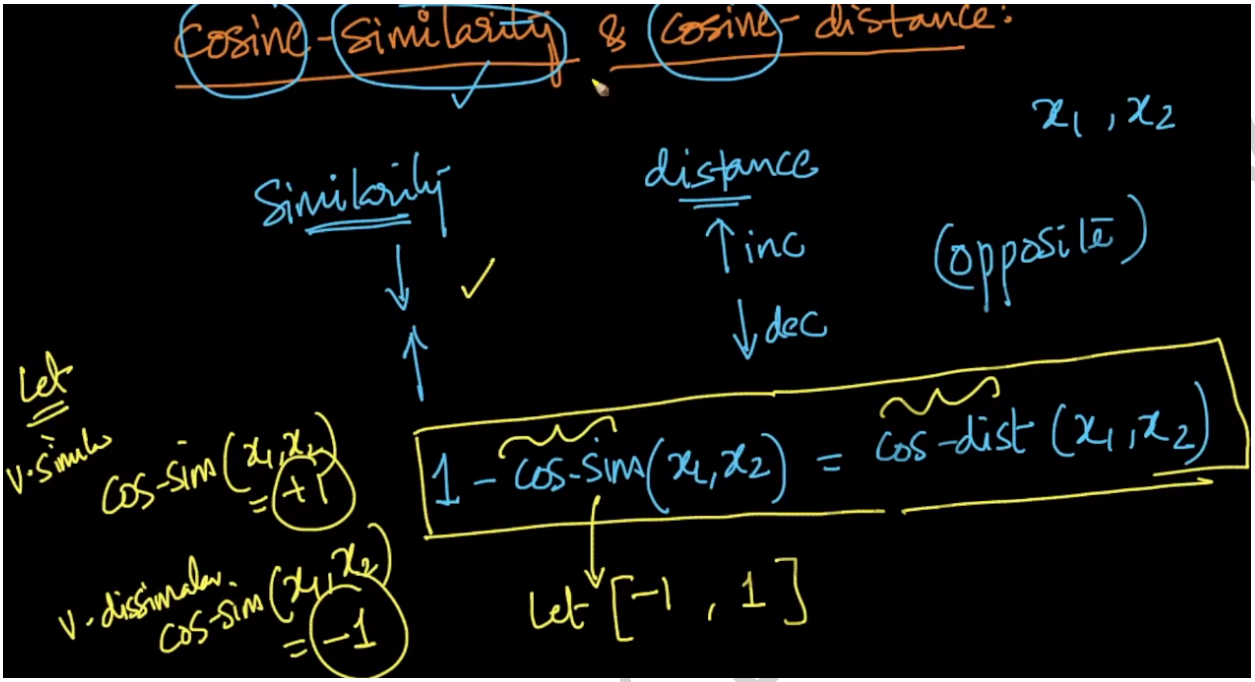
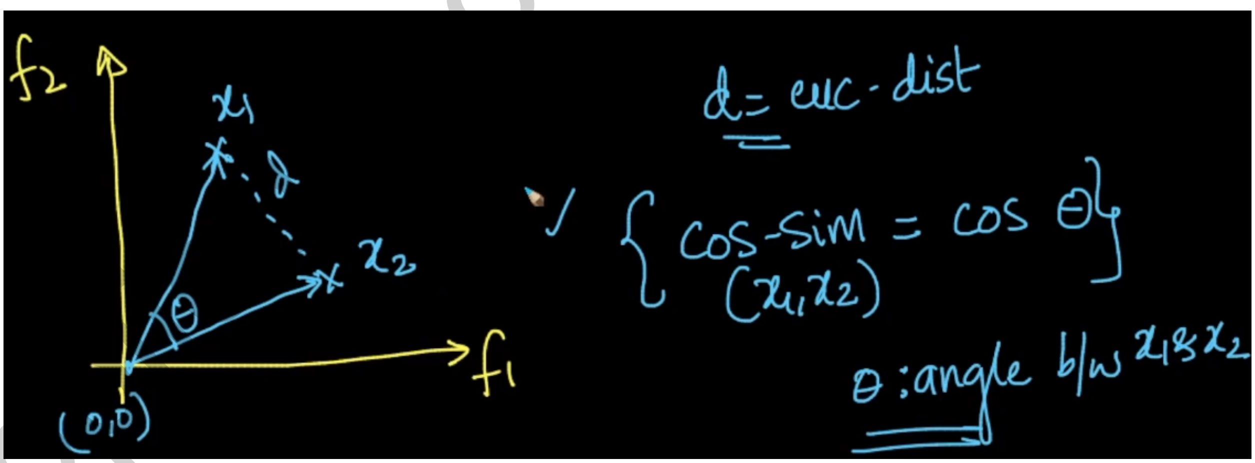
29.7 Cosine Distance & Cosine Similarity



Let us assume there are two vectors 'x_1' and 'x_2'. As the distance between the two points increases, the similarity between them decreases. Similarly if the similarity between two points increases, then the distance between them decreases.

