Mathematical Prerequisites

"Philosophy is written in that great book which ever lies before our eyes — I mean the universe — but we cannot understand it if we do not first learn the language and grasp the symbols, in which it is written. This book is written in the mathematical language, and the symbols are triangles, circles and other geometrical figures, without whose help it is impossible to comprehend a single word of it; without which one wanders in vain through a dark labyrinth."

Galileo Galilei, il Saggiatore, 1623

1) Linear Algebra.

2) Probability + Statistics.

3) Optimization.

References

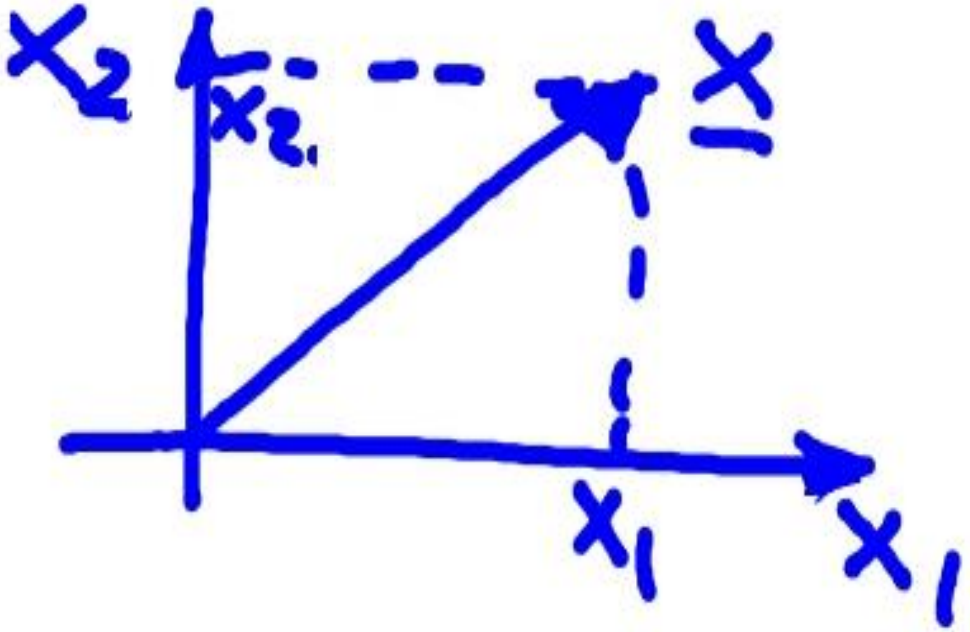
https://towardsdatascience.com/linear-algebra-and-probability-theory-review-for-ml-e3d2d70c5eb3

https://medium.com/@rohitrpatil/basic-linear-algebra-for-deep-learning-f537825b278f

https://mml-book.github.io/book/mml-book.pdf

1) Linear algebra.

$\mathbb{R}^2$



Vectors:

Generalize to any number of dimensions $\mathbb{R}^n$

Direction

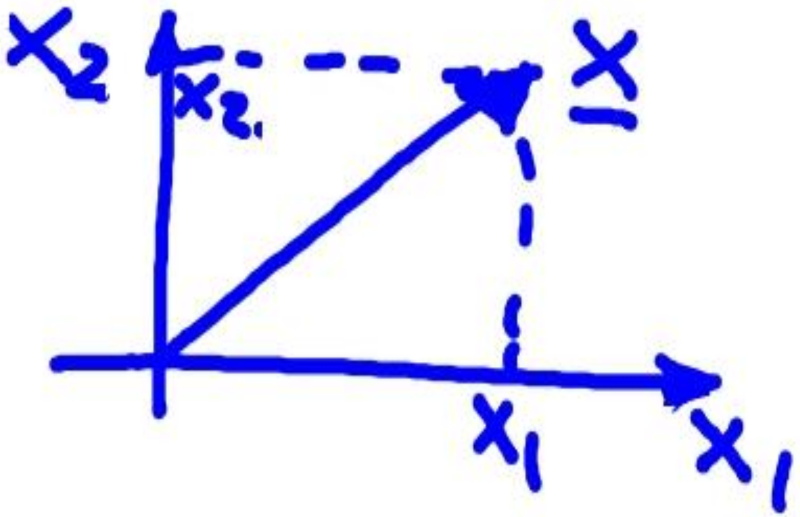
Length (magnitude).

