

Distance Measures

Question 1:

Consider the following three vectors u, v, w in a 6-dimensional space:

$$u = [1, 0.25, 0, 0, 0.5, 0]$$

$$v = [0.75, 0, 0, 0.2, 0.4, 0]$$

$$w = [0, 0.1, 0.75, 0, 0, 1]$$

Suppose $\cos(x,y)$ denotes the similarity of vectors x and y under the cosine similarity measure. Compute all three pairwise similarities among u, v, w .

1st ans

Given data is

$$u = [0.25, 0.0, 0.5]$$

$$v = [0.75, 0, 0, 0.2, 0.4, 0]$$

$$w = [0, 0.1, 0.75, 0, 0, 1]$$

$$|u| = \sqrt{1^2 + 0.25^2 + 0^2 + 0^2 + 0.5^2 + 0^2} = 1.145$$

$$|v| = \sqrt{0.75^2 + 0^2 + 0^2 + 0.2^2 + 0.4^2 + 0^2} = 0.873$$

$$|w| = \sqrt{0^2 + 0.1^2 + 0.75^2 + 0^2 + 0^2 + 1^2} = 1.25$$

$$\cos(u, v) = \frac{u \cdot v}{|u| \cdot |v|} = \frac{0.75 + 0.02}{1.145 \cdot 0.873} \rightarrow \theta = 18 \text{ degrees}$$

$$\cos(v, w) = \frac{v \cdot w}{|v| \cdot |w|} = \frac{0}{0.873 \cdot 1.25} \rightarrow \theta = 0 \text{ degrees}$$

$$\cos(u, w) = \frac{u \cdot w}{|u| \cdot |w|} = \frac{0.025}{1.145 \cdot 1.25} \rightarrow \theta = 89 \text{ degrees}$$

Question 2:

Here are five vectors in a 10-dimensional space:

1111000000 0100100101 0000011110 0111111111 1011111111

Compute the Jaccard distance (not Jaccard "measure") between each pair of the vectors.

2 ans

$$i = 1111000000$$

$$j = 0100100101$$

$$k = 0000011110$$

$$L = 0111111111$$

$$M = 1011111111$$

$$\text{Jaccard distance } (i, j) = 1 - (1/7) = 6/7$$

$$\text{Jaccard distance } (i, k) = 1 - (0/8) = 1$$

$$\text{Jaccard distance } (i, L) = 1 - (3/10) = 7/10$$

$$\text{Jaccard distance } (i, M) = 1 - (3/10) = 7/10$$

$$\text{Jaccard distance } (j, k) = 1 - (1/7) = 6/7$$

$$\text{Jaccard distance } (j, L) = 1 - (4/9) = 5/9$$

$$\text{Jaccard distance } (j, M) = 1 - (3/10) = 7/10$$

$$\text{Jaccard distance } (k, L) = 1 - (4/9) = 5/9$$

$$\text{Jaccard distance } (k, M) = 1 - (4/9) = 5/9$$

$$\text{Jaccard distance } (L, M) = 1 - (8/10) = 2/10$$

Question 3:

Here are five vectors in a 10-dimensional space:

1111000000 0100100101 0000011110 0111111111 1011111111

Compute the Manhattan distance (L_1 norm) between each two of these vectors.

3rd ans let $i = 1111000000$ $j = 0100100101$ $k = 0000011110$

$l = 01111111$ $m = 1011111111$

Manhattan distance of $i, j = 6$

Manhattan distance of $i, k = 8$

" " of $i, l = 7$

" " of $i, m = 7$

" " of $j, k = 6$

" " of $j, l = 5$

" " of $j, m = 7$

" " of $k, l = 5$

" " of $l, m = 2$

Question 4: The edit distance is the minimum number of character insertions and character deletions required to turn one string into another. Compute the edit distance between each pair of the strings **he**, **she**, **his**, and **hers**.

4th ans:-

The edit distance between **he** and **she** = 1

The edit distance between **he** and **his** = 3

" " **he** and **hers** = 2

" " **she** and **his** = 4

" " **she** and **hers** = 3

" " **his** and **hers** = 3

" " " "