HW_Midterm_exam Q1

##

Course : Knowledge Discovery and Data Mining (CS 513-A)

First Name : Mahvish

Last Name : Syed

Id : 10456845

#purpose : Explain distance formula and calculate distance.

##

#1 (10 Points)

Is the following function a proper distance function? Why? Explain your answer. Measure the distance between (0, 0, 0), (0, 1, 0), (0, 1, 1), and (1, 1, 1)

$$d(x, y) = \Sigma\ (|x_{i-}\ _{Y_i}|^{\ 2})$$

ANSWER :

Yes It is a proper distance function. This can be proven with given below explanation.

Data analysts define distance metrics to measure similarity. A distance metric
or distance function is a real-valued function d, such that for any coordinates x, y,
and z:

1. d(x, y) ≥ 0, and d(x, y) = 0 if and only if x = y
2. d(x, y) = d(y, x)
3. d(x, z) ≤ d(x, y) + d(y, z)

Property 1: assures us that distance is always nonnegative, and the only way for distance to be zero is for the coordinates (e.g., in the scatter plot) to be the same.

If both points are same then only given distance formula compute result as zero.
For eg : say point A = (0,0) and point B = (0,0)
To calculate distance between A and B using given distance formula:
D(A, B) = (x1 - x2) ^2 + (y1 – y2) ^ 2
= (0 - 0) ^2 + (0 - 0) ^ 2
= 0 + 0 = 0
Hence given distance formula $d(x, y) = \Sigma\ (|x_{i-}\ Y_i|\ ^2)$ satisfy property 1.

Property 2: indicates commutativity. In our example distance between x to y or y to x yields same result because at end their difference will be squared.

While calculating distance using given formula, distance can never be negative because we compute square of differences.

For eg : say point A = (0,0) and point B = (1,1)
To calculate distance between A and B using given distance formula:
D(A, B) = (x1 - x2) ^2 + (y1 – y2) ^ 2
= (0 - 1) ^2 + (0 - 1) ^ 2
= 1 + 1 = 2
Hence distance between two point is positive

Similarly while calculating distance between B and A using given distance formula , result be will same as above :
D(B, A) = (x1 - x2) ^2 + (y1 – y2) ^ 2
= (1 - 0) ^2 + (1 - 0) ^ 2
= 1 + 1 = 2
Hence distance between two points are same.

Hence given distance formula $d(x, y) = \Sigma\ (|x_{i-}\ Y_i|\ ^2)$ satisfy property 2.

Property 3 is the triangle inequality, which states that introducing a third point can never shorten the distance between two other points.
In our case, if new point introduce between A and B say C with points A = (0,0) B = (1,1) and C = (0.5,0.5)
Then distance between A and C
= (0-0.5)^2 + (0 – 0.5) ^ 2 = 0.5
Distance between C and B
= (1 – 0.5) ^ 2 + (1 – 0.5) ^ 2 = 0.5

Therefore distance between A and B = 0.5 + 0.5 = 1

From property 2 we get distance between A and B is 1

Hence given distance formula $d(x, y) = \Sigma\ (|x_{i-}\ _{Y_i}|\ ^2)$ satisfy property 3.

Therefore given formula is valid distance formula as it satisfy distance property 1 , 2 and 3.

Let say point A = (0, 0, 0), B= (0, 1, 0), C= (0, 1, 1), and D= (1, 1, 1).
So calculating distance between them using Euclidean distance formula :
dEuclidean(x, y) = sqrt (Σ (xi − yi) ^ 2)

a) Distance between A(0,0,0) and B(0,1,0)
= Sqrt((x1- x2) ^ 2 + (y1 – y2) ^ 2 + (z1 – z2) ^ 2)
= sqrt((0-0)^2 + (0-1)^2 + (0-0)^2)
= sqrt(0 + 1 + 0)
= 1.

b) Distance between B(0,1,0) and A(0,0,0)
According to distance property 2 , D(A, B) = D(B, A)
Therefore D(B, A) = 1

c) Distance between A(0,0,0) and C(0,1,1)

= Sqrt((x1- x2) ^ 2 + (y1 – y2) ^ 2 + (z1 – z2) ^ 2)

= sqrt((0-0)^2 + (0-1)^2 + (0-1)^2)

= sqrt(0 + 1 + 1)

= 1.41.

d) Distance between C(0,1,1)and A(0,0,0)

According to distance property 2 , D(A, C) = D(C, A)

Therefore D(C, A) = 1.41

e) Distance between A(0,0,0) and D(1,1,1)

= Sqrt((x1- x2) ^ 2 + (y1 – y2) ^ 2 + (z1 – z2) ^ 2)

= sqrt((0-1)^2 + (0-1)^2 + (0-1)^2)

= sqrt(1 + 1 + 1)

= 1.73.

f) Distance between D(1,1,1)and A(0,0,0)

According to distance property 2 , D(A, D) = D(D, A)

Therefore D(D, A) = 1.73

g) Distance between B(0,1,0) and C(0,1,1)

= Sqrt((x1- x2) ^ 2 + (y1 – y2) ^ 2 + (z1 – z2) ^ 2)

= sqrt((0-0)^2 + (1-1)^2 + (0-1)^2)

= sqrt(0 + 0 + 1)

= 1.

h) Distance between C(0,1,1)and B(0,1,0)

According to distance property 2 , D(C, B) = D(B, C)

Therefore D(C, B) = 1

i) Distance between B(0,1,0) and D(1,1,1)

= Sqrt((x1- x2) ^ 2 + (y1 – y2) ^ 2 + (z1 – z2) ^ 2)

= sqrt((0-1)^2 + (1-1)^2 + (0-1)^2)

= sqrt(1 + 0 + 1)

= 1.41.

j) Distance between D(1,1,1)and B(0,1,0)

According to distance property 2 , D(D, B) = D(B, D)

Therefore D(D, B) = 1.41

k) Distance between C(0,1,1) and D(1,1,1)

= Sqrt((x1- x2) ^ 2 + (y1 – y2) ^ 2 + (z1 – z2) ^ 2)

= sqrt((0-1)^2 + (1-1)^2 + (1-1)^2)

= sqrt(1 + 0 + 0)

= 1.

l) Distance between D(1,1,1)and C(0,1,1)

According to distance property 2 , D(D, C) = D(C, D)

Therefore D(D, C) = 1.