$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ Components

Row vector: $(x_1\ x_2 \dots x_N)$

Transpose of $\begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$. $\underline{x}^T$

Inner product:

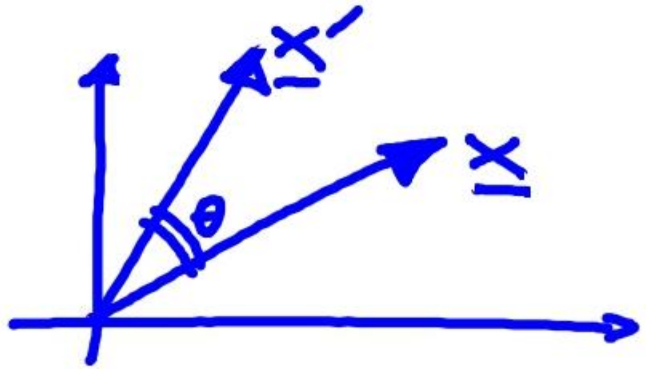


$$\underline{x}^T\underline{y} = (x_1 \cdots x_N)\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = x_1 y_1 + \cdots + x_N y_N.$$

$$\sqrt{\underline{x}^T\underline{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_N^2} \rightarrow \text{Length of } \underline{x}.$$

Norm

$\|\underline{x}\|$

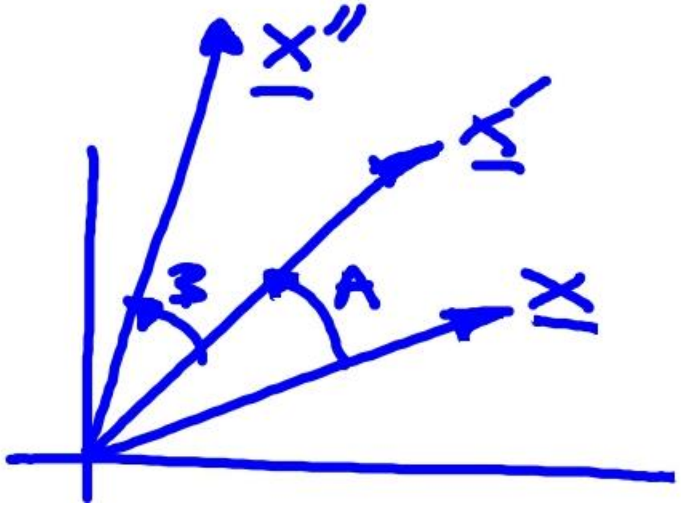


To obtain any other vector: Multiply it by a matrix A

$$
\underline{x}' = A\underline{x} = \underset{2\times 2}{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}} \underset{2\times 1}{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} = \underset{2\times 1}{\begin{pmatrix} \cos\theta\, x_1 - \sin\theta \cdot x_2 \\ \sin\theta\, x_1 + \cos\theta \cdot x_2 \end{pmatrix}}.
$$

Each element in the product:

Inner product of row in A × column of $\underline{x}$.



$$\underline{x}'' = B\,(A \cdot \underline{x})$$

Associativity.

$$\underline{x}'' = (B \cdot A) \cdot \underline{x}$$

To do $B \cdot A$, # columns in B
must be the same as # rows in A.

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix} = \begin{pmatrix} x_1 y_1 & x_1 y_2 & x_1 y_3 \\ x_2 y_1 & x_2 y_2 & x_2 y_3 \\ x_3 y_1 & x_3 y_2 & x_3 y_3 \end{pmatrix}.$$

$$\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

$$\begin{bmatrix} 5 & 3 \\ 7 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5\cdot 1 + 3\cdot 2 & 5\cdot 2 + 3\cdot 1. \\ \boxed{?} & \boxed{?} \end{bmatrix}$$

$$\underline{x}^T \underline{\theta}\, \underline{\theta}^T \underline{x} \quad \begin{cases} (\underline{x}^T \underline{\theta})(\underline{\theta}^T \underline{x}) \\ \underline{x}^T \underbrace{(\underline{\theta}\, \underline{\theta}^T)}_{\text{Matrix}} \underline{x} \end{cases}$$

Unit matrix : $\mathbb{I} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$

Inverse matrix of square matrix A: $AA^{-1} = A^{-1}A = \mathbb{I}$.

For 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $\quad A^{-1} = \dfrac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

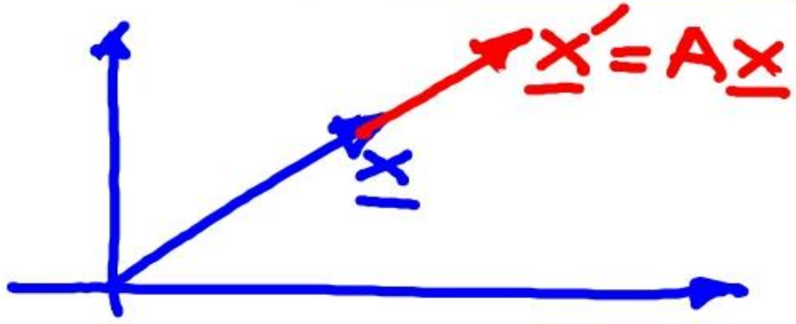
determinant of A

$$AA^{-1} = \frac{1}{ad-bc}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$= \frac{1}{ad-bc}\begin{pmatrix} ad-bc & \cancel{-ab+ba}^{\,0} \\ \cancel{cd-dc}_{\,0} & -bc+da \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Eigenvector of A:

$$\underline{x}' = A\underline{x} = \lambda \cdot \underline{x}$$



Eigenvalue of A.

