Euclidian Norm, Inequalities and Angle

Linear Algebra

Department of Computer Engineering

Sharif University of Technology

Hamid R. Rabiee rabiee@sharif.edu

Maryam Ramezani maryam.ramezani@sharif.edu

Overview

- Introduction
- Inequalities
- Euclidean Norm
- Euclidean Distance
- Angle

Introduction

The reason to use norms

- Machine learning uses vectors, matrices, and tensors as the basic units of representation
- Two reasons to use norms:

1. To estimate how **big** a vector/matrix/tensor is
    - How big is the difference between two tensors is
2. To estimate how **close** one tensor is to another
    - How close is one image to another

Euclidean Norm

Definition

For $v \in V$, we define the norm of v, denoted $||v||$, by:

$$||v|| = \sqrt{< v, v >}$$

Example

Norm of $P_n(x)$ in the term of inner product $\langle p_n(x), q_n(x) \rangle = \int_0^1 p_n(x) q_n(x) dx$:

$$||P_n(x)|| = \sqrt{\int_0^1 P_n^2(x) dx}$$

Euclidean Norm

Definition

- Euclidean Norm (2−norm, l_2 norm, length)

$$||x|| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} = \sqrt{x^T x}$$

  - A vector whose length is 1 is called a unit vector
  - **Normalizing**: divide a non−zero vector by its length which is a unit vector in the same direction of original vector
  - It is a nonnegative scalar
  - In $\mathbb{R}^2$ follows from the Pythagorean Theorem.
  - What about $\mathbb{R}^3$?
  - What is the shape of $||x||_2 = 1$?

Inequalities

Root Mean Square Value (RMS)

Definition

Mean-square (MS) value of n-vector x is:

$$\frac{x_1^2+x_2^2+\ldots+x_n^2}{n} = \frac{||x||^2}{n}$$

Root-mean-square value (RMS)

$$rms(x) = \sqrt{\frac{x_1^2+x_2^2+\ldots+x_n^2}{n}} = \frac{||x||}{\sqrt{n}}$$

The RMS value of a vector x is useful when comparing norms of vectors with different dimensions. rms(x) gives "typical" value of $|x_i|$

Example

$rms(1) = 1$ (independent of n)

if all the entries of a vector are the same, (a) then the RMS value of the vector is $|a|$

Theorem

Suppose that k of the numbers $|x_1|, |x_2|, \dots, |x_n|$ are $\geq a$ then k of the numbers $x_1^2, x_2^2, \dots, x_n^2$ $are \geq a^2$

So $||x||^2 = x_1^2 + x_2^2 + \ \dots + x_n^2 \geq ka^2$ $so\ we\ have\ k \leq \frac{||x||^2}{a^2}$

Number of x_i with $|x_i| \geq a$ is no more than $\frac{||x||^2}{a^2}$

Question

- What happens when $\frac{||x||^2}{a^2} \geq n$?
- No entry of a vector can be larger in magnitude than the norm of the vector. (why?)

Chebyshev Inequality

Important

Chebyshev inequality is easier to interpret in terms of the RMS value of a vector.

$$\frac{k}{n} \leq \left(\frac{rms(x)}{a}\right)^2$$

Example

How many entries of x can have value more than $5rms(x)$?

Important

The Chebyshev inequality partially justifies the idea that the RMS value of a vector gives an idea of the size of a typical entry: It states that not too many of the entries of a vector can be much bigger (in absolute value) than its RMS value

Standard Deviation

Theorem

- For n−vector x, $avg(x) = 1^T(\frac{x}{n})$
- De−meaned vector is $\tilde{x} = x - avg(x)1$ $\quad\quad (so, avg(\check{x}) = 0)$
- Standard deviation of x is:

$$std(x) = rms(\check{x}) = \frac{\left\|x - \left(\frac{1^T x}{n}\right)1\right\|}{\sqrt{n}}$$

- $Std(x)$ gives "typical" amount x_i vary from $avg(x)$
- $Std(x) = 0$ only if $x = \alpha 1$ for some α
- A basic formula

$$rms(x)^2 = avg(x)^2 + std(x)^2$$

Chebyshev Inequality for std

Theorem

x is an $n-vector$ with mean $avg(x)$, standard deviation $std(x)$

Rough idea: most entries of x are not too far from the mean
By Chebyshev inequality, fraction of entries of x with $|x_i - avg(x)| \geq \alpha\, std(x)$ is no more than

$$\frac{1}{\alpha^2}\ (for\ \alpha > 1)$$

- The fraction of entries of x within θ standard deviations of $avg(x)$ is at least $(1 - \frac{1}{\theta^2})$ for $\theta > 1$

Vector Standardization

Definition

$$z = \frac{1}{std(x)}\,(x - avg(x)\mathbf{1}).$$

- It has $mean\ (\mu) = 0\ and\ std(\sigma) = 1$
- Its entries are sometimes called the z−scores associated with the original entries of x.
- The standardized values for a vector give a simple way to interpret the original values in the vectors.

