Foundation of Machine Learning

Revisiting Prerequisite of Maths: Matrix Theory

Sayanta Adhikari

MTech(RA)
Department of Artificial Intelligence
Indian Institute of Technology, Hyderabad

Table of Contents

Table of Contents

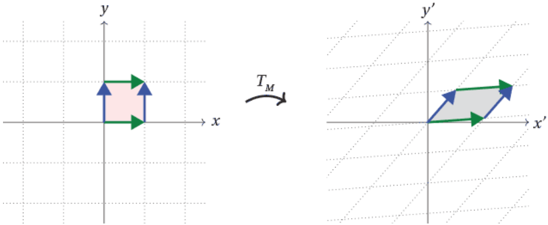
Definition

With $m, n \in \mathbb{N}$ a real-valued (m,n) matrix $\mathbf{A}$ is a collection of elements $a_{ij}, i = 1, \cdots, m,\ j = 1, \cdots, n$ which is ordered according to a rectangular scheme consisting of m rows and n columns.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} a_{ij} \in \mathbb{R} \tag{1}$$

Interpretation

A matrix can be visualized as a function that performs **Linear Transformation**. A linear function $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ can be represented as a matrix $T_M \in \mathbb{R}^{m \times n}$.



Matrix changes in two ways:

- by Stretching
- by Rotating

Vector Spaces

- Its a set of all possible vectors (collection of real numbers) that is closed under two operation: *finite summation* and *scalar multiplication*. By closed, I mean that when we preform that operation on the elements of the space the output is also part of that space.
- One of the example of vector space is $\mathbb{R}^n$. This is the one we would be considering by default throughout this session.

Basis Vectors

- **Basis Vectors Set**: Given a Vector Space V, its is the smallest possible set of linearly independent vectors from the same vector space, such that all the vectors in the vector space can be represented as a linear combination of these vectors.

$$\{b_1, b_2, \cdots, b_n\} \textit{ be the basis vectors for vector space } V$$

$$v \in V \textit{ then},$$

$$v = \sum_{i=1}^{n} \alpha_i b_i, \textit{ where } \alpha_i \in \mathbb{R}$$

- Basis Vectors Set is not unique, there can be multiple sets of basis vectors but for every set the number of elements in the set will remain constant, known as the **dimension** of the vector space.

Table of Contents

The inner product is a very general function that takes two vectors and maps them to a real number, and this mapping follows some particular properties. Here, we will discuss a particular inner product known as **dot product** or **scalar product**.

Dot Product:

Its a function $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.
Given $x, y \in \mathbb{R}^n$
$dot(x, y) = x^T y = \sum_{i=1}^{n} x_i y_i$

- Dot product can be seen as taking the projection of one vector onto another unit-length vector.
- Orthogonal vectors dot product is zero.

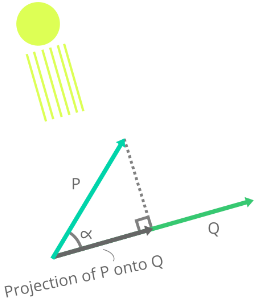


Table of Contents

Vector Norms

A **norm** on $\mathbb{R}^n$ is a function $\|.\| : \mathbb{R}^n \to \mathbb{R}^+$ which assigns each vector with a positive $\mathbb{R}$ number such that the mapping holds the following properties:

- Absolutely Homogeneous: $\|\lambda x\| = |\lambda|\|x\|$
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$
- Positive definite: $\|x\| \geq 0$ and $\|x\| = 0 \Leftrightarrow x = 0$

A specific kind of norm on $\mathbb{R}^n$ is L_p Norm. It is defined for all values of $p \geq 1$ as:

$$\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}} \tag{2}$$

For $p = 2$, it becomes the well known *Euclidean Norm*.
Generally, norms are thought of a measure of length for vectors.

