}]_{m\times p}[b_{ij}]_{p\times n} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mp} \end{vmatrix} \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{p1} & b_{p2} & \dots & b_{pn} \end{vmatrix} = [c_{ij}]_{m\times n}$$

Check the shape pattern …

$$c_{ij} = row_i(A)col_j(B)$$

Transpose of a Matrix

Transpose swap rows and columns

$$C_{m\times n} = A^T_{n\times m}$$

$$c_{ij} = a_{ji}$$

Some useful identities exist

$$(A+B)^T = A^T + B^T$$

$$(AB)^T = B^T A^T$$

A matrix is symmetric if $A = A^T$

Matrix Multiplication Properties You Should Know

Indeed, might not even match as dimensions!

Commutative property (does not hold): $AB \neq BA$

Distributive properties (left + right): $A(B + C) = AB + AC$

$(B + C)D = BD + CD$

Product with a scalar: $c(A) = (A)c$

Transpose of a product: $(AB)^T = B^T A^T$

Associativity: $A(BC) = (AB)C$

https://en.wikipedia.org/wiki/Matrix_multiplication

Matrix Trace and Determinant

The determinant is computed from the element of a <u>square</u> matrix

$$A = [a_{ij}]_{n \times n};$$

$$\det(A) = \sum_{j=1}^{n} a_{ij} A_{ij}; \quad i = 1, \ldots, n;$$

$$A_{ij} = (-1)^{i+j} \det(M_{ij})$$

$$\det(AB) = \det(A)\det(B)$$

No need to bother right now …

- Trace $\quad A = [a_{ij}]_{n \times n}; tr[A] = \sum_{j=1}^{n} a_{jj}$

Matrix Inverse

Does not exist for all matrices, necessary (but not sufficient) that the matrix is square

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}, \det \mathbf{A} \neq 0$$

If $\det \mathbf{A} = 0$, $\mathbf{A}$ does not have an inverse.

Linear Independence

- A set of *n*-dimensional vectors $\mathbf{x_i} \in \mathbf{R}^n$, are said to be linearly independent if none of them can be written as a linear combination of the others.
- In other words,

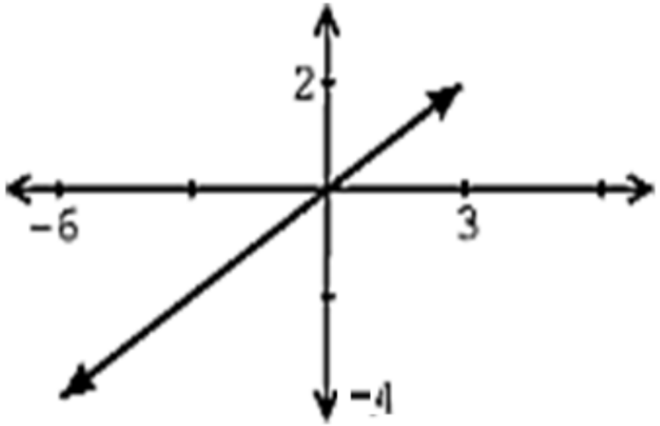
$$c_1 x_1 + c_2 x_2 + \ldots + c_k x_k = 0$$

$$\textit{iff} \qquad c_1 = c_2 = \ldots = c_k = 0$$

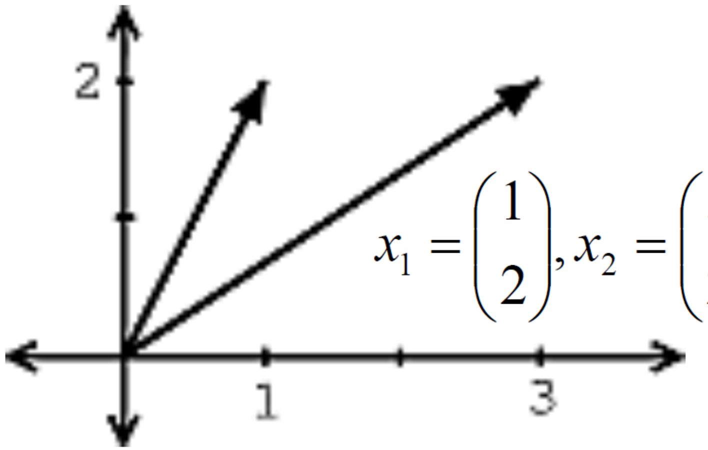
Linear Independence

- Another approach to reveal a vectors independence is by graphing the vectors.

$$x_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, x_2 = \begin{pmatrix} -6 \\ -4 \end{pmatrix}$$



Not linearly independent vectors



$$x_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, x_2 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

Linearly independent vectors

Span

- A span of a set of vectors $\mathbf{x_1}$, $\mathbf{x_2}$, … , $\mathbf{x_k}$ is the set of vectors that can be written as a linear combination of $\mathbf{x_1}$, $\mathbf{x_2}$, … , $\mathbf{x_k}$.

$$span(x_1, x_2, \ldots, x_k) = \{c_1 x_1 + c_2 x_2 + \ldots + c_k x_k \mid c_1, c_2, \ldots, c_k \in \Re\}$$

Basis

- A basis for $\mathbf{R}^n$ is a set of vectors which:
  - Spans $\mathbf{R}^n$, i.e. any vector in this *n*-dimensional space can be written as linear combination of these basis vectors.
  - Are linearly independent
- Clearly, any set of *n*-linearly independent vectors form basis vectors for $\mathbf{R}^n$.

Orthogonal/Orthonormal Basis

- An **orthonormal basis** of an a vector space V with an inner product, is a set of basis vectors whose elements are mutually orthogonal and of magnitude 1 (unit vectors).

- Elements in an **orthogonal basis** do not have to be unit vectors, but must be mutually perpendicular. It is easy to change the vectors in an orthogonal basis by scalar multiples to get an orthonormal basis, and indeed this is a typical way that an orthonormal basis is constructed.

- Two vectors are **orthogonal** if they are perpendicular, i.e., they form a right angle, i.e. if their inner product is zero.

$$a^T \cdot b = \sum_{i=1}^{n} a_i b_i = 0 \quad \Rightarrow \quad a \perp b$$

- The standard basis of the n-dimensional Euclidean space $\mathbf{R}^n$ is an example of orthonormal (and ordered) basis.

Matrix Eigenvalues and Eigenvectors

A eigenvalue λ and eigenvector $\mathbf{u}$ satisfies

$$\mathbf{Au} = \lambda \mathbf{u}$$

where $\mathbf{A}$ is a square matrix.

- Multiplying $\mathbf{u}$ by $\mathbf{A}$ scales $\mathbf{u}$ by λ

Matrix Eigenvalues and Eigenvectors

Rearranging the previous equation gives the system

$$\mathbf{Au} - \lambda\mathbf{u} = (\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = 0$$

which has a solution if and only if $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

- The eigenvalues are the roots of this determinant which is polynomial in λ.
- Substitute the resulting eigenvalues back into $\mathbf{Au} = \lambda\mathbf{u}$ and solve to obtain the corresponding eigenvector.