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 Compute all three pairwise similarities among u, v ,w

Solution :-

vectors are:

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

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

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

cosine similarity $(u, v) = \frac{u \cdot v}{|u| \cdot |v|}$

$= \frac{\sum u \cdot v}{\sqrt{\sum u^2} \times \sqrt{\sum v^2}}$

$$\cos(u,v) = \frac{1 \times 0.75 + 0.25 \times 0 + 0 \times 0 + 0.5 \times 0.2 + 0 \times 0.4}{\sqrt{1^2 + 0.25^2 + 0^2 + 0.5^2 + 0} \cdot \sqrt{0.75^2 + 0^2 + 0^2 + 0.2^2 + 0.4^2}}$$

$$= \frac{0.75 + 0.1}{1.1456 \times 0.8732} = 0.95$$

$$\cos(u,w) = \frac{0.25 \times 0.1}{\sqrt{1.3185} + \sqrt{1.3225}}$$

$$= \frac{0.025}{1.145 \cdot 1.254} = \frac{0.025}{1.437} = 0.02$$

$$\cos(v,w) = \frac{0}{0.8732 \times 1.254} = \frac{0}{\quad}$$

Question 2:

Here are five vectors in a 10-dimensional space:

1111000000 0100100101 0000011110 0111111111 1011111111

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

Solution :-

2) Five vector are

A: 1111000000

B: 0100100101

C: 0000011110

D: 0111111111

E: 1011111111

(A) Jaccard distance = 1 - Jaccard similarity

Between A & B

Jaccard sim = $\frac{|A \cap B|}{|A \cup B|} = 1/7$

Jaccard Distance = $1 - 1/7 = 6/7$

(B) Between A & C

Jaccard Sim = 0

Jaccard distance = 1 - 0 = 1

(C) Between A & D

Jaccard Sim = 3/10

Jaccard distance = $1 - 3/10 = 7/10$

(D) between A & E

Jaccard Sim = 3/10

Jaccard distance = $1 - 3/10$

$= 3\ 7/10$

(e) between B & C

Jaccard sim = 1/7

Jaccard distance: 1 - 1/7

= 6/7

(f) between B & D

Jaccard sim = 4/9

Jaccard distance: 1 - 4/9 = 5/9

(g) between B & E

Jaccard Sim = 3/10

Jaccard distance = 1 - 3/10 = 7/10

(h) between C & D

Jaccard sim = 4/9

~~Jaccard~~ Jaccard Distance: 5/9

(i) between C & E

Jaccard sim = 4/9

Jaccard Distance = 1 - 4/9 = 5/9

(j) between D & E vectors

Jaccard Sim = 8/10

Jaccard distance = 1 - 5/10

= 2/10 = 1/5

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 :-

(3)

A: 1111000000

B: 0100100101

C: 0000011110

D: 0111111111

E: 1011111111

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

Between A & B

- Manhattan Distance = 6

Between A & C

Manhattan Distance = 8

Between A & D

Manhattan Distance = 7

Between A & E

Manhattan Distance = 7

Between B & C

Manhattan Distance = 5

Between B & D

Manhattan distance = 5

Between B & E

Manhattan Distance = 7

Between C & D

Manhattan distance = 5

between C & E

Manhattan Distance = 5.

between D & 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 :-

④ "He", 'she', 'his', 'hers'

Edit distance = x+y-2 [LCS(x,y)]

1st pair 'He' & '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' & 'hers'

LCS = 2

edit distance = 2+4-2(2)

= 2

4th pair 'she' and 'his'

LCS = 1

edit distance = 3+3-2(1)

= 4