

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 .

SOLUTION

Assignment 3 Distance measures, Nearest-Neighbor Learning

Question 1

3 vectors are:

$$u = [1, 0.25, 0, 0, 0.5, 0]$$
$$v = [0.75, 0, 0, 0.2, 0.4, 0]$$
$$w = [0, 0.1, 0.75, 0, 0, 1]$$
$$\cos(u,v) = \frac{u \cdot v}{|u| \cdot |v|} = \frac{\sum u_i v_i}{\sqrt{\sum u_i^2} \sqrt{\sum v_i^2}}$$
$$\cos(u,v) = \frac{1 \times 0.75 + 0.25 \times 0 + 0 \times 0 + 0 \times 0.2 + 0.5 \times 0.4 + 0}{\sqrt{1^2 + 0.25^2 + 0^2 + 0^2 + 0.5^2 + 0} \cdot \sqrt{0.75^2 + 0^2 + 0^2 + 0.2^2 + 0.4^2 + 0}}$$
$$= \frac{0.75 + 0.2}{1.1456 \times 0.8932} = 0.75$$
$$\cos(u,w) = \frac{0.25 \times 0.1}{\sqrt{1.3125} \times \sqrt{1.3725}}$$
$$= \frac{0.025}{1.146 \times 1.171} = 0.02$$
$$\cos(v,w) = \frac{0}{0.8932 \times 1.354} = 0$$

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.

SOLUTION

Question 2
Five vectors are

A = 1111000000
B = 0100100101
C = 0000011110
D = 0111111111
E = 1011111111

Jaccard distance = $1 - \text{Jaccard similarity}$

(a) Between A and B
Jaccard sim = $\frac{|A \cap B|}{|A \cup B|} = \frac{1}{7}$
Jaccard distance = $1 - \frac{1}{7} = \frac{6}{7}$

(b) Between A and C
Jaccard sim = 0
Jaccard distance = $1 - 0 = 1$

(c) Between A and D
Jaccard sim = $\frac{3}{10}$
Jaccard distance = $1 - \frac{3}{10} = \frac{7}{10}$

(d) between A and E
Jaccard sim = $\frac{3}{10}$
Jaccard distance = $1 - \frac{3}{10} = \frac{7}{10}$

g between B and C

$$\text{Jaccard sim} = 1/7$$

$$\begin{aligned}\text{Jaccard distance} &= 1 - 1/7 \\ &= \underline{\underline{6/7}}\end{aligned}$$

h between B and D

$$\text{Jaccard sim} = 4/9$$

$$\begin{aligned}\text{Jaccard distance} &= 1 - 4/9 \\ &= \underline{\underline{5/9}}\end{aligned}$$

i between B and E

$$\text{Jaccard sim} = 3/10$$

$$\begin{aligned}\text{Jaccard distance} &= 1 - 3/10 \\ &= \underline{\underline{7/10}}\end{aligned}$$

h between C and D

$$\text{Jaccard sim} = 4/9$$

$$\begin{aligned}\text{Jaccard distance} &= 1 - 4/9 \\ &= \underline{\underline{5/9}}\end{aligned}$$

i between C and E

$$\text{Jaccard sim} = 4/9$$

$$\begin{aligned}\text{Jaccard distance} &= 1 - 4/9 \\ &= \underline{\underline{5/9}}\end{aligned}$$

j between D and F vectors

$$\text{Jaccard sim} = 8/10$$

$$\begin{aligned}\text{Jaccard distance} &= 1 - 8/10 \\ &= 2/10 = \underline{\underline{1/5}}\end{aligned}$$

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.

SOLUTION

Question 3

A = 1111000000
B = 0100100101
C = 0000011110
D = 0111111111
E = 1011111111

Manhattan distance is absolute sum of differences, b/w vectors.

Between A and B
Manhattan distance = 6

Between A and C
Manhattan distance = 8

Between A and D
Manhattan distance = 7

Between A and E
Manhattan distance = 7

Between B and C
Manhattan distance = 5

between B and D

Manhattan distance = 5

between B and E

Manhattan distance = 7

between C and D

Manhattan distance = 5

between C and E

Manhattan Distance = 5

between D and E

Manhattan Distance = 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**.

SOLUTION

Question 4
'He', 'she', 'his', 'hers'
Edit distance = $x + y - 2[LCS(x, y)]$

1st pair 'He' and 'she'
 $LCS = 2$
edit distance = $2 + 3 - 2(2)$
 $= 1$

2nd pair 'he' and 'his'
 $LCS = 1$
edit distance = $2 + 3 - 2(1)$
 $= 3$

3rd pair 'he' and 'hers'
 $LCS = 2$
edit distance = $2 + 4 - 2(2)$
 $= 2$

4th pair 'she' and 'his'

$$\text{LCS} = 1$$

$$\begin{aligned}\text{edit distance} &= 3 + 3 - 2(1) \\ &= 4\end{aligned}$$

5th pair 'she' and 'hers'

$$\text{LCS} = 2$$

$$\begin{aligned}\text{edit distance} &= 3 + 4 - 2(2) \\ &= 3\end{aligned}$$

6th pair 'his' and 'hers'

$$\text{LCS} = 2$$

$$\begin{aligned}\text{edit distance} &= 3 + 4 - 2(2) \\ &= 3\end{aligned}$$