The relationship between cosine similarity and cosine distance is given as

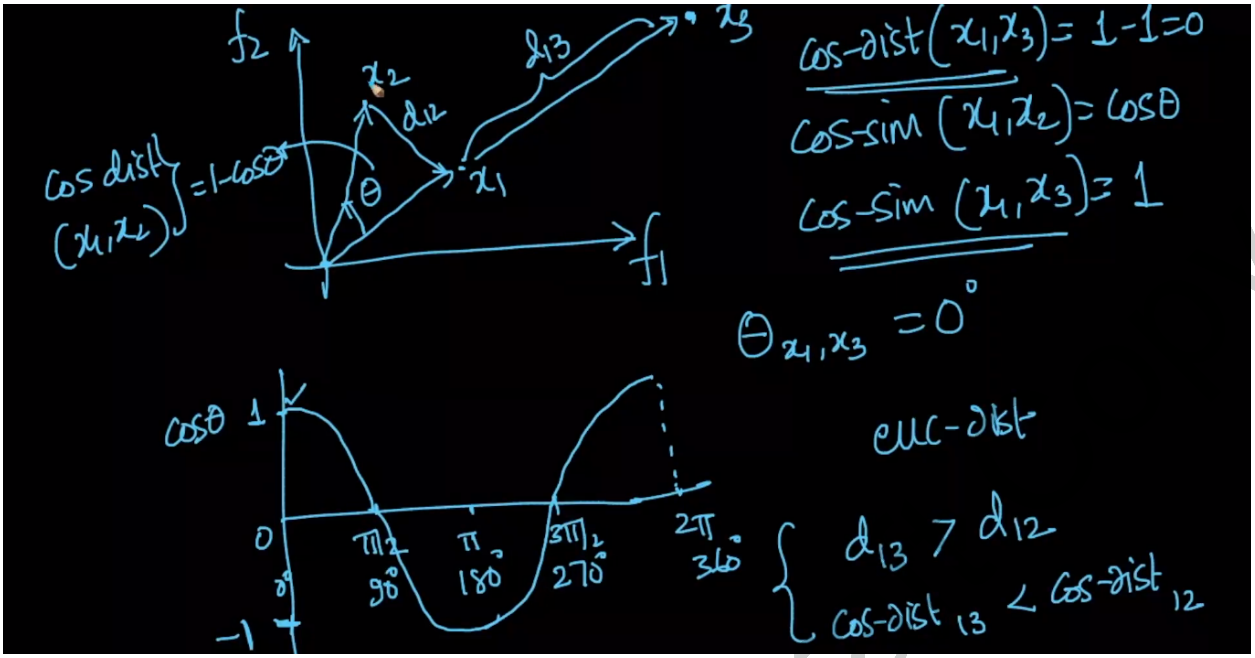
1 - cosine_similarity(x_1, x_2) = cosine_distance(x_1, x_2)



The cosine similarity between 'x_1' and 'x_2' is given by the cosine of the angle between the vectors 'x_1' and 'x_2'.

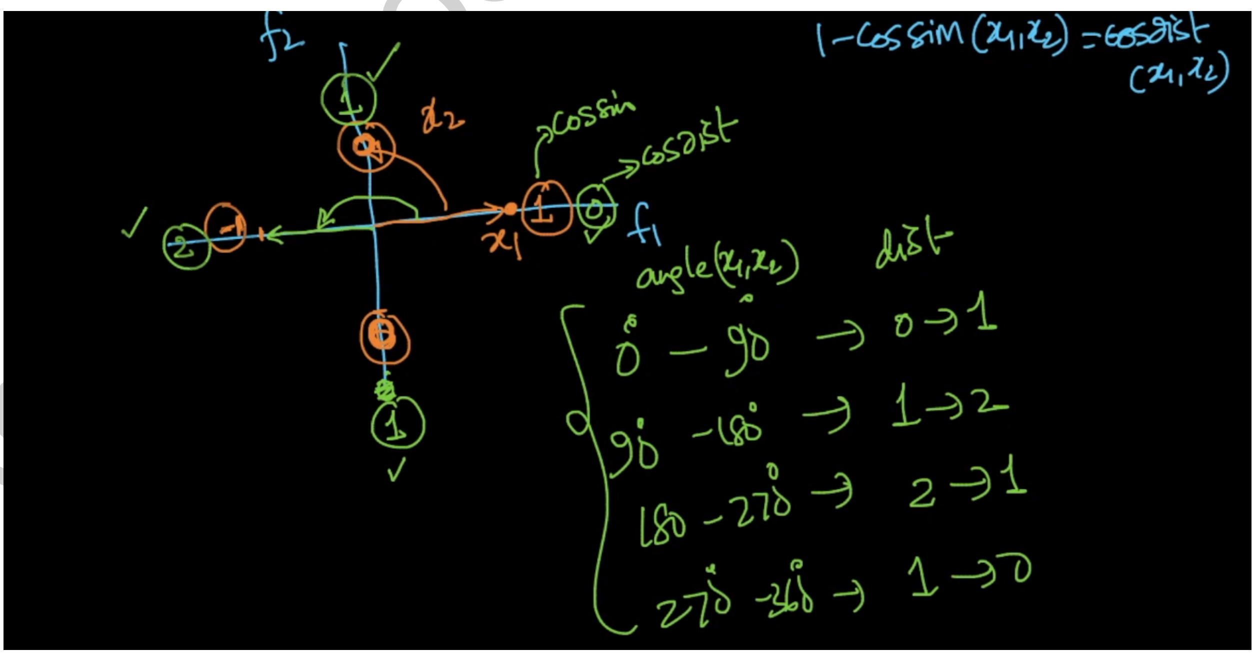
cosine_similarity(x_1, x_2) = cosθ where 'θ' is the angle between the vectors 'x_1' and 'x_2'.