Cauchy-Schwartz Inequality

Theorem

For two n-vectors a and b, $|a^T b| \leq \ ||a||\,||b||$

Written out:

$$|a_1 b_1 + \cdots + a_n b_n| \leq (a_1^2 + \ \ldots + a_n^2)^{\frac{1}{2}} (b_1^2 + \ \ldots + b_n^2)^{\frac{1}{2}}$$

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right) \left(\sum_{i=1}^{n} y_i^2\right)$$

Proof:

Cauchy–Schwartz Inequality

Theorem

For two n–vectors a and b, $|a^T b| \leq ||a||\,||b||$

Written out:

$$|a_1 b_1 + \cdots + a_n b_n| \leq (a_1^2 + \ \ldots + a_n^2)^{\frac{1}{2}} (b_1^2 + \ \ldots + b_n^2)^{\frac{1}{2}}$$

$$(\sum_{i=1}^{n} x_i y_i) \leq (\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i^2)$$

It is clearly true if either a or b is 0.

So, assume $\alpha = ||a||$ and $\beta = ||b||$ are non–zero

We have

$$\begin{aligned} 0 &\leq \left|\left|\beta a - \alpha b\right|\right|^2 \\ &= \left|\left|\beta a\right|\right|^2 - 2\,(\beta a)^T(\alpha b) + \left|\left|\alpha b\right|\right|^2 \\ &= \beta^2 \left|\left|a\right|\right|^2 - 2\,\beta\alpha(a^T b) + \alpha^2 \left|\left|b\right|\right|^2 \\ &= 2\,\left|\left|a\right|\right|^2 \left|\left|b\right|\right|^2 - 2\,\left|\left|a\right|\right| \left|\left|b\right|\right| (a^T b) \end{aligned}$$

Divide by 2 $\left|\left|a\right|\right|\left|\left|b\right|\right|$ to get $a^T b \leq \left|\left|a\right|\right|\left|\left|b\right|\right|$

Apply to $-a, b$ to get other half of Cauchy–Schwartz inequality.

Cauchy–Schwarz inequality holds with equality when one of the vectors is a multiple of the other

If and only if a and b are linear dependent

Complexity

- **Norm:** $2n$ flops
  - $O(n)$
- **RMS:** $2n$ flops
  - $O(n)$
- **Distance:** $3n$ flops
  - $O(n)$
- **Angle:** $6n$ flops
  - $O(n)$
- **Standardizing:** $5n$ flops
  - $O(n)$
- **Correlation Coefficient:** $10n$ flops
  - $O(n)$
- **Standard Deviation:** $4n$ flops
  - $O(n)$
  - Can reduce to $3n$ flops.

$$std(x)^2 = rms^2 - avg(x)^2$$

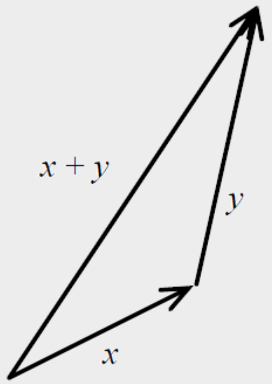
Triangle Inequality

Theorem

Consider a triangle in two or three dimensions:

$$||x+y|| \leq ||x|| + ||y||$$

Proof:

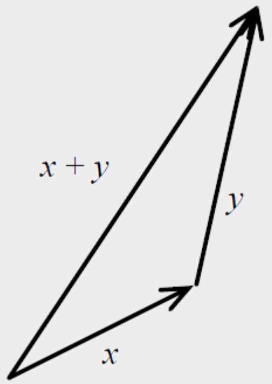


Triangle Inequality

Theorem

Consider a triangle in two or three dimensions:

$$||x + y|| \le ||x|| + ||y||$$



Verification of triangle inequality:

$$||x + y||^2 = ||x||^2 + ||y||^2 + 2\, x^T y$$

$$\le ||x||^2 + ||y||^2 + 2\, ||x||\,||y||$$

$$= (||x|| + ||y||)^2$$

$$\Rightarrow ||x + y|| \le ||x|| + ||y||$$

Cauchy–Schwartz Inequality

Euclidean Norm

Vector Norm Properties

Important Properties:

1. Absolute Homogenity / Linearity:

$$||\alpha x|| = |\alpha| \, ||x||$$

2. Subadditivity / Triangle Inequality:

$$||x + y|| \leq ||x|| + ||y||$$

3. Positive definiteness / Point separating:

$$if \ ||x|| = 0 \ then \ x = 0$$
$$(from\ 1 \ \& \ 3): For\ every\ x, ||x|| = 0 \ iff \ x = 0$$

4. Non-negativity:

$$||x|| \geq 0$$

Proof:

Norm of sum

Theorem

If x and y are vectors:

$$||x + y|| = \sqrt{||x||^2 + 2\, x^T y + ||y||^2}$$

Proof:

$$||x + y||^2 = (x + y)^T (x + y)$$

$$= x^T x + x^T y + y^T x + y^T y$$

$$= ||x||^2 + 2\, x^T y + ||y||^2$$

Theorem

Take any inner product $\langle\cdot,\cdot\rangle$ and define $f(x) = \sqrt{\langle x, x\rangle}$. Then f is a norm.

