Distance Measures

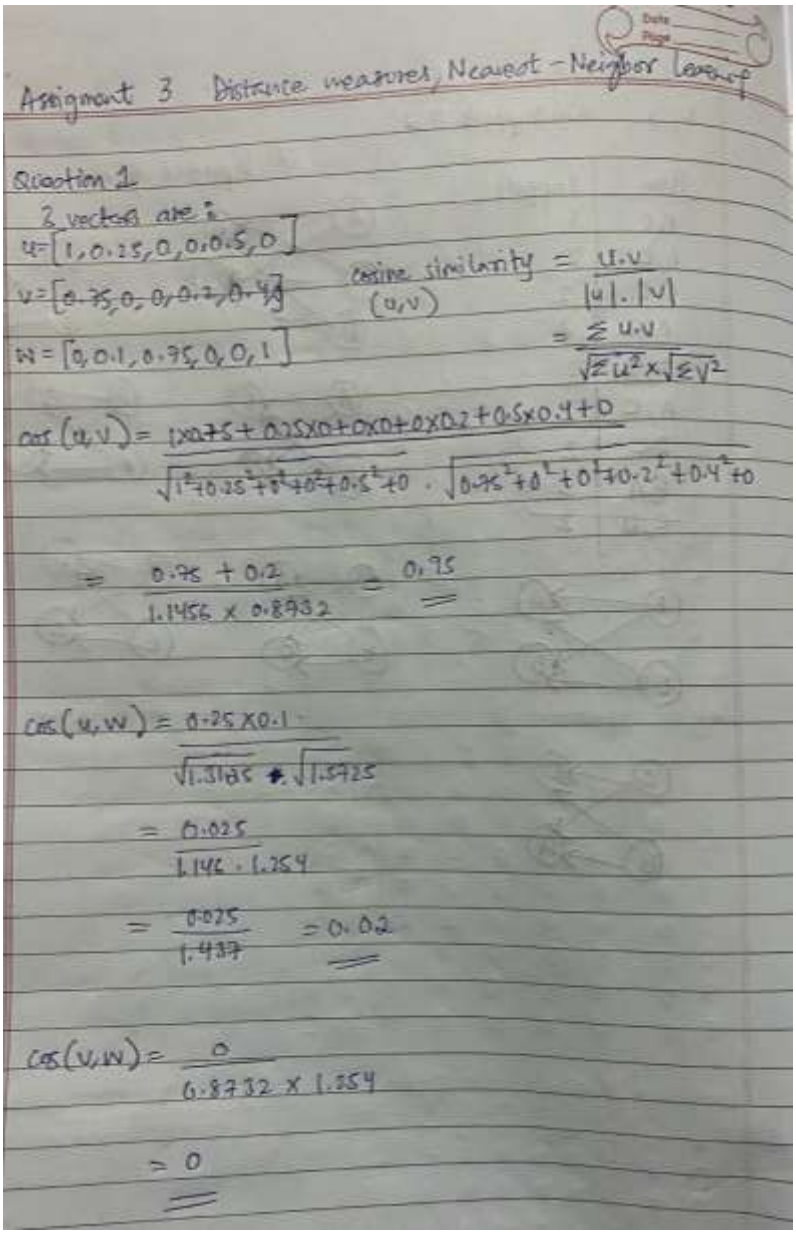
QUESTION 1:

