

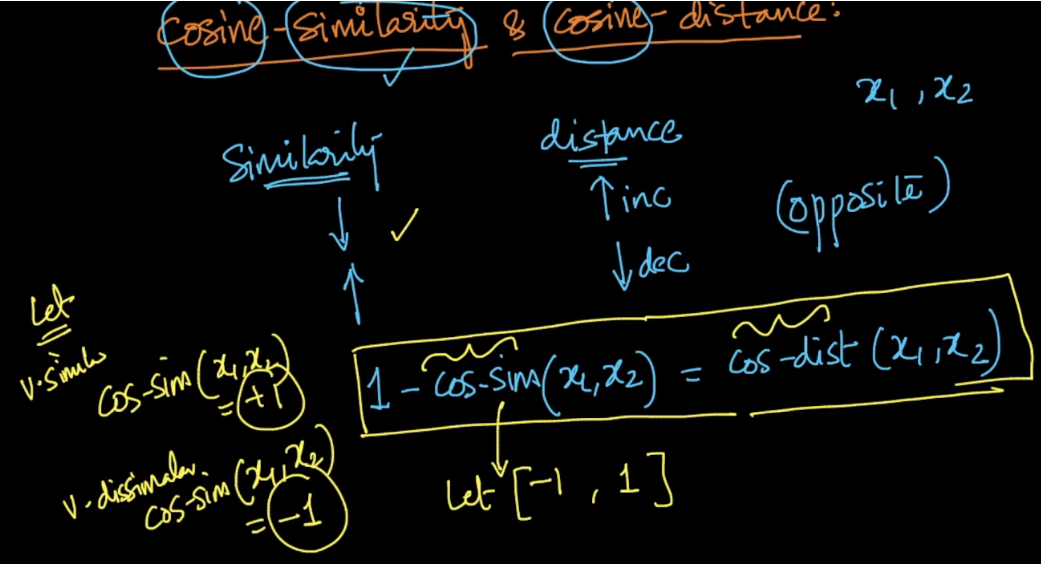
Above image shows the relation between cosine-similarity & cosine-distance:

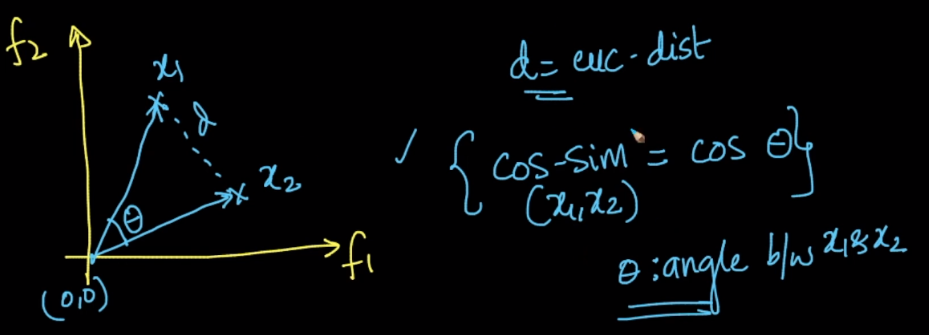
* Cosine distance increases when cosine similarity decreases
* Cosine distance decreases when cosine similarity increases

Therefore mathematical relation between them can be stated as:

1 – cos-similarity(x1, x2) = cos-dis(x1, x2).

Here cos-similarity is cos(theta) between vector x1 and x2.

