



Semester : VII

Subject : Big Data Analytics

Academic Year: 2024 – 2025

Module 5:

Hamming Distance

Hamming Distance

=> Hamming distance is used for the boolean vectors, that is which contain only 0 and 1

=> The no. of items in which the two items differ is the Hamming Distance between them

Example:

$$P_1 = 1 \ 0 \ 1 \ 0 \ 1$$

$$P_2 = 1 \ 1 \ 1 \ 1 \ 0$$

$$d(P_1, P_2) = 3$$

Cosine Distance

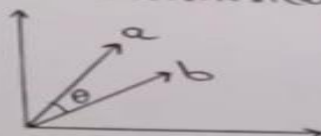
Cosine Distance

=> Cosine Distance is the angle between two vectors

=> Less the angle between two vectors more is the similarity and vice versa.

Note: The angle ranges from 0° to 180°

Application: Recommendation System





Example

Q Consider the following two vectors in the Euclidean Space
 $x = [1, 2, -1]$ and $y = [2, 1, 1]$
Calculate cosine distance between x & y
 $\Rightarrow \cos \theta = \frac{\text{Dot Product of vectors}}{\text{L2-norm of both vectors}}$
 $x \cdot y = (1 \times 2) + (2 \times 1) + ((-1) \times 1)$
 $= 2 + 2 - 1 = 3$
 $\text{L2-norm for } x = \sqrt{(1)^2 + (2)^2 + (-1)^2}$
 $= \sqrt{1 + 4 + 1} = \sqrt{6}$
 $\text{L2 Norm for } y = \sqrt{(2)^2 + (1)^2 + (1)^2} = \sqrt{4 + 1 + 1} = \sqrt{6}$

$$\cos \theta: 3/\sqrt{6}\sqrt{6}=1/2$$

$$\cos \theta = 60^\circ$$