Cosine Similarity is the measure of the similarity between two given non zero vectors.



Let us consider the 3 vectors ‘x_1’, ‘x_2’ and ‘x_3’. According to the figure above,
cosine_similarity(x_1, x_2) = cosθ
cosine_similarity(x_1, x_3) = cos(0) = 1
cosine_distance(x_1, x_3) = 1 - cosine_similarity(x_1, x_3) = 1 - 1 = 0
In case, if the angle between the vectors ‘x_1’ and ‘x_2’ is known, then
cosine_similarity(x_1, x_2) = cosθ
In case, if the angle between the vectors ‘x_1’ and ‘x_2’ is unknown, then
cosine_similarity(x_1, x_2) = $x_1.x_2/(||x_1||_2*||x_2||_2)$

If the vectors ‘x_1’ and ‘x_2’ are unit vectors, then **$||x_1||_2 = ||x_2||_2 = 1$**



If the angle between the two given vectors 'x_1' and 'x_2' lies in between 0^0 and 90^0 (or) 270^0 and 360^0(ie., $\theta \in [0, \pi/2]$ or $\theta \in [3\pi/2, 2\pi]$), then the cosine similarity lies in between 0 and 1. The cosine distance also lies in between 0 and 1.

If the angle between the two given vectors 'x_1' and 'x_2' lies in between 90^0 and 270^0((ie., $\theta \in [\pi/2, 3\pi/2]$)), then the cosine similarity lies in between -1 and 0. The cosine distance lies in between 1 and 2.

Note: Cosine Distance measures the angular difference between the vectors 'x_1' and 'x_2'. Cosine Similarity is used as a metric for measuring distance when the magnitude of the vectors does not matter.

Relationship between Euclidean Distance and Cosine Distance

If 'x_1' and 'x_2' are the two given vectors, then

$$||x_1-x_2||^2 = (x_1-x_2)^T(x_1-x_2) = ||x_1^2|| + ||x_2^2|| - 2x_1^Tx_2 \text{ ------- } (1)$$

$x_1^Tx_2$ is nothing but the dot product of the two vectors 'x_1' and 'x_2'.

Formula: *If we have two vectors 'A' and 'B', then*

$A^TB = A.B = ||A||*||B||*\cos(\theta)$ where 'θ' is the angle between the vectors 'A' and 'B'.

So $x_1^Tx_2 = x_1.x_2 = ||x_1||*||x_2||*\cos(\theta)$ -------- **(2)**

In the above equation, 'θ' is the angle between the vectors 'x_1' and 'x_2'.

After substituting (2) in (1), we get

$$||x_1-x_2||^2 = ||x_1^2|| + ||x_2^2|| - (2*||x_1||*||x_2||*\cos(\theta)) \text{ -------- } (3)$$

Let us assume the given two vectors 'x_1' and 'x_2' are the unit vectors, then

$$||x_1|| = ||x_2|| = 1 \text{ -------- } (4)$$

After substituting (4) in the RHS of (3), we get

$$||x_1-x_2||^2 = 1 + 1 - (2*\cos(\theta))$$

$$||x_1-x_2||^2 = 2 - (2*\cos(\theta)) = 2(1 - \cos(\theta)) \text{ ---- } (5)$$

Here $||x_1-x_2||^2 \rightarrow$ Euclidean Distance between the vectors 'x_1' and 'x_2'.

$\cos(\theta)$ $\rightarrow$ Cosine Similarity between the vectors 'x_1' and 'x_2'.

$1-\cos(\theta)$ $\rightarrow$ Cosine Distance between the vectors 'x_1' and 'x_2'.

So now the equation (5) becomes

[euclidean_distance(x_1, x_2)]^2 = 2*cosine_distance(x_1, x_2)

The above derived relationship is valid and applicable only if the two given vectors 'x_1' and 'x_2' are the unit vectors. (ie., $||x_1||_2 = ||x_2||_2 = 1$)