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.

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

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

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

$$\cos(u,v) = \frac{u \times v}{|u| \times |v|} = \frac{0.75 \times 0.02}{1.145 \times 0.873} = 18°$$

$$\cos(v,w) = \frac{v \times w}{|v| \times |w|} = \frac{0}{0.873 \times 1.25} = 0°$$

$$\cos(u,w) = \frac{u \times w}{|u| \times |w|} = \frac{0.025}{1.145 \times 1.25} = 89°$$

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.

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

- Jaccard dis(A,B)=1-1/7=6/7
- Jaccard dis(A,C)=1-0/8=1
- Jaccard dis(A,D)=1-3/10=7/10
- Jaccard dis(A,E)=1-3/10=7/10

- Jaccard dis(B,C)=1-1/7=6/7
- Jaccard dis(B,D)=1-4/9=5/9
- Jaccard dis(B,E)=1-3/10=7/10
- Jaccard dis(C,D)=1-4/9=5/9
- Jaccard dis(C,E)=1-4/9=5/9
- Jaccard dis(D,E)=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.

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

- Manhattan dis(A,B)=6
- Manhattan dis(A,C)=8
- Manhattan dis(A,D)=7
- Manhattan dis(A,E)=7
- Manhattan dis(B,C)=6
- Manhattan dis(B,D)=5
- Manhattan dis(B,E)=7
- Manhattan dis(C,D)=5
- Manhattan dis(C,E)=5
- Manhattan dis(D,E)=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**.

- Edit dis(he,she)=1
- Edit dis(he,his)=3
- Edit dis(he,hers)=2
- Edit dis(she.his)=4
- Edit dis(she,hers)=3
- Edit dis(his,hers)=3