16/1/2019 Bhattacharya Distance :-

It measures the similarity between two probability distributions.

[It is used to compare two normalized histograms]

Let the two normalized histograms be:

$$x = (x_1, x_2, \ldots x_n)$$

$$y = (y_1, y_2, \ldots y_n)$$

Consider two new vectors x', y' s.t.

$$\left.\begin{aligned} x' &= (\sqrt{x_1}, \sqrt{x_2}, \ldots \sqrt{x_n}) \\ y' &= (\sqrt{y_1}, \sqrt{y_2}, \ldots \sqrt{y_n}) \end{aligned}\right\} \text{———— Eqn } 1$$

Now, find the dot product of x' & y'

$$x' \cdot y' = |x'||y'| \cos\theta \text{ ———— Eqn } 2$$

substituting the values ~~Eqn~~ from Eqn 1 to Eqn 2;

We have;

$$\sqrt{x_1 y_1} + \sqrt{x_2 y_2} + \ldots \sqrt{x_n y_n} = \left(\sqrt{x_1 + x_2 + x_3 + \ldots x_n}\right)\left(\sqrt{y_1 + y_2 + \ldots y_n}\right) \cos$$

———— 3

Since; $x_1 + x_2 + \dots x_n = 1$

$y_1 + y_2 + \dots y_n = 1$;

[As x & y denotes two probability distributions;

$$\Sigma x_i = \Sigma y_i = 1$$

where; $x = (x_1, x_2 \dots x_n)$ &

$y = (y_1, y_2, \dots y_n)$].

From the above substitution;

Eqn 3 becomes;

$$\sqrt{x_1 y_1} + \sqrt{x_2 y_2} + \dots \sqrt{x_n y_n} = 1.\cos\theta$$

$$\therefore \cos\theta = \sum_{i=1}^{n} \sqrt{x_i y_i}$$

Bhattacharya Coefficient;

$$\boxed{B(x,y) = \sum_{i=1}^{n} \sqrt{x_i y_i}}$$

Bhattacharya Distance:

$$\boxed{d_B(x,y) = -\ln B(x,y)}$$

Range;

$$\boxed{\begin{array}{l} 0 \le B(x,y) \le 1 \\ 0 \le d_B < \infty \end{array}}$$

Hellinger Distance

$$\boxed{d_h(x,y) = 1 - B(x,y)}$$

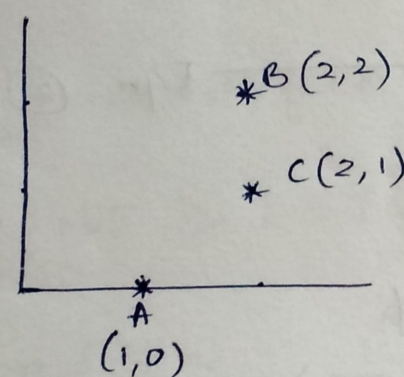
Bhattachaya Distance

Is Bhattachaya Distance a metric ? No.

Proof:-

To prove that Bhattachaya Distance is not a metric, it is sufficient to prove the triangular inequality.

Consider the following example:



From the formula for Bhattachaya Distance,

$$\boxed{d_B(x,y) = -\ln B(x,y)}$$

where, x & y are two vectors;

d_B : Bhattachaya Distance &

$$B(x,y) = \cos\theta$$
$$= \sum_{i=1}^{n} \sqrt{x_i y_i}$$

$$d_B(A,B) = -\ln \sum \sqrt{x_i y_i}$$

$$\cancel{-\ln(\quad)} = -\ln(\sqrt{2.1} + \sqrt{2.0})$$

$$= \cancel{-\ln(\sqrt{2.1 + 2.0})} = -\ln(\sqrt{2})$$

$$= -\ln(\sqrt{2})$$

$$= -0.3465 \quad \text{——— Eqn } 1$$

$$d_B(A,C) = -\ln \sum \sqrt{x_i y_i}$$

$$= -\ln \sqrt{2}$$

$$= -0.3465 \quad \text{——— Eqn } 2$$

$$d_B(B,C) = -\ln \sum \sqrt{x_i y_i}$$

$$= -\ln(\sqrt{2.2} + \sqrt{2.1})$$

$$= -\ln(\sqrt{4} + \sqrt{2})$$

$$= -\ln(2 + \sqrt{2})$$

$$= -1.2279 \quad \text{——— Eqn } 3$$

As per triangular inequality:-

$$d(A,B) \leq d(A,C) + d(B,C) \quad \text{——— eqn.}$$

We need to prove that the above inequality is not satisfied for values of AB, BC & AC.

