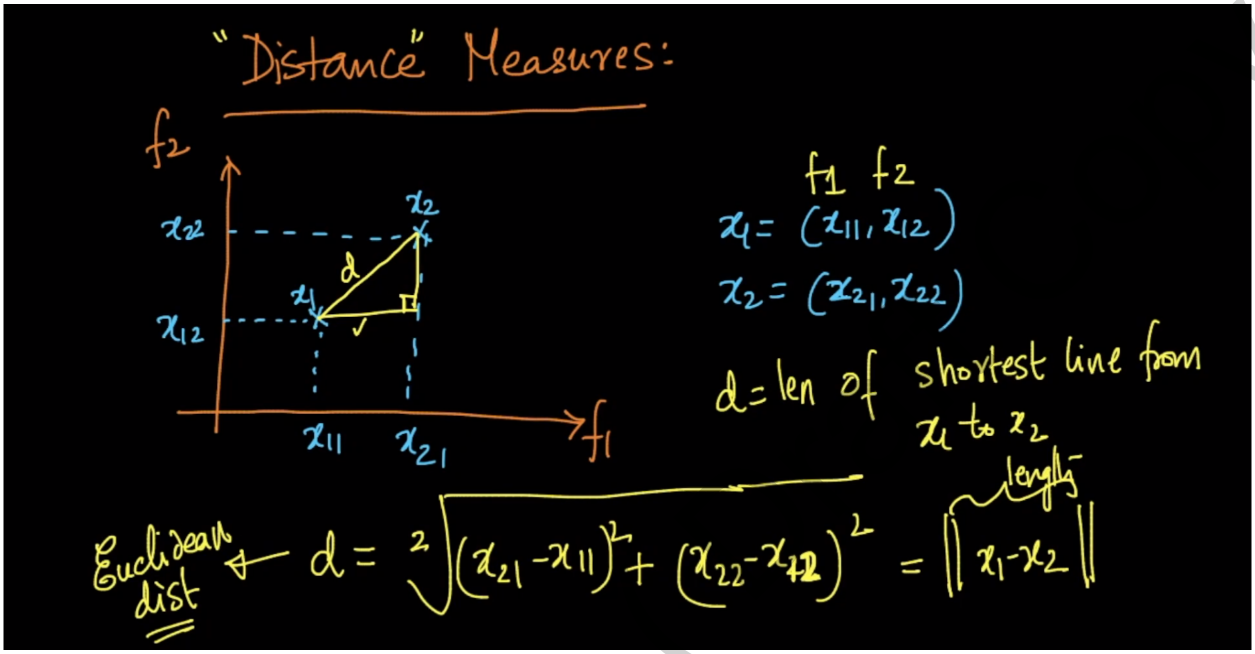
29.6 Distance Measures: Euclidean(L2), Manhattan(L1), Minkowski, Hamming

Euclidean Distance (L2)



Let us assume, we have a 2-dimensional space and we have two points ‘x_1’ and ‘x_2’.

$x_1 = (x_{11}, x_{12})$

$x_2 = (x_{21}, x_{22})$

d → Length of the shortest line from ‘x_1’ to ‘x_2’

$d = sqrt((x_{21}-x_{11})^2 + (x_{22}-x_{12})^2) = ||x1-x2||$

Here ‘d’ → euclidean distance between the points ‘x_1’ and ‘x_2’.

Let us assume we are working on a d-dimensional space.

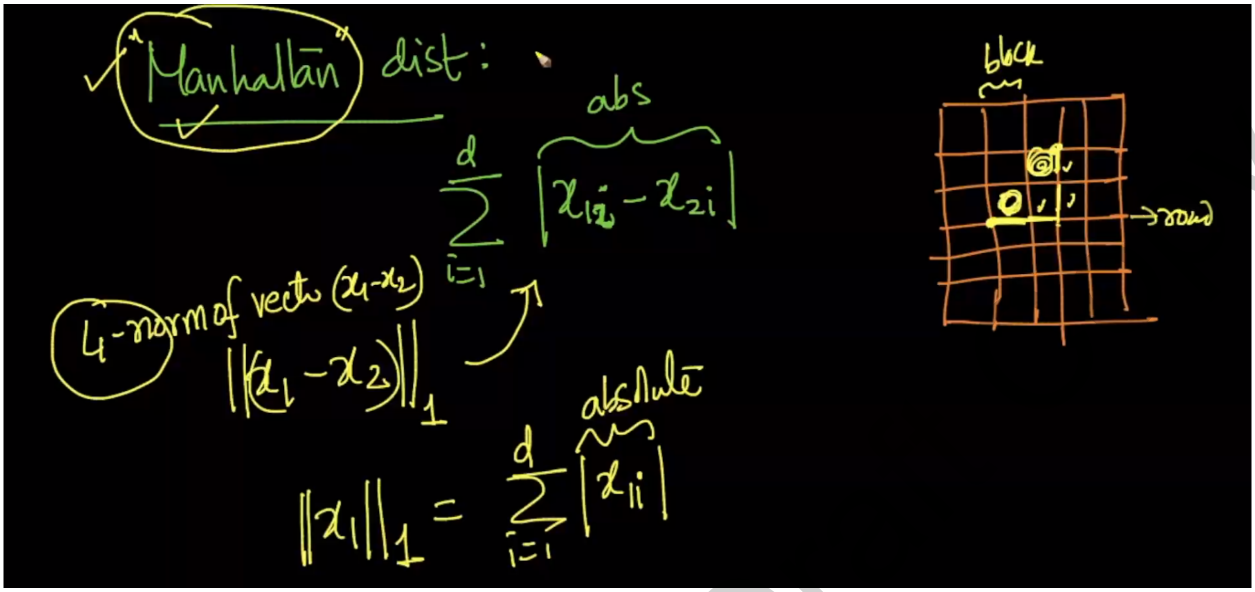
$x_1 \in R^d$, $x_2 \in R^d$

Euclidean Distance = $sqrt(\Sigma_{i=1}^{d} (x_{1i}-x_{2i})^2)$

$||x_1-x_2||_2$ → L2 norm of the vector (x_1-x_2)

$||x_i||$ → Distance of the point ‘x_i’ from origin = **$sqrt(\Sigma_{i=1}^{d} x_i^2)$**

Manhattan Distance (L1)



The Manhattan distance (d) between $x_1(x_{11}, x_{12})$ and $x_2(x_{21}, x_{22})$ is given as

Manhattan Distance = $\boldsymbol{\Sigma_{i=1}^{2} |x_{1i}-x_{2i}|}$

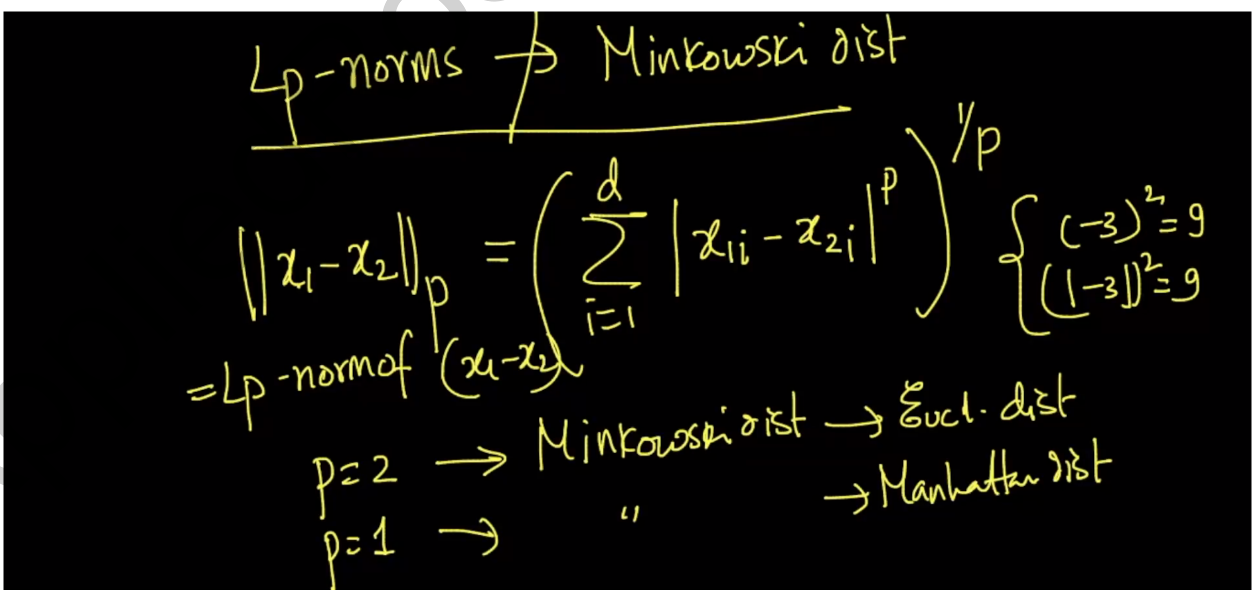
For a d-dimensional space, it is given as

Manhattan Distance = $\boldsymbol{\Sigma_{i=1}^{d} |x_{1i}-x_{2i}|}$

$||x_1-x_2|| \rightarrow$ L1 Norm of the vector (x_1-x_2)

$||x_1||_1 = \boldsymbol{\Sigma_{i=1}^{d} |x_{1i}|}$

Minkowski Distance (Lp)



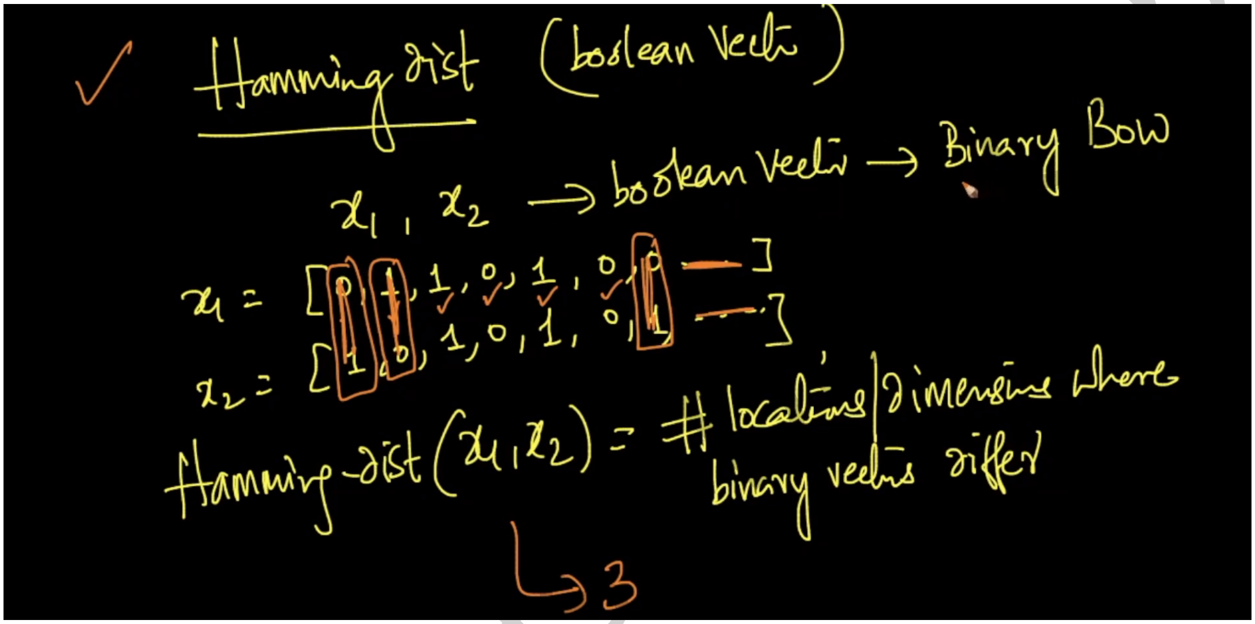
The Minkowski distance is the 'L$_p$' norm of the vector (x_1-x_2). The Minkowski distance in a d-dimensional space is given as

$\mathbf{||x_1-x_2||_p = (\Sigma_{i=1}^{d} |x_{1i}-x_{2i}|_p)^{1/p}}$ (Here p>0 always)

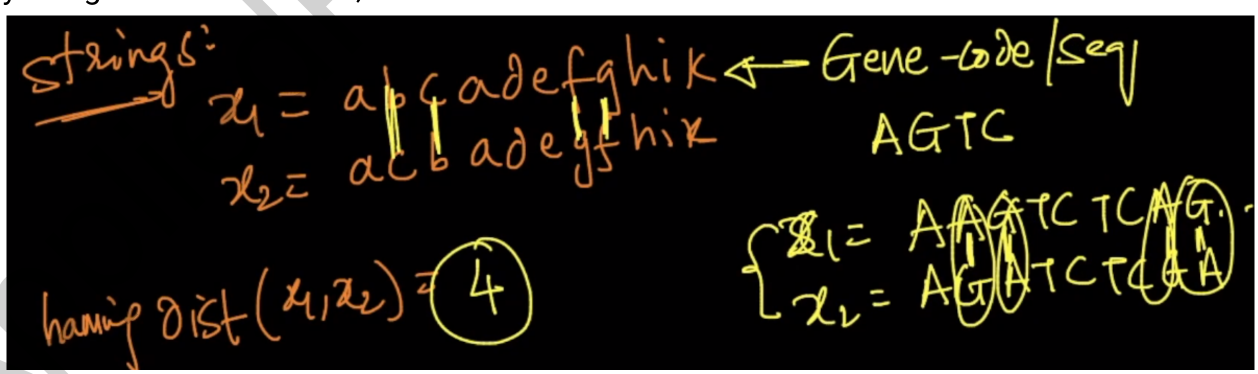
If p = 1, Minkowski Distance → Manhattan Distance

If p = 2, Minkowski Distance → Euclidean Distance

Hamming Distance



Hamming Distance between the two given vectors is the number of dimensions/features in which the two vectors differ. It is recommended to use Hamming Distance only on Binary Vectors. One best example is to apply Hamming Distance on the Binary BOW data. You also can apply it on Count Bsed BOW, but still you'll get the same result, as that obtained on the BOW data.



Hamming Distance between the two given strings is the total number of differing components in those strings. In the above example, we see there are 4 characters differing(2nd, 3rd, 7th, 8th locations). So the hamming distance is 4.

Note: Hamming Distance is no way related to the Minkowski Distance.