We find eigenvalues by solving $\boxed{\det(A - \lambda \mathbb{I}) = 0.}$

$$A = \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix} \qquad \det(A - \lambda \mathbb{I}) = \det\begin{pmatrix} 5-\lambda & -2 \\ -2 & 2-\lambda \end{pmatrix}$$

$$= (5-\lambda)(2-\lambda) - 4 = \lambda^2 - 7\lambda + 6 = 0 \quad \begin{array}{l} \nearrow \lambda_1 = 6 \\ \searrow \lambda_2 = 1. \end{array}$$

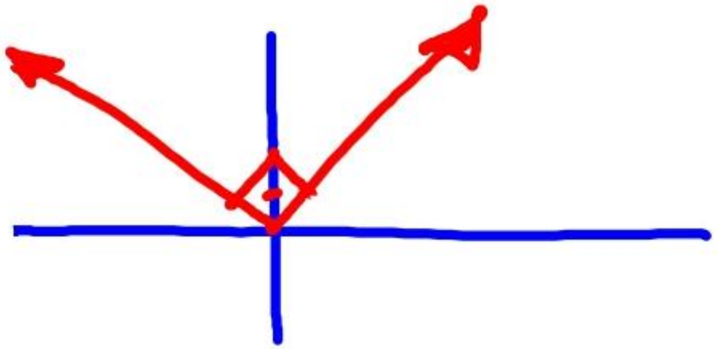
Eigenvectors:

For λ_1: $A\underline{x} = 6\underline{x} \Rightarrow \begin{pmatrix} 5 & -2 \\ -2 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 6\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$\Rightarrow 5x_1 - 2x_2 = 6x_1 \Rightarrow x_1 = -2x_2.$

$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2k \\ k \end{pmatrix}$ Normalize: $\|x\|^2 = 4k^2 + k^2 = 5k^2 = 1$

Eigenvector: $\underline{x}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 1 \end{pmatrix}$. For the other eigenvalue: $\underline{x}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}$



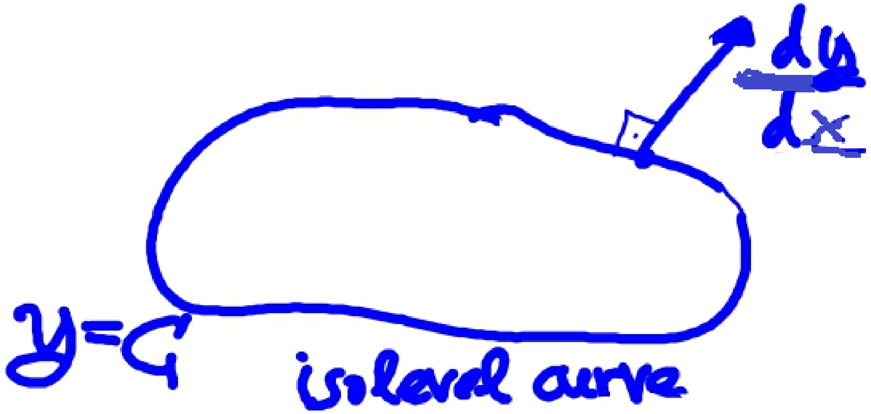
For all symmetric matrices:
Eigenvalues real.
Eigenvectors orthogonal.

Vector-matrix differentiation.

$$\frac{d\mathbf{y}}{dx} = \begin{pmatrix} \frac{dy_1}{dx} \\ \frac{dy_2}{dx} \\ \vdots \\ \frac{dy_N}{dx} \end{pmatrix}$$

$$\frac{dy}{d\mathbf{x}} = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_N} \end{pmatrix}$$

gradient of y.



$$\frac{d(\underline{a}^T\underline{x})}{d\underline{x}} = \frac{d(a_1x_1 + a_2x_2)}{d\underline{x}} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \underline{a}.$$

$$\frac{d(A^T\underline{x})}{d\underline{x}} = A$$

$$\frac{d(\underline{x}^T A \underline{x})}{d\underline{x}} = (A^T + A)\underline{x}$$

For symmetric A: $2A\underline{x}$