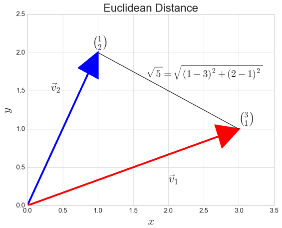
Length of a vector is nothing but the norm of the vector.
Euclidean Length/Distance:

$$\|x\|_2 := \sqrt{x^T x} \tag{3}$$

This gives us a relation between L_2-norms and dot product.

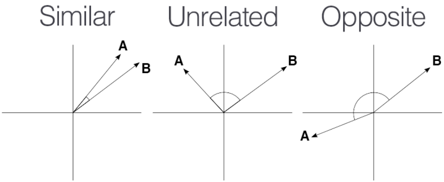
Table of Contents

Euclidean Distance: $d(x, y) := \|x - y\| = \sqrt{(x-y)^T(x-y)}$



Cosine Distance/Similarity:

$$cos(\theta) = \frac{x^T y}{\|x\|_2 \|y\|_2} \tag{4}$$



Both give us a notion of the difference between the two vectors.

Table of Contents

Rank

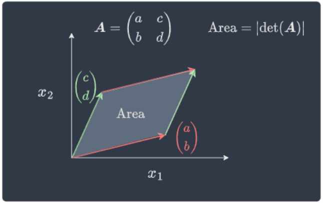
The number of linearly independent columns of a matrix $A \in \mathbb{R}^{m \times n}$ equals the number of linearly independent rows and is called the **rank** of A. Rank of $A \in \mathbb{R}^{m \times n}$ can be at-most be equal to $min(m, n)$.

Determinant

The determinant of a square matrix $A \in \mathbb{R}^{n \times n}$ is a function that maps A onto a real number.
For a square matrix $A \in \mathbb{R}^{n \times n}$, it holds that A is invertible if and only if $det(A) \neq 0$.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{5}$$

$$det(A) = \frac{1}{ad - bc} \tag{6}$$



The absolute value of the Determinant gives us the volume under the parallelepiped, i.e., in 2-D, a parallelogram.

Trace

The trace of a square matrix $A \in \mathbb{R}^{n \times n}$ is defined as the sum of the diagonal elements of A.

$$tr(A) := \sum_{i=1}^{n} a_{ii} \tag{7}$$

Properties:

- $tr(A + B) = tr(A) + tr(B)$ for $A, B \in \mathbb{R}^{n \times n}$
- $tr(\alpha A) = \alpha tr(A)$, $\alpha \in \mathbb{R}$ for $A \in \mathbb{R}^{n \times n}$
- $tr(I_n) = n$
- Trace is invariant under cyclic permutations, i.e., $tr(AKL) = tr(KLA)$ for matrices $A \in \mathbb{R}^{a \times k}$, $K \in \mathbb{R}^{k \times l}$, $L \in \mathbb{R}^{l \times a}$.

Table of Contents

Eigenvalues & Eigenvectors

Consider a square matrix $A \in \mathbb{R}^{n \times n}$. There exists a eigenvalue $\lambda \in \mathbb{R}$ and corresponding eigenvectors $x \in \mathbb{R}^n \backslash \{0\}$ such that they satisfy the following equation (known as *eigenvalue equation*):

$$Ax = \lambda x$$

Characteristic Polynomial: $\det(A - \lambda I_n) = 0$ is a characteristic polynomial for matrix A and is denoted by $p_A(\lambda)$. $\lambda \in \mathbb{R}$ is a eigenvalue of $A \in \mathbb{R}^{n \times n}$ if and only if λ is a root of this polynomial.

