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.

Ans:

$$cos(u, v) = 0.95/(\sqrt{1}.3125 * \sqrt{0}.7625)$$

$$cos(u, v) = 0.95$$

$$cos(u, w) = 0.025/(\sqrt{1}.3125 * \sqrt{1}.5725)$$

$$cos(u, w) = 0.017$$

$$cos(v, w) = 0.025/(\sqrt{0}.7625 * \sqrt{1}.5725)$$

$$cos(v, w) = 0.023$$

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.

Ans*:*

Let:

v1 = 1111000000

v2 = 0100100101

v3 = 0000011110

v4 = 0111111111

v5 = 1011111111

d(v1,v2) = 1 - (1/7) = 6/7

d(v1,v3) = 1

d(v1,v4) = 1 - (3/10) = 7/10

d(v1,v5) = 1 - (3/10) = 7/10

d(v2,v3) = 1 - (1/7) = 6/7

d(v2,v4) = 1 - (4/9) = 5/9

d(v2,v5) = 1 - (3/10) = 7/10

d(v3,v4) = 1 - (4/9) = 5/9

d(v3,v5) = 1 - (4/9) = 5/9

d(v4,v5) = 1 - (8/10) = 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.

Ans:

Let:

v1 = 1111000000

v2 = 0100100101

v3 = 0000011110

v4 = 0111111111

v5 = 1011111111

d(v1,v2) = (1-0) + (1-1) + (1-0) + (1-0) + (0-1) + (0-0) + (0-0) + (0-1) + (0-0) + (0-1) = 6

d(v1,v3) = 8

d(v1,v4) = 7

d(v1,v5) = 7

d(v2,v3) = 6

d(v2,v4) = 5

d(v2,v5) = 7

d(v3,v4) = 5

d(v3,v5) = 5

d(v4,v5) = 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`**.

Ans:

d(he, she) = 1

d(he, his) = 3

d(he, hers) = 2

d(she, his) = 4

d(she, hers) = 3

d(his, hers) = 5