### **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:

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| = √𝟏 𝟐 + 𝟎. 𝟐𝟓𝟐 + 𝟎 𝟐 + 𝟎 𝟐 + 𝟎. 𝟓 𝟐 + 𝟎 𝟐 = 𝟏. 𝟏𝟒𝟓

|v| = √𝟎. 𝟕𝟓𝟐 + 𝟎 𝟐 + 𝟎 𝟐 + 𝟎. 𝟐 𝟐 + 𝟎. 𝟒 𝟐 + 𝟎 𝟐 = 𝟎. 𝟖𝟕𝟑

|w| = √𝟎 𝟐 + 𝟎. 𝟏 𝟐 + 𝟎. 𝟕𝟓𝟐 + 𝟎 𝟐 + 𝟎 𝟐 + 𝟏 𝟐 = 𝟏. 𝟐𝟓

cos (u, v) = u\*v |𝑢|∗|𝑣| = 0.75 + 0.02 1.145 \* 0.873 → 𝜽 = 18 𝑑𝑒𝑔𝑟𝑒𝑒𝑠.

cos (v, w) = v\*w |v|∗|w| = 0 0.873\*1.25 → 𝜽 = 0 𝑑𝑒𝑔𝑟𝑒𝑒𝑠.

cos (u, w) = u\*w |𝑢|∗|w| = 0.025 1.145 \* 1.25 → 𝜽 = 89 𝑑𝑒𝑔𝑟𝑒𝑒𝑠.

**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:

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

Jaccard Distance (A, B) = 1 – (1/7) = 6/7

Jaccard Distance (A, C) = 1 – (0/8) = 1

Jaccard Distance (A, D) = 1 – (3/10) = 7/10

Jaccard Distance (A, E) = 1 – (3/10) = 7/10

Jaccard Distance (B, C) = 1 – (1/7) = 6/7

Jaccard Distance (B, D) = 1 – (4/9) = 5/9

Jaccard Distance (B, E) = 1 – (3/10) = 7/10

Jaccard Distance (C, D) = 1 – (4/9) = 5/9

Jaccard Distance (C, E) = 1 – (4/9) = 5/9

Jaccard Distance (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.

Solution:

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

Manhattan distance of A, B = 6

Manhattan distance of A, C = 8

Manhattan distance of A, D = 7

Manhattan distance of A, E = 7

Manhattan distance of B, C = 6

Manhattan distance of B, D = 5

Manhattan distance of B, E = 7

Manhattan distance of C, D = 5

Manhattan distance of C, E = 5

Manhattan distance of 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**.

Solution:

The edit distance between he and she = 1

The edit distance between he and his = 3

The edit distance between he and hers = 2

The edit distance between she and his = 4

The edit distance between she and hers = 3

The edit distance between his and hers = 3