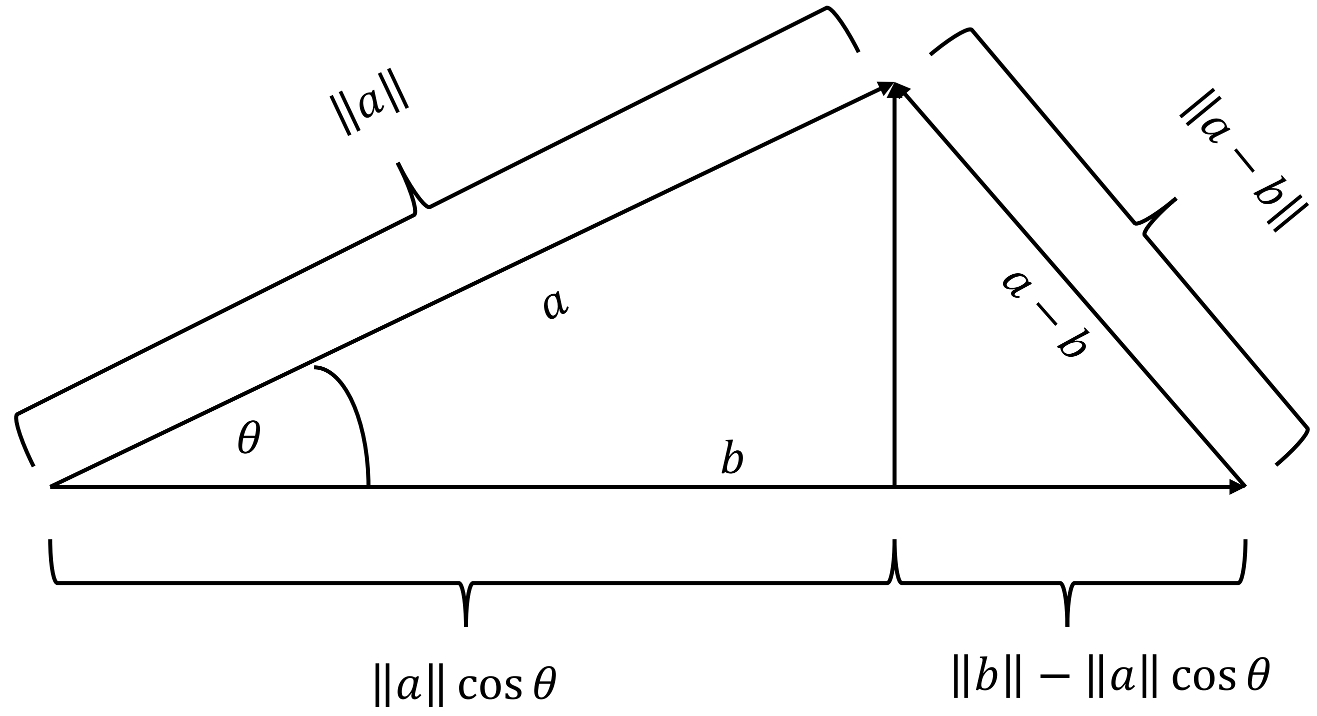
Cosine similarity


$\|a\|$
$\|a-b\|$
a
$a-b$
θ
b
$\|a\|\cos\theta$
$\|b\|-\|a\|\cos\theta$

- We want to prove that $\cos\theta = \frac{a \cdot b}{\|a\|\|b\|}$

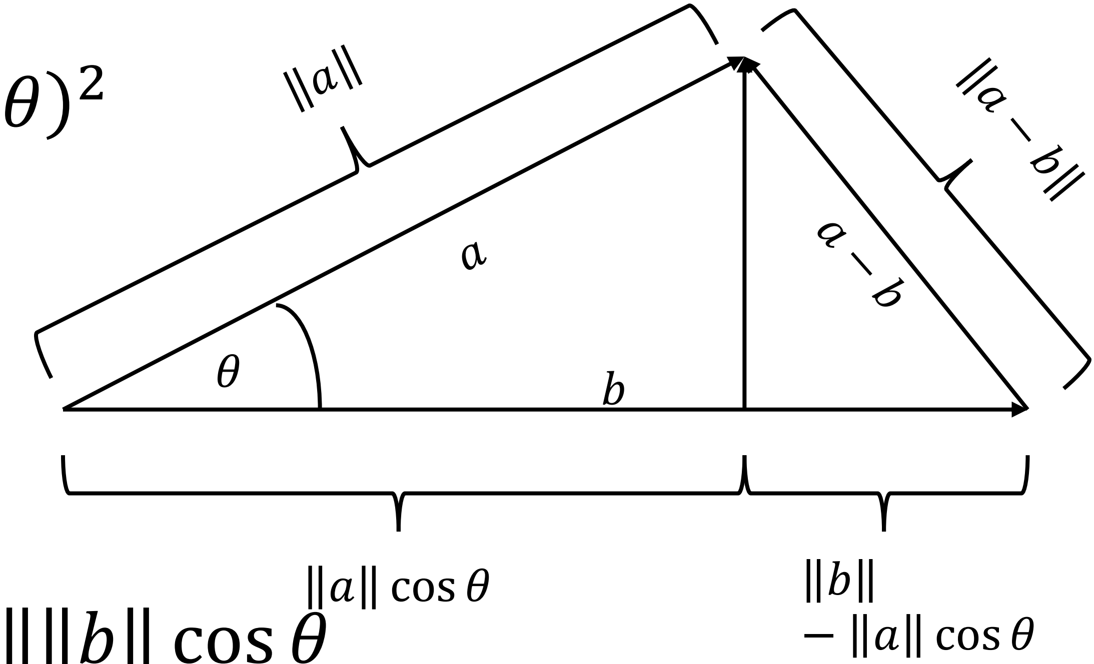
$$\|a-b\|^2 = \sum_{i=1}^{m}(a_i - b_i)^2$$