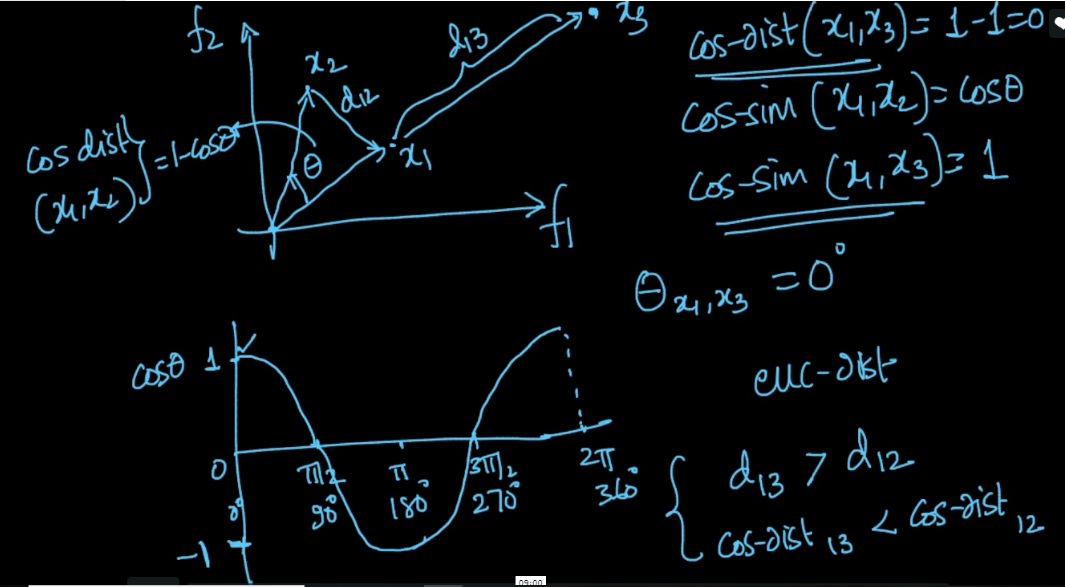


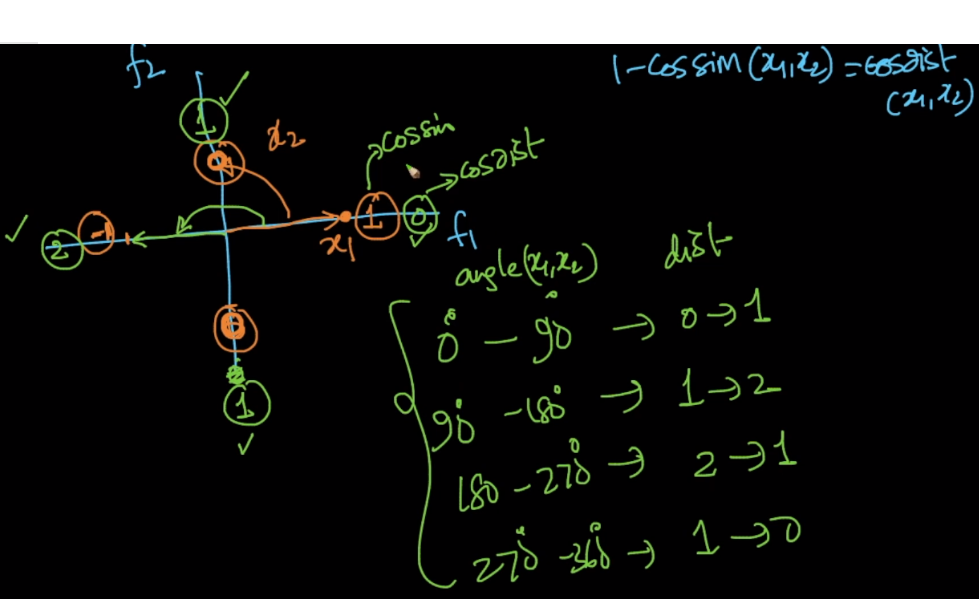


Above image shows that for Euclidean distance we consider **d** (gemotric dist between x1 and x2), and for cosine distance we consider **theta** (angle between vector x1 and x2).

Below image shows that x1 and x3 are in same direction having angle 0 between them,

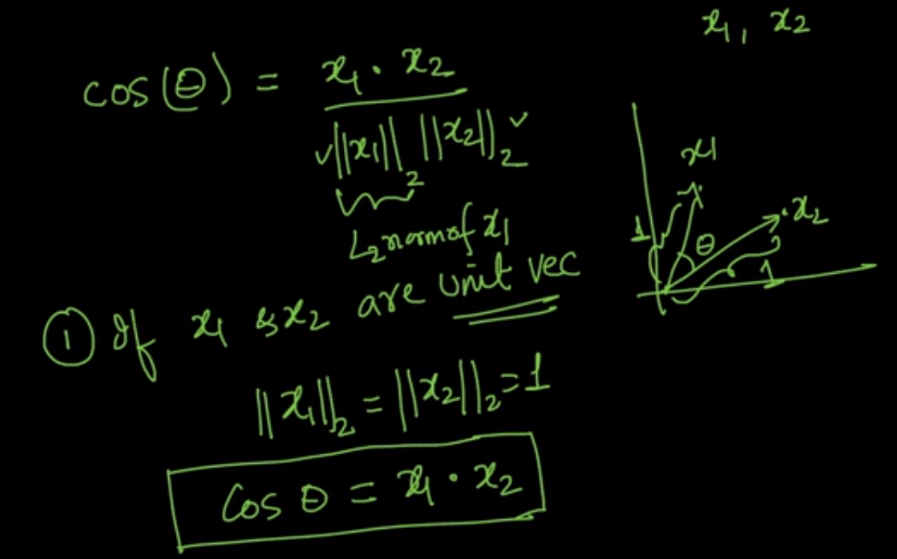
Therefore cos-dist(x1, x3) = 1 – cos(0) = 1 – 1 =0





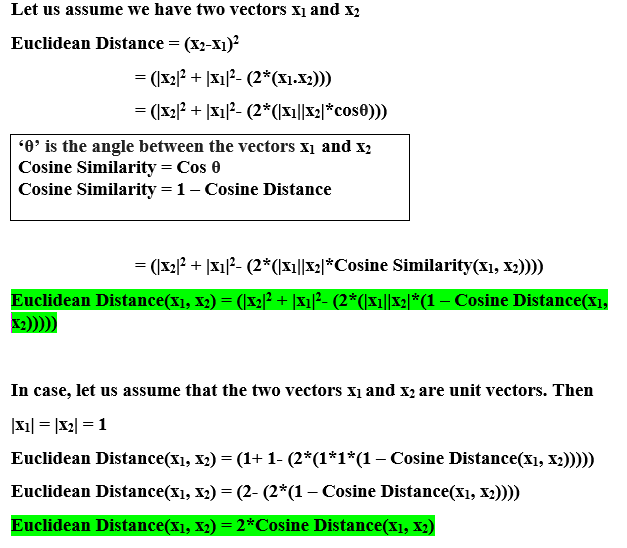
Above image shows cosine distance for different angle between two vectors.

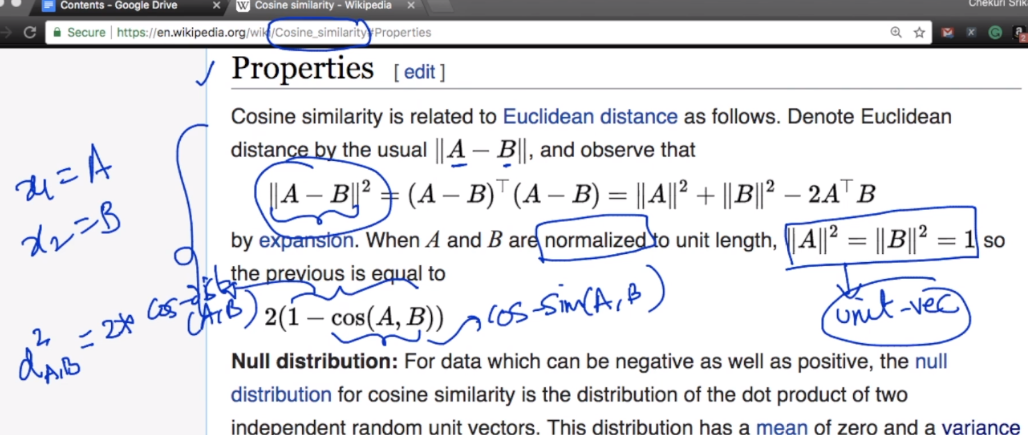
|  |  |  |
| --- | --- | --- |
| Angle between vectors | Cosine similarity: cos(theta) | Cosine distance:  1 – cos-similarity |
| 0 | 1 | * 1. 0 |
| 90 | 0 | 1. 1 |
| 180 | -1 | (1+1) 2 |
| 270 | 0 | 1. 1 |