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



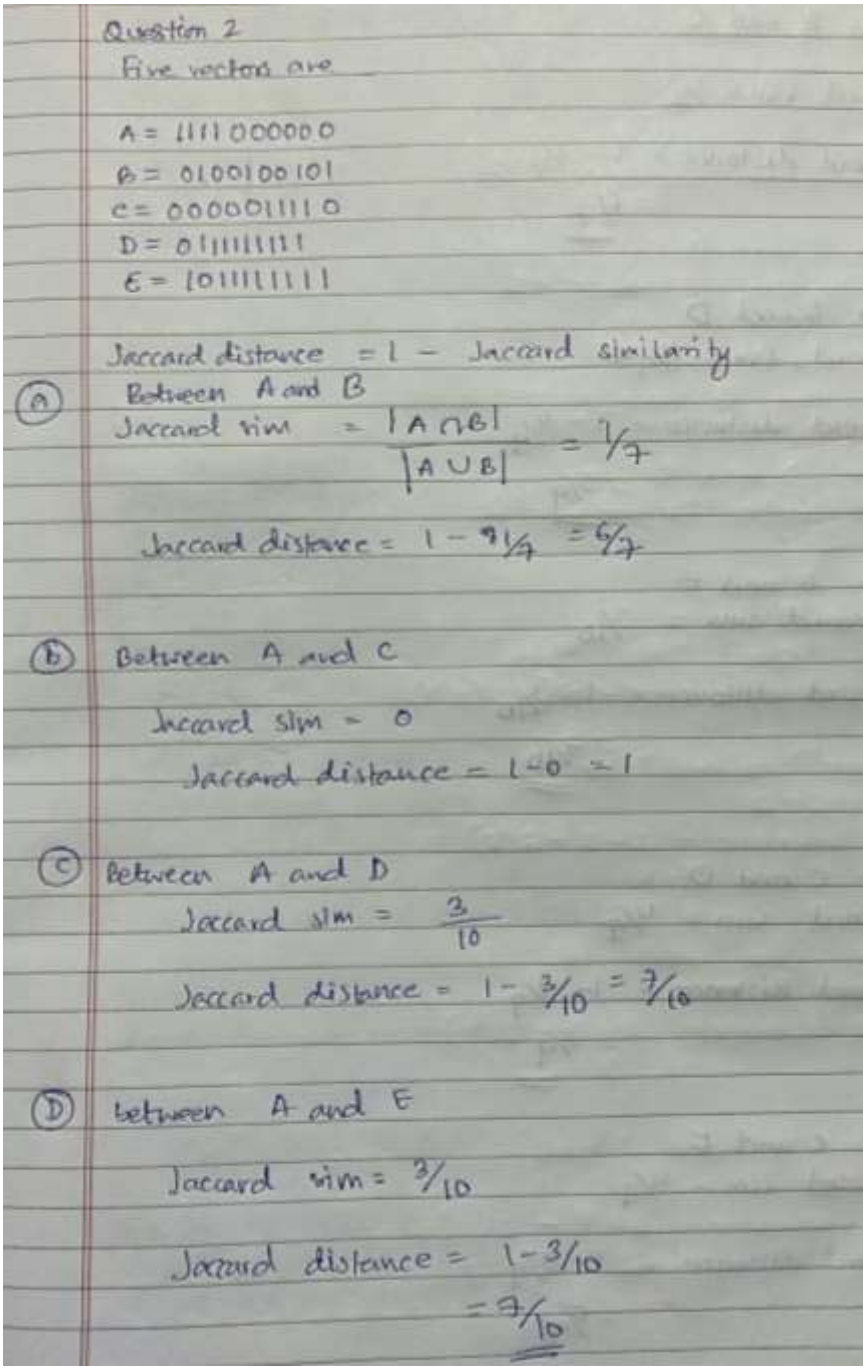
QUESTION 2:

