Euclidian Norm and Inequalities

CE282: Linear Algebra

Computer Engineering Department

Sharif University of Technology

Hamid R. Rabiee

Maryam Ramezani

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

Root Mean Square Value (RMS)

Definition

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

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

Root-mean-square value (RMS)

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

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

$rms(\mathbf{1}) = 1$ (independent of n)
if all the entries of a vector are the same, (a) then the RMS value of the vector is $|a|$

Chebyshev Inequality

Theorem

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

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

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

Question

- What happens when $\frac{\left||x|\right|^2}{a^2} \geq n$?
- 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 $5rms(x)$?

Chebyshev Inequality

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

Definition

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

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

  - $Std(x)$ gives "typical" amount x_i vary from $avg(x)$
  - $Std(x) = 0$ only if $x = \alpha 1$ for some α
  - Greek letters μ , σ commonly used for mean, standard deviation
  - 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 $\theta > 1$

Properties of std

Important

1. ***Adding a constant:*** For any vector x and any number a, we have $std(x + a1) = std(x)$. Adding a constant to every entry of a vector doesn't change its standard deviation.

2. ***Multiplying by a scalar:*** For any vector x and any number a, we have
$std(ax) = |a|\ std(x)$. Multiplying a vector by a scalar, multiplies the standard deviation by the absolute value of the scalar.

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:

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

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

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

We have

$$0 \leq ||\beta a - \alpha b||^2$$
$$= ||\beta a||^2 - 2\,(\beta a)^T(\alpha b) + ||\alpha b||^2$$
$$= \beta^2 ||a||^2 - 2\,\beta\alpha\left(a^T b\right) + \alpha^2 ||b||^2$$
$$= 2\,||a||^2 ||b||^2 - 2\,||a||\,||b||\left(a^T b\right)$$

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

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

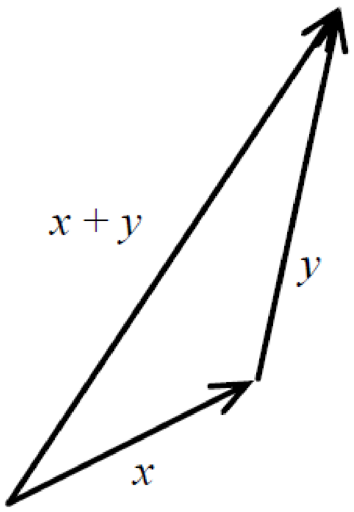
Cauchy-Schwarz inequality holds with equality when one of the vectors is a multiple of the other

Triangle Inequality

Theorem

Consider a triangle in two or three dimensions:

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



Verification of triangle inequality:

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

$$\leq ||x||^2 + ||y||^2 + 2\,\underline{||x||\,||y||}$$

Cauchy-Schwartz Inequality

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

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

Euclidean Norm

Definition

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

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

  - It is a nonnegative scalar
  - In $\mathbb{R}^2$ follows from the Pythagorean Theorem.
  - What about $\mathbb{R}^3$?
  - What is the shape of $||x||_2 = 1$?

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:

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

Inner product and norm

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 sum

Theorem

If x and y are vectors:

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

Proof:

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

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

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

Norm of block vectors

Important

Suppose a,b,c are vectors:

$$||(a,b,c)||^2 = a^T a + b^T b + c^T c = ||a||^2 + ||b||^2 + ||c||^2$$

So, we have

$$||(a,b,c)|| = \sqrt{||a||^2 + ||b||^2 + ||c||^2} = \big|\big|(||a||, ||b||, ||c||)\big|\big|$$

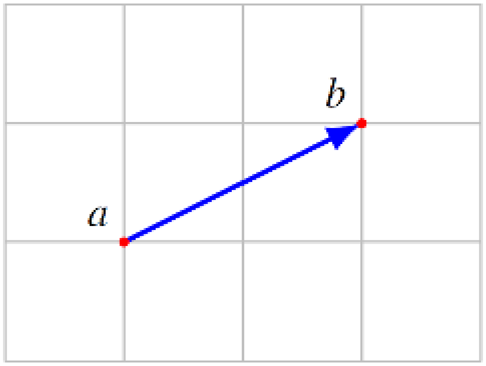
(Parse RHS very carefully!)

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

Euclidean Distance

Distance:

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



RMS deviation between the two vectors:

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

Euclidean Distance

Distance between two n-vectors shows the vectors are “close” or “nearby” or “far”.

Example

$$u = \begin{bmatrix} 1.8 \\ 2 \\ -3.7 \\ 4.7 \end{bmatrix}, v = \begin{bmatrix} 0.6 \\ 2.1 \\ 1.9 \\ -1.4 \end{bmatrix}, w = \begin{bmatrix} 2.0 \\ 1.9 \\ -4.0 \\ 4.6 \end{bmatrix}$$

The distance between pairs of them are:

$$||u - v|| = 8.368, \qquad ||u - w|| = 0.387, \qquad ||v - w|| = 8.533$$

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. $dist(x, y) = dist(y, x)$

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:

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

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

Complexity

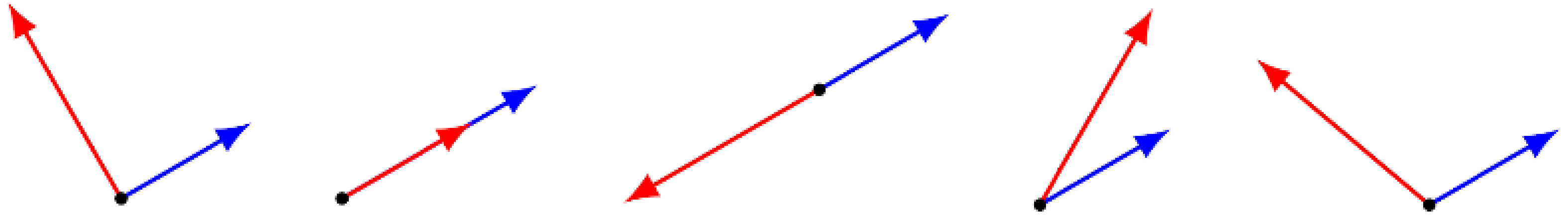
- **Norm:** $2n$ flops
  - $O(n)$
- **RMS:** $2n$ flops
  - $O(n)$
- **Distance:** $3n$ flops
  - $O(n)$
- **Angle:** $6n$ flops
  - $O(n)$
- **Standardizing:** $5n$ flops
  - $O(n)$
- **Correlation Coefficient:** $10n$ flops
  - $O(n)$
- **Standard Deviation:** $4n$ flops
  - $O(n)$
  - Can reduce to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

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

Classification of Angles

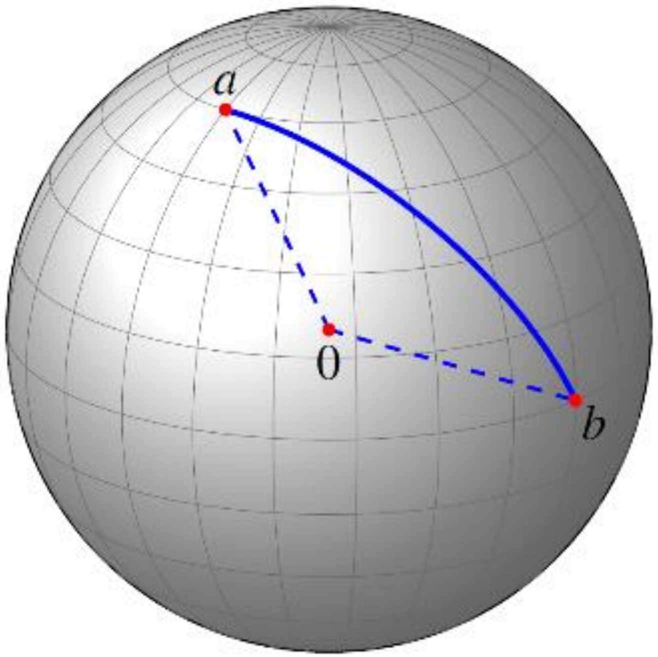
$$\theta = \angle(a, b)$$

- $\theta = \frac{\pi}{2} = 90°$: a and b are orthogonal, written $a \perp b$ $(a^T b = 0)$
- $\theta = 0$: a and b are aligned $\left(a^T b = ||a||\,||b||\right)$
- $\theta = \pi = 180°$: a and b are anti-aligned $\left(a^T b = -||a||\,||b||\right)$
- $\theta \leq \frac{\pi}{2} = 90°$: a and b make an acute angle $\left(a^T b \geq 0\right)$
- $\theta \geq \frac{\pi}{2} = 90°$: a and b make an obtuse angle $(a^T b \leq 0)$



Spherical distance:

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



Applications

Correlation Coefficient:

$$We\ have\ vectors\ a, b\ and\ de-meaned\ vectors\ \tilde{a} = a\ - avg(a)1\ , \tilde{b} = b\ - avg(b)1$$

$$correlation\ coefficient\ \left(between\ a, b\ with\ \tilde{a}\ \neq 0, \tilde{b}\ \neq 0\right)\ is\ \rho = \frac{\tilde{a}^T \tilde{b}}{||\tilde{a}||\ \left|\left|\tilde{b}\right|\right|}$$

$$\rho = \cos\angle(\tilde{a},\ \tilde{b})$$

- $\rho = 0 \Rightarrow a, b\ are\ uncorrelated$
- $\rho > 0.8\ (or\ so) \Rightarrow a, b\ are\ highly\ correlated$
- $\rho < -0.8\ (or\ so) \Rightarrow a, b\ are\ highly\ anti-correlated$

Very roughly: highly correlated means $a_i\ and\ b_i$ are typically both above (below) their means together.

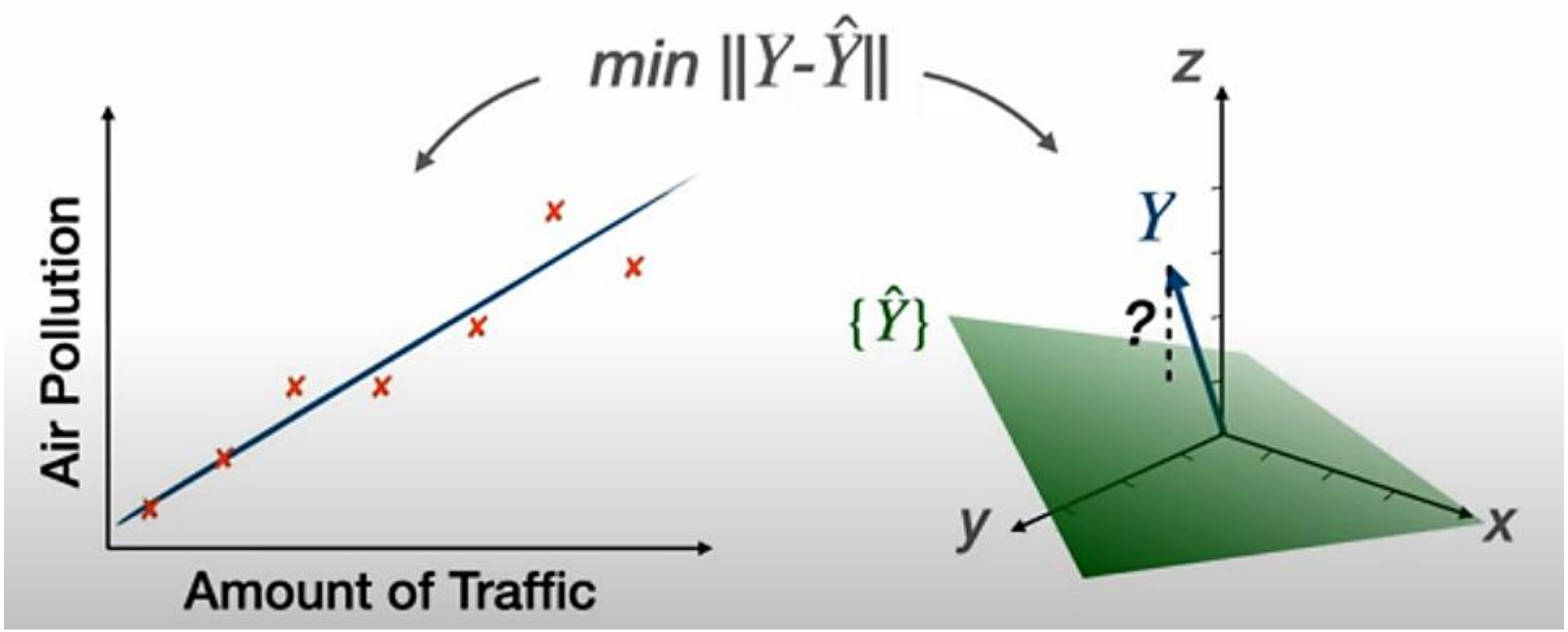
Document dissimilarity by angles

- measure dissimilarity by angle of word count histogram vectors
- pairwise angles (in degrees) for 5 Wikipedia pages shown below

	Veterans Day	Memorial Day	Academy Awards	Golden Globe Awards	Super Bowl
Veterans Day	0	60.6	85.7	87.0	87.7
Memorial Day	60.6	0	85.6	87.5	87.5
Academy A.	85.7	85.6	0	58.7	85.7
Golden Globe A.	87.0	87.5	58.7	0	86.0
Super Bowl	87.7	87.5	86.1	86.0	0

ML Application

The best linear regression model comes from choosing the closest $\hat{Y}$ *to* Y based on

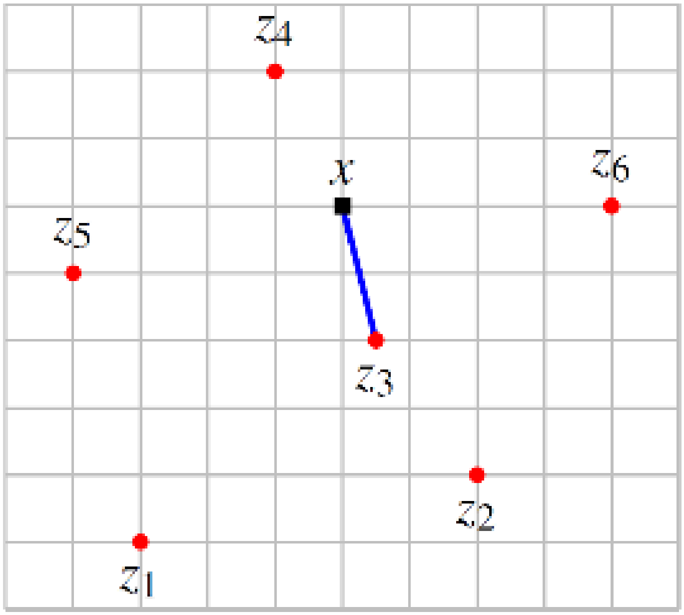


Feature Distance and Nearest Neighbors:

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

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

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



❖ Number of flops and order?

Reference

- Linear Algebra and Its Applications David C. Lay
- Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares
- https://www.youtube.com/watch?v=76B5cMEZA4Y