Examples of Eigenvalues & Eigenvectors

$$A_1 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} cos(\frac{\pi}{6}) & -sin(\frac{\pi}{6}) \\ sin(\frac{\pi}{6}) & cos(\frac{\pi}{6}) \end{bmatrix}$$

$$A_4 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$A_5 = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

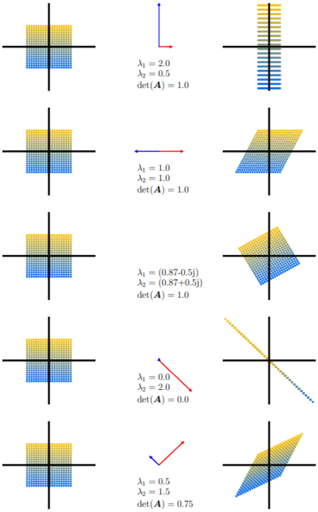
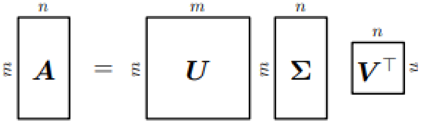


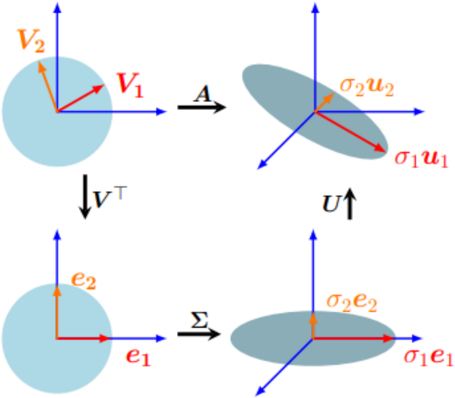
Table of Contents

Singular Value Decomposition

Singular Value Decomposition (SVD) Theorem: Let $A^{m \times n}$ be a rectangular matrix of rank $r \in [0, min(m, n)]$. The SVD of A is a decomposition of form

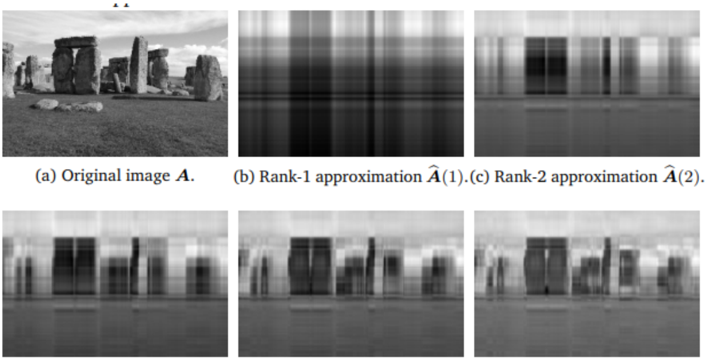


with two orthogonal matrix $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$. Moreover, Σ is an $m \times n$ matrix with $\Sigma_{ii} = \sigma_i \geq 0$ and $\Sigma_{ij} = 0, i \neq j$.
These diagonal entries $\sigma_i, i = 1, 2, \cdots, r$, of Σ are called singular values. And by conversion, singular values are ordered, i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$.



- Decomposition of a Linear Transformation (A) into a sequence of three specific linear transformations (U, V, Σ).
- U and V are matrices that only do rotations.
- Σ is a matrix that performs scaling and augmentation(or reduction) in dimensionality.

Example of SVD



(a) Original image $\boldsymbol{A}$. (b) Rank-1 approximation $\widehat{\boldsymbol{A}}(1)$. (c) Rank-2 approximation $\widehat{\boldsymbol{A}}(2)$.

(d) Rank-3 approximation $\widehat{\boldsymbol{A}}(3)$. (e) Rank-4 approximation $\widehat{\boldsymbol{A}}(4)$. (f) Rank-5 approximation $\widehat{\boldsymbol{A}}(5)$.

Figure: Image Reconstruction with SVD

The Original Image requires $1432 \times 1910 = 2735120$ numbers where as the Rank-5 Approximation requires a total of $5 \times (1432 + 1910 + 1) = 16715$ numbers - just 0.6% i=of the original.

References

- Most topics are taken from **Mathematics for Machine Learning, Marc Peter**. PDF LINK
- **BOOK:** Introduction to Linear Algebra, Gilbert Strang, 5th Edition
- Essence of Linear Algebra, 3Blue1Brown, Playlist Link
- Understand Linear Algebra. Gilbert Strang Lectures
- SVD Visualization First 6 videos only.