$$\|a-b\|^2 = \sum_{i=1}^{m}\left(a_i^2 + b_i^2 - 2a_i b_i\right)$$

$$\|a-b\|^2 = \sum_{i=1}^{m} a_i^2 + \sum_{i=1}^{m} b_i^2 - 2\sum_{i=1}^{m} a_i b_i$$

$$\|a-b\|^2 = \|a\|^2 + \|b\|^2 - 2a \cdot b$$

- $\|a-b\|^2 = (\|a\|\sin\theta)^2 + (\|b\| - \|a\|\cos\theta)^2$



- $= \|a\|^2\sin^2\theta + \|b\|^2 + \|a\|^2\cos^2\theta - 2\|a\|\|b\|\cos\theta$
- $= \|a\|^2(\cos^2\theta + \sin^2\theta) + \|b\|^2 - 2\|a\|\|b\|\cos\theta$
- $= \|a\|^2 + \|b\|^2 - 2\|a\|\|b\|\cos\theta$
- $\|a\|^2 + \|b\|^2 - 2a\cdot b = \|a\|^2 + \|b\|^2 - 2\|a\|\|b\|\cos\theta$
- $2a\cdot b = 2\|a\|\|b\|\cos\theta$

$$\cos\theta = \frac{a\cdot b}{\|a\|\|b\|}$$

Geometric interpretation of covariance and correlation

- Remember we treated each observation of the dataset as an element in d-dimensional vector space where you have d attributes.
- However, you could do this for the attributes as well. Suppose you have n observations then attribute X could be thought of as a vector in n dimensional space.
- $D = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}, X = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$

- However, you could do this for the attributes as well. Suppose you have n observations then attribute X could be thought of as a vector in n dimensional space.
- $D = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}, X = (x_1, x_2, \dots, x_n)^T$
- Further, let
- $Z = X - 1 \cdot \hat{\mu} = \begin{pmatrix} x_1 - \hat{\mu} \\ x_2 - \hat{\mu} \\ x_3 - \hat{\mu} \\ \vdots \\ x_n - \hat{\mu} \end{pmatrix}$ This is the mean subtracted attribute vector.

Estimated variance

- $\hat{\sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \frac{1}{n-1}Z^T Z = \frac{\|Z\|^2}{n-1}$
- Estimated variance is the dot product of the centered attribute vector with itself divided by n-1

Estimated Covariance and correlation

- $\hat{\sigma}_{12} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{1i} - \hat{\mu}_1)(x_{2i} - \hat{\mu}_2)$
- $\hat{\rho}_{12} = \frac{\sum_{i=1}^{n} (x_{1i} - \hat{\mu}_1)(x_{2i} - \hat{\mu}_2)}{\sqrt{\sum_{i=1}^{n} (x_{1i} - \hat{\mu}_1)^2} \sqrt{\sum_{i=1}^{n} (x_{2i} - \hat{\mu}_2)^2}}$
- $\hat{\sigma}_{12} = \frac{1}{n-1} Z_1^T \cdot Z_2$ (the dot product between the two centered attribute vectors, divided by n-1)
- $\hat{\rho}_{12} = \frac{Z_1^T \cdot Z_2}{\|Z_1\| \|Z_2\|}$
- What is $\hat{\rho}_{12}$?
- It is the cosine of the angle between the centered attribute vectors of attributes 1,2.

Interesting post that I saw on twitter.

unusual_whales
@unusual_whales

Subscribe

Bessent: Markets are going down because Japan's bond market just suffered a six-standard-deviation move in ten-year bonds over the past two days... This has nothign to do with Greenland.

2:31 PM · Jan 20, 2026 · **1.3M** Views

- If the 10-year Japanese bond yield data follows a normal distribution, then observing a data point that is 6 sigma's away has roughly 1 in 507 million chance.
- However, in practice these financial data does not follow a normal distribution rather they have fat tailed distribution where extreme values are more frequent.
- But it is still rare.