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.

Given data is:

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]

|u| = $\sqrt{1^2 + 0.25^2 + 0^2 + 0^2 + 0.5^2 + 0^2}$ = $\sqrt{1 + 0.625 + 0.25}$ = 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

a) cos(u, v) = $\frac{u.v}{|u||v|} = \frac{0.75+0.02}{1.145\ x\ 0.873} = 0.95$, Θ = 18^0

b) cos(v, w) = $\frac{0 \text{ x } 0.75 + 0 \text{ x } 0.1 + 0 \text{ x } 0.75 + 0.2 \text{ x } 0 + 0.4 \text{ x } 0 + 0}{0.873\ x\ 1.25} = 0$, Θ = 90^0

c) cos(u, w) = $\frac{1 \text{ x } 0 + 0.025 + 0 + 0 + 0 + 0}{1.145\ x\ 1.25} = 0.017$, Θ = 89^0

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.

Given five vectors in 10 – D

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

Jaccard distance:

(A, B) = 1 - $\frac{1}{7} = \frac{6}{7}$, (A, C) = 1 - $\frac{0}{8}$ = 1, (A, D) = 1 - $\frac{3}{10} = \frac{7}{10}$, (A, E) = 1 - $\frac{3}{10} = \frac{7}{10}$

(B, C) = 1 - $\frac{1}{7} = \frac{6}{7}$, (B, D) = 1 - $\frac{4}{9} = \frac{5}{9}$, (B, E) = 1 - $\frac{3}{10} = \frac{7}{10}$

(C, D) = 1 - $\frac{4}{9} = \frac{5}{9}$, (C, E) = 1 - $\frac{4}{9} = \frac{5}{9}$

(D, E) = 1 - $\frac{8}{10} = \frac{2}{10}$

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.

(A, B) = |(1 - 0)| + |(1 – 1)| + |(1 - 0)| + |(1 - 0)| + |(0 -1)| + |(0 - 1)| + |(0 - 0)| +|(0 - 0)| + |(0 - 1)| + |(0 - 1)| = 6; (A, C) = 8; (A, D) = 7; (A, E) = 7

(B, C) = 6; (B, D) = 5; (B, E) = 7

(C, D) = 5; (C, E) = 5

(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.

(He, She) edit distance = 1
(He, His) edit distance = 3
(He, Hers) edit distance = 2
(She, His) edit distance = 4
(She, Hers) edit distance = 3
(His, Hers) edit distance = 3