Proximity measures:

In order to classify patterns, they need to be compared against each other and against a standard when a new pattern is present and it is necessary to classify it.

* ### Distance measure:

A distance measure is used to find the dissimilarity between pattern representations. Patterns which are more similar should be closer. The distance function could be a metric (or) a non-metric.

A metric is a measure for which the following properties hold:

i) positive reflexivity $d(x,x) = 0$

1. $d(x,y) \geq 0$ - non-negativity
2. $d(x,y) = 0 \Leftrightarrow x = y$ identity.

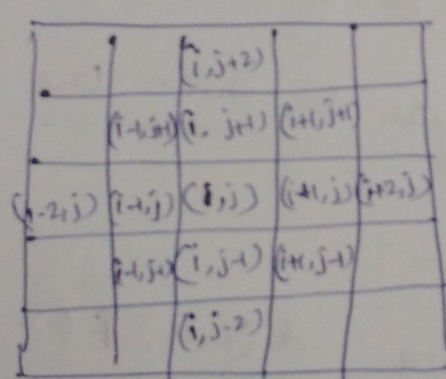
ii) Symmetry $d(x,y) = d(y,x)$

iii) Triangular inequality $d(x,y) \leq d(x,z) + d(z,y)$

The popularly used distance metric called 'Minkowski metric' is of the form

$$L_p(X, Y) = \left[\sum_{i=1}^{d}(|x_i - y_i|)^p\right]^{1/p}$$ - which is also called as Lp Norm.

When p is 1; It is called the Manhattan distance / L_1 distance | Taxicab distance / City-blo distance



$X = \langle x_1, x_2, \ldots x_d \rangle$;

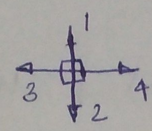
$Y = \langle y_1, y_2, \ldots y_d \rangle$;

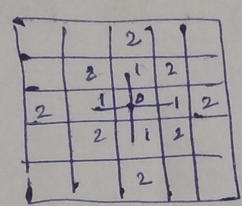
when $d = 2$ $X = \langle x_1, x_2 \rangle$

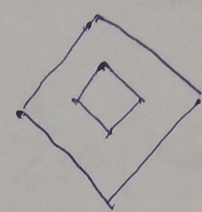
$Y = \langle y_1, y_2 \rangle$

$L_1(X,Y) = |x_1 - y_1| + |x_2 - y_2|$

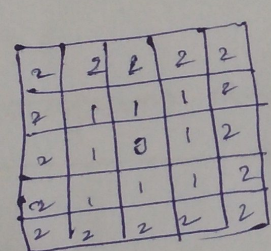
Based on 4-connectivity ; the L_1 norm is called as city block distance

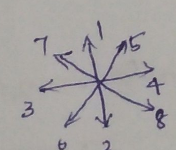


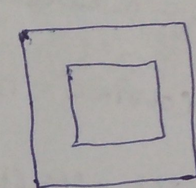




Based on 8-connectivity ; the L_1 norm is called as Chess board distance max distance







Euclidean distance

when $P = 2$; It is called Euclidean distance

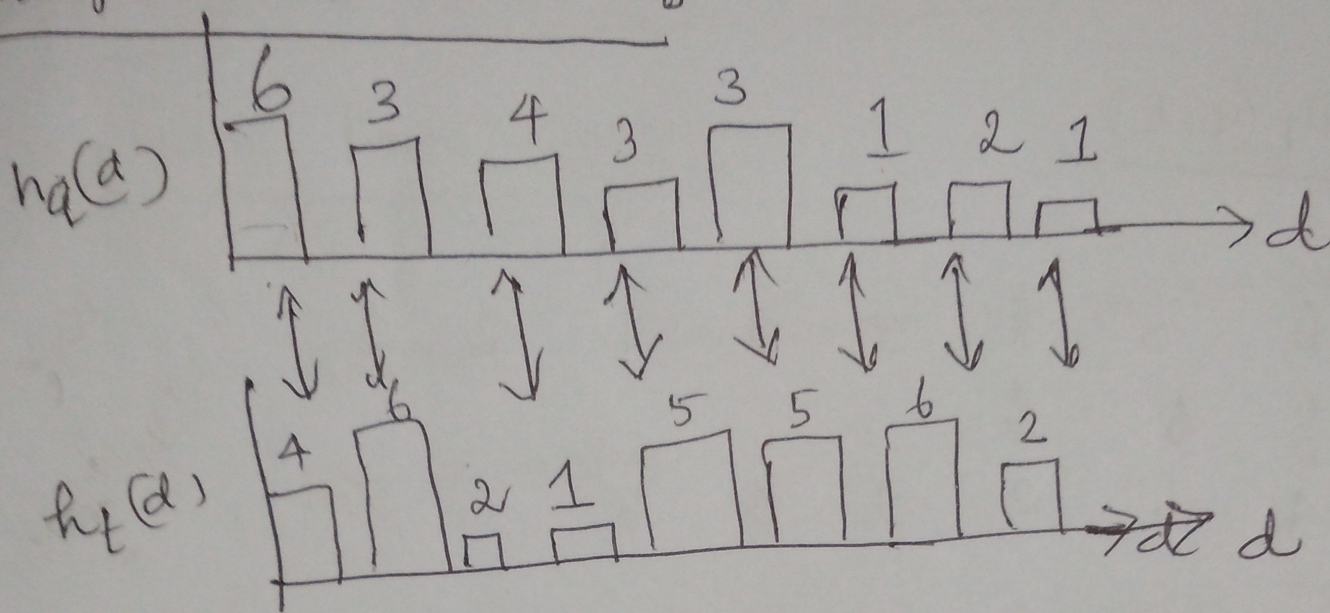
$$L_2(X,Y) = \left[|x_1 - y_1|^2 + |x_2 - y_2|^2 \right]^{1/2}$$

$$= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

L_∞ : max (Chess board distance)

$L_{-\infty}$ = min

istogram L_1 distance:



$$D_1(q,t) = 1 - \frac{\sum_{i=1}^{d} \min\left[h_q(i), h_t(i)\right]}{\min\left[|h_q|, |h_t|\right]}$$

If histograms are normalized ~~[|hq| = |ht|]~~

$$D_1(q,t) = \sum_{i=1}^{d} \left[h_q(i) - h_t(i) \right]$$

According to $D_1(q,t)$

$$q = 6 \quad 3 \quad 4 \quad 3 \quad 3 \quad 1 \quad 2 \quad 1$$

$$t = 4 \quad 6 \quad 2 \quad 1 \quad 5 \quad 5 \quad 6 \quad 2$$

$\min(6,4)$.

$$\frac{4+3+2+1+3+1+2+1}{\min(111,111)} = \frac{17}{1} = 17$$

This captures similarity between two histograms q and t $D_1(q,t) = 17$

According to $D_2(q,t)$:

$$D_2(q,t) = (6-4)^2 + (3-6)^2 + (4-2)^2 + (3-1)^2 + (3-5)^2 + (1-5)^2 + (2-6)^2 + (1-2)^2$$

$$= \sqrt{4+9+4+4+4+16+16+1}$$

$$= \sqrt{62}$$

$$= 7.87$$