Proof

Note

Every inner product gives rise to a norm, but not every norm comes from an inner product. (Think about norm 2 and norm max)

Norm of block vectors

Important

Suppose a,b,c are vectors:

$$\left\|\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right\|^2 = a^T a + b^T b + c^T c = ||a||^2 + ||b||^2 + ||c||^2$$

So, we have

$$\left\|\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right\| = \sqrt{||a||^2 + ||b||^2 + ||c||^2} = \left\|\begin{bmatrix} ||a|| \\ ||b|| \\ ||c|| \end{bmatrix}\right\|$$

(Parse RHS very carefully!)

❖ The norm of a stacked vector is the norm of the vector formed from the norms of sub-vectors.

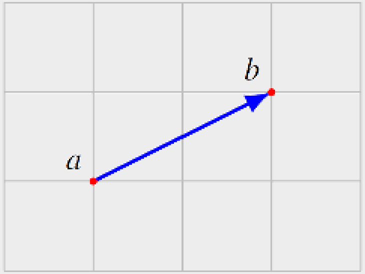
Euclidean Distance

Euclidean Distance

Distance between two n−vectors shows the vectors are "close" or "nearby" or "far".

Distance:

$$dist(a,b) = ||a - b||$$



RMS deviation between the two vectors:

$$rms(a - b) \;=\; \frac{||a-b||}{\sqrt{n}}$$

Comparing Norm and Distance

Norm

(Normed Linear Space)

1. $||x - y|| \geq 0$
2. $||x - y|| = 0 \Rightarrow x = y$
3. $||\lambda(x - y)|| = |\lambda| \, ||x - y||$

Distance Function

(Metric Space)

1. $dist(x, y) \geq 0$
2. $dist(x, y) = 0 \Rightarrow x = y$
3. $dist(x, y) = dist(y, x)$

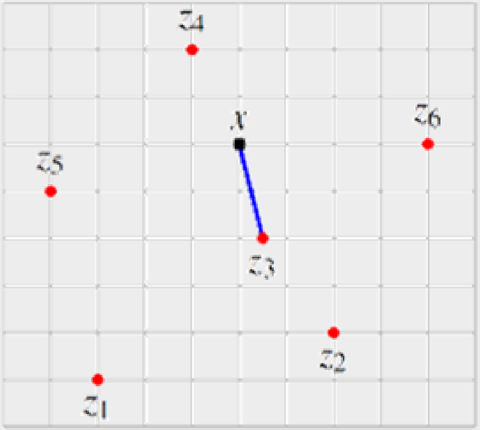
Feature Distance and Nearest Neighbors:

$$if\ x, y\ are\ feature\ vectors\ for\ two\ entities, ||x - y||\ is\ the\ feature\ distance$$

$if\ z_1, z_2, \ldots, z_m\ is\ a\ list\ of\ vectors, z_j is\ the\ nearest\ neighbor\ of\ x\ if$:

$$||x - z_j|| \leq ||x - z_i||, \qquad i = 1, 2, \ldots, m$$

Number of flops and order?



Angle

Angle

Definition

Angle between two non−zero vectors a, b is defined as:

$$\angle(a, b) = \arccos\left(\frac{a^T b}{||a||\,||b||}\right)$$

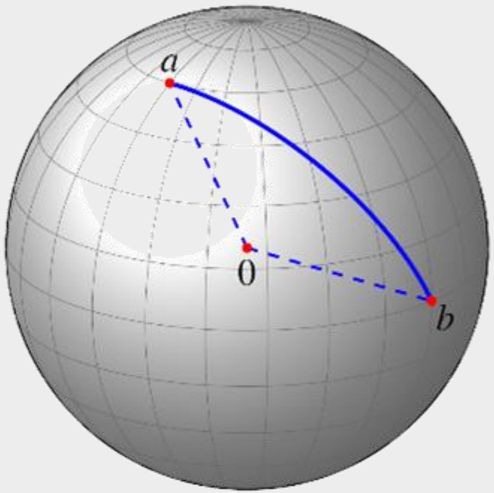
$\angle(a, b)$ is the number in $[0, \pi]$ that satisfies:

$$a^T b = ||a||\,||b||\cos(\angle(a, b))$$

Coincides with ordinary angle between vectors in 2D and 3D

Spherical distance:

$$\text{if } a, b \text{ are on sphere with radius } R, \text{distance along the sphere is } R\ \angle(a,b)$$



References

- Chapter 1: Advanced Linear and Matrix Algebra, Nathaniel Johnston
- Chapter 6: Linear Algebra David Cherney
- Linear Algebra and Optimization for Machine Learning
- Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares