# Norm & Distances

**Norm:** Euclidean norm of an n-vector $x$ is $\|x\|$ is square root of sum of squares.

$$\boxed{\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}} \quad \text{or} \quad \boxed{\|x\| = \sqrt{x^T x}}$$

- When $x$ is scalar, (1-vetor) Euclidean norm $= |x|$ (absolute value of $x$)

Norm of a vector - numerical measure of its magnitude.

Small vector → vector with norm as smaller number. large vector. vice versa.

**Properties of Norm:** $x, y$ → vectors of same size ; $\beta$ → scalar.

1) Non negative homogeniety : $\boxed{\|\beta x\| = \cancel{\|\beta\| |x|} \; |\beta| \, \|x\|}$
   - multiplying vector by a scalar multiplies the norm by abs. value of scalar.

2) Triangle inequality : $\boxed{\|x+y\| \le \|x\| + \|y\|}$

3) Non negetivity : $\boxed{\|x\| \ge 0}$

4) Definiteness : $\boxed{\|x\| = 0 \text{ iff } x = 0}$

} +ve definiteness

General Norm: any real valued function of an n-vector that satisfies above 4.

$$\boxed{\mathrm{rms}(x) = \sqrt{\frac{x_1^2 + \dots + x_n^2}{n}} = \frac{\|x\|}{\sqrt{n}}}$$

- RMS value of a vector is a way to measure the "typical" size of its entries.

Norm of a sum: $\boxed{\|x\| + \|y\| = \sqrt{\|x\|^2 + 2x^T y + \|y\|^2}}$

Norm of block vectors: These are basically vectors inside vectors.

$$\boxed{\|d\|^2 = d^T d = a^T a + b^T b + c^T c = \|a\|^2 + \|b\|^2 + \|c\|^2}$$

The norm of a stacked vector is the norm of the vector formed from the norms of the subvectors.

Chebyshev Inequality: $x$ = n-vector ; $a > 0$.

'k' entries of $x$ satisfy $|x_i| \ge 0$

Then k of its entries follows $x_i^2 \ge a^2$

This follows that, $\boxed{\|x\|^2 = x_1^2 + \dots + x_n^2 \geq k a^2}$

since k of the numbers in the sum are atleast $a^2$ and n-k are non negative.

$$\boxed{\text{chebyshev inequality} = k \leq \frac{\|x\|^2}{a^2}}$$

- This inequality basically helps us understand how many entries in a vector can be 'large' compared to the rest.
  - k no. of entries that are larger than a certain value a.
  - if a is bigger than vector size; then no entry can be larger than the norm.
  - if we pick a that is bigger than typical size of no. in vectors, then the inequality says, k, the no. of large entries, will be small or even 0.

**Euclidean Distance:** $\boxed{\text{dist}(a,b) = \|a-b\|}$

rms (a-b) is the RMS deviation between a and b.

- this is basically saying how far apart they are on average.

Triangle inequality: triangle with vertices at positions a, b, c

→ edge lengths are $\|a-b\|$, $\|b-c\|$, $\|c-a\|$

by the triangle inequality: $\|a-c\| = \|(a-b)+(b-c)\| \leq \|a-b\| + \|b-c\|$

i.e, third edge length is no longer than sum of other 2.


- Feature distance: x & y → feature vectors, then: $\|x-y\|$ = feature distance.
  - This gives measure of how diff the obj are (in terms of fea. values)

- RMS prediction error: $y$ → time series of some quantity.

  $\hat{y}$ → estimation or prediction of time series $y$.

  smaller the value the better.

  Then $y - \hat{y}$ = prediction error & $\text{rms}(y-\hat{y})$ = rms prediction error.

- Nearest neighbour: $z_1 \dots z_m$ = collection of m n-vectors. ; $x$ = another n-vec.

  if $\|x - z_j\| \leq \|x - z_i\|$, $i = 1, \dots, m$.

  then we can say $z_j$ is the nearest neighbour (closest vector) to $x$.

**Heterogenous Vector Entries:**

square of distance between 2 n-vectors x & y is given by:

$$\boxed{\|x-y\|^2 = (x_1-y_1)^2 + \dots + (x_n-y_n)^2}$$

- This gives equal importance to every feature in the vector.

Ex: if you are comparing 2 objects, diff in 1 feature (weight) is treated just as important as diff in another feature (height) when calc overall distance.

- **same units for features:** This method works well when all the features (entries in vector) represent the same type of quantity in the same units.

Ex: if you are ~~counting~~ comparing word count in doc, where each entry is a count of how often a word appears, it makes sense to treat each word count equally.

- **different units for features:** if you are comparing house size in sqmr & bedrooms no. you have to be careful because 1 feature has big values (size) & than the other, it can distort the distance calculation.

* **choose units carefully:** # scale the units.

Ex: if you are comparing houses : house size (given in 1000's) (so 1600 becomes 1.6)
no. of bedrooms will remain an integer.

→ by doing so, both features have similar magnitude making it easy for comparison.

House 1 : (1.6, 2)
House 2 : (1.5, 2)
House 3 : (1.6, 4)

} small difference ∴ similar
} large diff in bedrooms ∴ Not similar.

- without this scaling, we will get errors saying 1 & 2 are far & 2 & 3 are close.

## Standard Deviation:

Associated De-meaned vector : $\tilde{x} = x - \text{avg}(x)\mathbf{1}$

(Here we subtract the avg (x) from every entry of x)

- This is useful in understanding how the entries in the original vector, deviate from $\bar{x}$

$$\boxed{\text{std}(x) = \sqrt{\frac{(x_1-\text{avg}(x))^2 + \dots + (x_n - \text{avg}(x))^2}{n}} = \frac{\|x - (\mathbf{1}^T x/n)\mathbf{1}\|}{\sqrt{n}}}$$

- std is 0 only if all entries are ~~e.~~ equal.
- std is small, when the entries of the vector are nearly same.

avg, rms & std : $\boxed{\text{rms}(x)^2 = \text{avg}(x)^2 + \text{std}(x)^2}$

Ex: Return time series: n-vector : represents returns (as %) of invest over n Time periods.

- mean return → avg of the vector. (simply called the return)
- risk → std of the vector measures how much the returns vary from p to p.
- risk-return plot → multiple invest can be compared by plotting y axis (mean return) & x-axis (risk)
- desirable investments have ↑ return & low risk.

Chebyshev inequality for standard deviation:

→ This inequality is a mathematical tool that helps estimate how many entries in a set of data can deviate significantly from the avg. value.

fraction of entries in the dataset that deviate $\bar{x}$ by more than 'a'

$$\boxed{\frac{K}{n} \le \left[\frac{std(x)}{a}\right]^2}$$

Ex: lets take a dataset with returns on an investment. avg $\bar{x}$ = 8% & risk = 3%.

By using the chebyshev inequality, we can estimate how many periods might result in a loss (return ≤ 0%)

- Here, we set a = 8%, because we are interested in how far returns can dev from the mean.

$$\therefore \left(\frac{3}{8}\right)^2 = \frac{9}{64} = 0.141 = 14.1\%$$

This means that at most, 14.1% of periods can have a return either below 0% or above 16%.

Properties of Standard Deviation:

1. Adding a constant : $std(x + a1) = std(x)$
2. Multiplying by a scalar : $std(ax) = |a|\, std(x)$

Standardization: Standardized version of $x$ :

$$\boxed{z = \frac{1}{std(x)}(x - avg(x)1)}$$

This has mean value as 0 & std value as 1.

→ these entries are called the z-scores.

$x_y = 1.4$ (means that $x_y$ is 1.4 stds away from the mean of entries of $x$)

Ex: $x$ → gives values of some medical test of n patients admitted to the hospital, the standardized values of z-scores tells us how low or high ..

Ex: $x_{32} = -3.2$, very low measurement.

$x_{22} = 0.3$, quite close to the avg value.

**Angle:** The cauchy - schwaz inequality: $\boxed{|a^T b| \leq \|a\|\|b\|}$, expanded, it is:

$$|a_1b_1 + \dots + a_nb_n| \leq (a_1^2 + \dots + a_n^2)^{1/2} (b_1^2 + \dots + b_n^2)^{1/2}$$

(1) Zero case: If either vector a/b is '0', then both sides of inequality are 0.
$\therefore$ the inequality trivially holds.

(2) Non-zero case: $\alpha = \|a\|$ ; $B = \|b\|$ (represent magnitudes of a & b)

observation: $\|Ba - \alpha b\|^2 \geq 0$ always, $\because$ norm of any vector is non-ve.

$$\Rightarrow \quad B^2\|a\|^2 + \alpha^2\|b\|^2 - 2\alpha B\, a^T b \geq 0$$

$$\Rightarrow \quad \|B\|^2\|a\|^2 + \|\alpha\|^2\|b\|^2 - 2\|a\|\|b\| a^T b \geq 0$$

$$\Rightarrow \quad 2\|a\|^2\|b\|^2 - 2\|a\|\|b\|\, a^T b \geq 0$$

$$\Rightarrow \quad 2\|a\|\|b\| (\|a\|\|b\| - a^T b)$$

$$\Rightarrow \quad \|a\|\|b\| \geq a^T b \quad \Rightarrow \quad |a^T b| \leq \|a\|\|b\|$$ cauchy Schwarz

**Angle between 2 vectors:** Angle between 2 non-zero vectors a, b:

$$\angle(a,b) = \theta \quad \boxed{\cos\theta = \frac{a^T b}{\|a\|\|b\|}} \qquad \arccos\theta \in [0, \pi]$$

- This is a symmetric function : $\angle(a,b) = \angle(b,a)$
- Scaling with +ve value has no effect : $\angle(\alpha a, Bb) = \angle(a,b)$

Acute & obstuse Angles:

- orthogonal vectors: $a^T b = 0$ ; which means $\theta = \pi/2 = 90°$.
- aligned vectors: $a^T b = \|a\|\|b\|$, which means $\theta = 0$.
- anti-aligned vectors: $a^T b = -\|a\|\|b\|$, which means $\theta = 180°$.
- acute angles: $\angle(a,b) < 90°$ (inner product is +ve value)
- obtuse angle: $\angle(a,b) > 90°$. (inner product is -ve value)
- document dissimilarity via angles: If n-vectors x & y represent word counts for 2 documents, $\angle(x,y)$ can be used as measure of dissimilarity.
  → either word counts / histograms can be used.

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2x^T y.$$
$$= \|x\|^2 + \|y\|^2 + 2\|x\|\|y\|\cos\theta$$

→ $\theta = 0°$ $\|x+y\| = \|x\| + \|y\|$

→ $\theta = 90°$ $\|x+y\| = \sqrt{\|x\|^2 + \|y\|^2}$ ← pythagorean theorem.

**Correlation coefficient:** (measures how closely 2 sets of data vary together.

Step 1: Demeaning the vectors $\longrightarrow$ $\tilde{a} = a - \text{avg}(a)\,1$, $\tilde{b} = b - \text{avg}(b)\,1$ } this step centers them aro^n.

Step 2: correlation coefficient:

$$\boxed{\rho = \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\| \|\tilde{b}\|}}$$

← dot product of the vectors, which measures how closely their entires align.

← product of lengths which normalizes the result.

This is equivalent to $\cos\theta$. ; $\theta$ is small == correlation is high.

Step 3: You can also express the correlation using standardized vectors: (vectors divided by their standard deviation)

$$u = \frac{\tilde{a}}{\text{std}(a)}, \quad v = \frac{\tilde{b}}{\text{std}(b)} \qquad \rho = \frac{u^T v}{n}$$ ← length of vectors.

Range of $\rho$: Cauchy-Swarz ensures that the value lies between $-1$ & $1$.

$\rho = 1$, when vectors are perfectly aligned (+ve multiples of each other)

$\rho = -1$, vectors are anti-aligned (-ve multiples of each other)

$\rho = 0$, vectors are uncorrelated. (don't show linear relationship)

Ex: **Standard Deviation of Sum of 2 vectors:**

The formula for the above is: $\boxed{\text{std}(a+b) = \sqrt{\text{std}(a)^2 + 2\rho\,.\,\text{std}(a)\,\text{std}(b) + \text{std}(b)^2}}$

when $\rho = 1$: $\text{std}(a+b) = \text{std}(a) + \text{std}(b)$ (vectors clearly correlated)

when $\rho = 0$: $\text{std}(a+b) = \sqrt{\text{std}(a)^2 + \text{std}(b)^2}$ (vectors are uncorrelated)

when $\rho = -1$: $\text{std}(a+b) = |\text{std}(a) - \text{std}(b)|$ (vectors perfectly negetively correlated)

**Hedging Investments:**

- Applied in finance where 2 assets a & b are considered, both having same ~~return (avg)~~ average return ($\mu$) & risk (std) ($\sigma$). The correlation is denoted by $\rho$.

① Blended investment: 50% of each asset has return time series: $c = \frac{a+b}{2}$

② Average return: $\text{avg}(c) = \text{avg}\left(\frac{a+b}{2}\right) = \frac{\text{avg}(a) + \text{avg}(b)}{2} = \mu$

③ Risk (std): $\text{std}(c) = \sigma . \sqrt{\frac{1+\rho}{2}}$

Units for Heterogenous Vector Entries:

* choose units such that the typical values of different entries in the vector are of similar magnitude.

→ This ensures each entry contributes fairly to metrics like correlation or std.

## Complexity:

- Norm of a veltor: 1. n multiplications to square each entry.
  2. n−1 additions to sum the squared entries.
  3. One square root (computationally expensive) ∴ Total : 2n flops.

- RMS value : same as computing norm (except div by $\sqrt{n}$ which takes 2 additional ")
- Distance betw. vectors : subtracting corresponding elements + squaring differences + summing them & taking square root. ∴ 3n flops.

- Angle between 2 vectors : 6n flops • Demeaning n-vector: 2n flops
- Standard Deviation: 1. Demeaning the vector (2n flops)
  2. computing RMS of demeaned vector (2n flops
  } 4n flops.

→ To ↑ efficiency to 3n flops we can use formula: $std(x) = \sqrt{rms(x)^2 - avg(x)^2}$

- Standardizing n-vector: 5n flops • correlation coefficient: 10n flops.

- Nearest neighbour search:
  1. computec the distance between 2 vectors (3n flops)
  2. ∴ computing distance between x & all K vectors is 3Kn flops.