From 1, 2 & 3;

$$-0.3465 \leq -0.3465 + -1.2279$$

$$-0.3465 \leq -1.5744,$$ ~~is not valid~~/

is incorrect/false.

[ie, $0.3465 \geq 1.5744$ is false].

Hence, proved that Bhattacharya distance is not a metric *

16/1/2019

iv) Mutual neighbourhood distance (MND)

The similarity between two patterns A and B is

$S(A,B) = f(A,B,\epsilon)$

↳ ~~ε~~ is called the context and

corresponds to the surrounding points.

Steps:

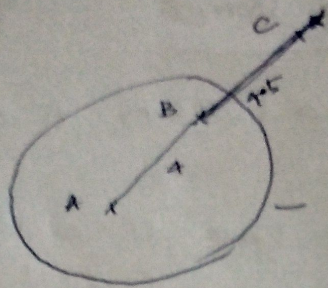
i) with respect to each point, all other data points are numbered from 1 to N-1 in increasing order of some distance measure, S.T the nearest neighbour is assigned value 1 and the farthest point is assigned the value N-1.

ii) If we denote by NN(u,v), the no. of data points v with respect to u, the (MND) is defined as,

$MND(u,v) = NN(u,v) + NN(v,u)$

*. This is symmetric, but not satisfying Triangular Inequality hence it is not metric.

Example:

i) 

– least MND distance is A and B.

Fig: 1

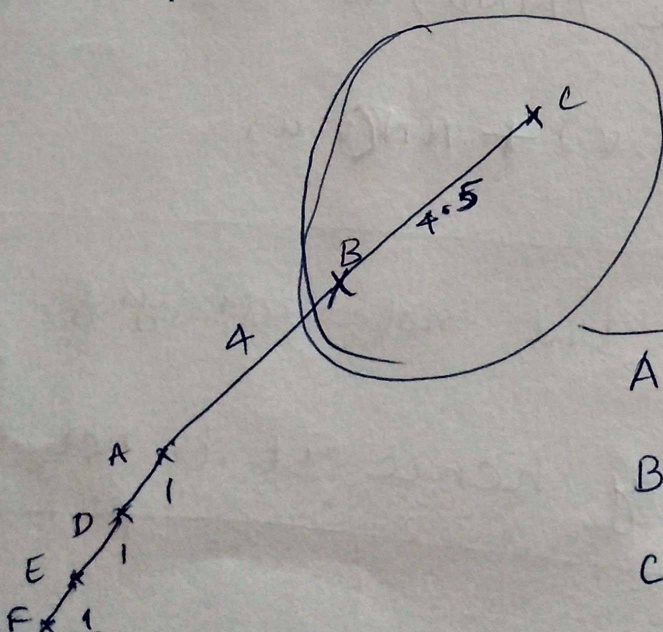
Consider Fig. 1, The ranking of the point A, B and C can be represented as

	1	2
A	B	C
B	A	C
C	B	A

$MND(A,B) = NN(A,B) + NN(B,A)$
$= 1 + 1$
$= 2$

$MND(B,C) = 3$
$MND(A,C) = 4$

$$\boxed{\begin{array}{c} MND(A,C) \leq MND(A,B) + MND(B,C) \\ 4 \not\leq 2 + 3 \end{array}}$$



– least MND distance is B and C.

Fig. 2

	1	2	3	4	5
A	D	E	F	B	C
B	A	C	D	E	F
C	B	A	D	E	F
D					
E					
F					

(5) least MND distance is B and C. This happens by changing the content.

$MND(A, B) = NN(A, B) + NN(B, A)$

$= 4 + 1 = 5$

$MND(B, C) = 3$ → least

$MND(A, C) = 7$

It can be seen that in case 1, the least MND distance is between A and B, whereas in case 2, it is between B and C. This happens by changing the context.