PARSHWANATH CHARITABLE TRUST'S



A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering
Data Science



Semester: VII Subject: Big Data Analytics Academic Year: 2024 – 2025

Module 5:

Hamming Distance

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 Humming Distance between
them

Example:

P_1 = 10101

P_2 = 11110

d(p,, P_2) = 3

Cosine Distance

Distance

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 range from 0° to 180°

Application: Recommendation System

PARSHWANATH CHARITABLE TRUST'S

A.P. SHAH INSTITUTE OF TECHNOLOGY

Department of Computer Science and Engineering Data Science



Example

```
Q Consider the following two vectors in
     the Euclidean Space
     X=[1,2,-1] and J=[2,1,1]
    Calculak cosine distance between X dy
 => cose = Dot Product of vectors
                L2-norm of both vectors
   x. y=(1x2)+(2x1)+((-1)x1)
        = 2 + 2 - 1 = 3
L2 normforx = \( (1)^2 + (2)^2 + (-1)^2
L2 Norm for y = J(2)2+(1)2+(1)2 = J4+1+1
```

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

COSθ=60degree