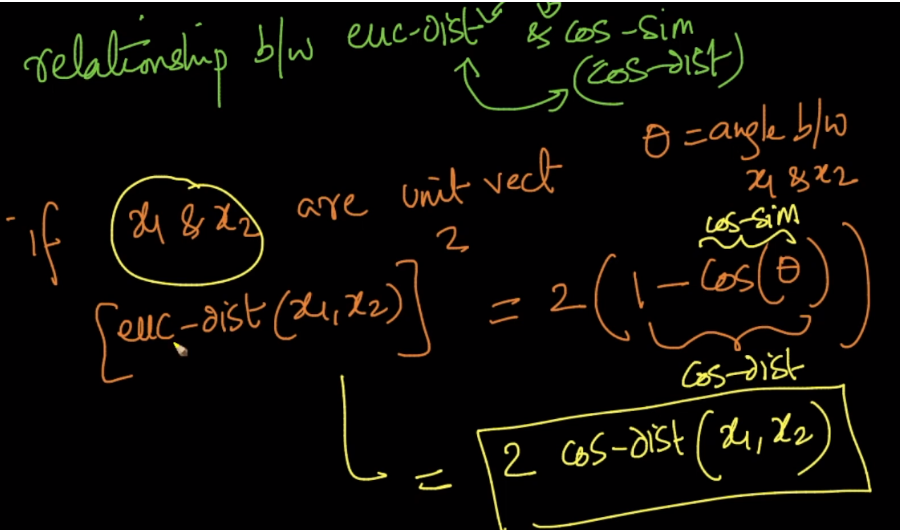


pto

**Below images shows relation between Euclidean distance and cosine distance.**







Hence we can say if x1 and x2 are unit vectors then

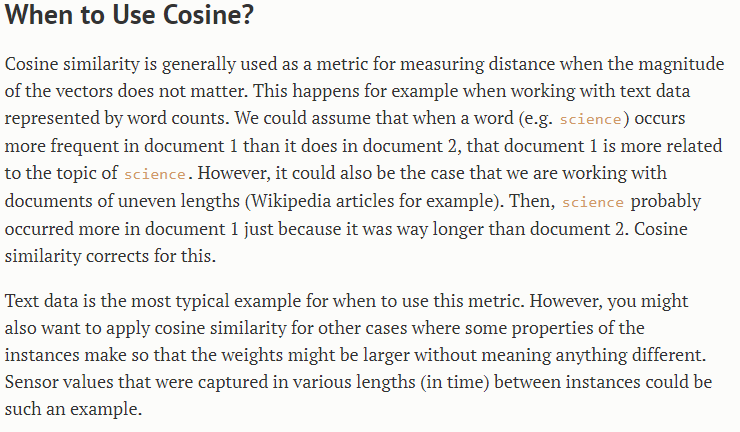
**Square of Euclidean distance = 2 \* cosine-distance.**

Notes:

* we use cosine similarity because cosine similarity is better at catching the semantic of each text, the direction the text points can be thought as its meaning so texts with similar meanings will be similar.  
  when modeling texts as vectors you will have many dimensions, thousands, the euclidean distance is not very good for very high dimensional data
* **Must Read:** Euclidean VS cosine distance: It explains where to use what with real example.

<https://cmry.github.io/notes/euclidean-v-cosine>

Conclusion for this blog is mentioned in below image.



* how can we apply cosine similarity concept if vector is not unit vector.If we are converting to unit vector,aren't we pulling all the data points within hyperUnitCube?And also loosing information?I have read all the comments but not got fully.Pls clarify in detail the above doubts.

1)the cosine similarity work better for unit vector as it gives the angle between two vectors. as the length is same of both the vectors then we can say that they differ by this much angle in similarity only.

2) no we are not loosing information here, as our task here is to check the cosine similarity , we only want to find the similarity by taking the angle between them and nothing else is required. So from the point of view of measuring similarity we just want to find the angle nothing else.