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



(e) between B and C

Jaccard sim = 1/7

Jaccard distance = 1 − 1/7

= 6/7

(f) between A and D

Jaccard sim = 4/9

Jaccard distance = 1 − 4/9

= 5/9

(g) between B and E

Jaccard sim = 3/10

Jaccard distance = 1 − 3/10

= 7/10

h between C and D

Jaccard sim = 4/9

Jaccard distance = 1 − 4/9

= 5/9

i between C and E

Jaccard sim = 4/9

Jaccard distance = 1 − 4/9

= 5/9

j between D and E vector

Jaccard sim = 8/10

Jaccard distance = 1 − 8/10

= 2/10 = 1/5

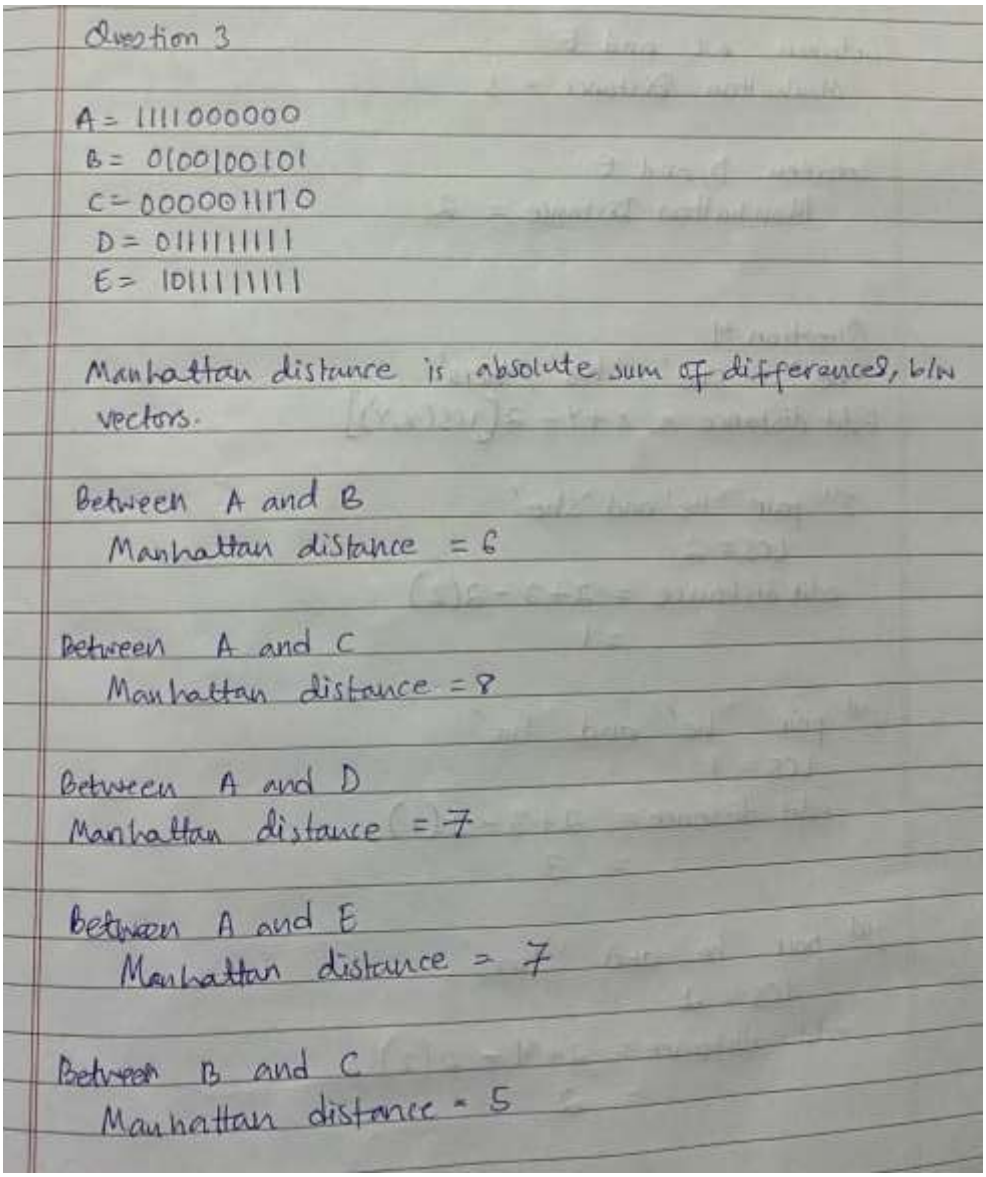
QUESTION 3:

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



between B and D
Manhattan distance = 5

between B and E
Manhattan distance = 7

between C and D
Manhattan distance = 5

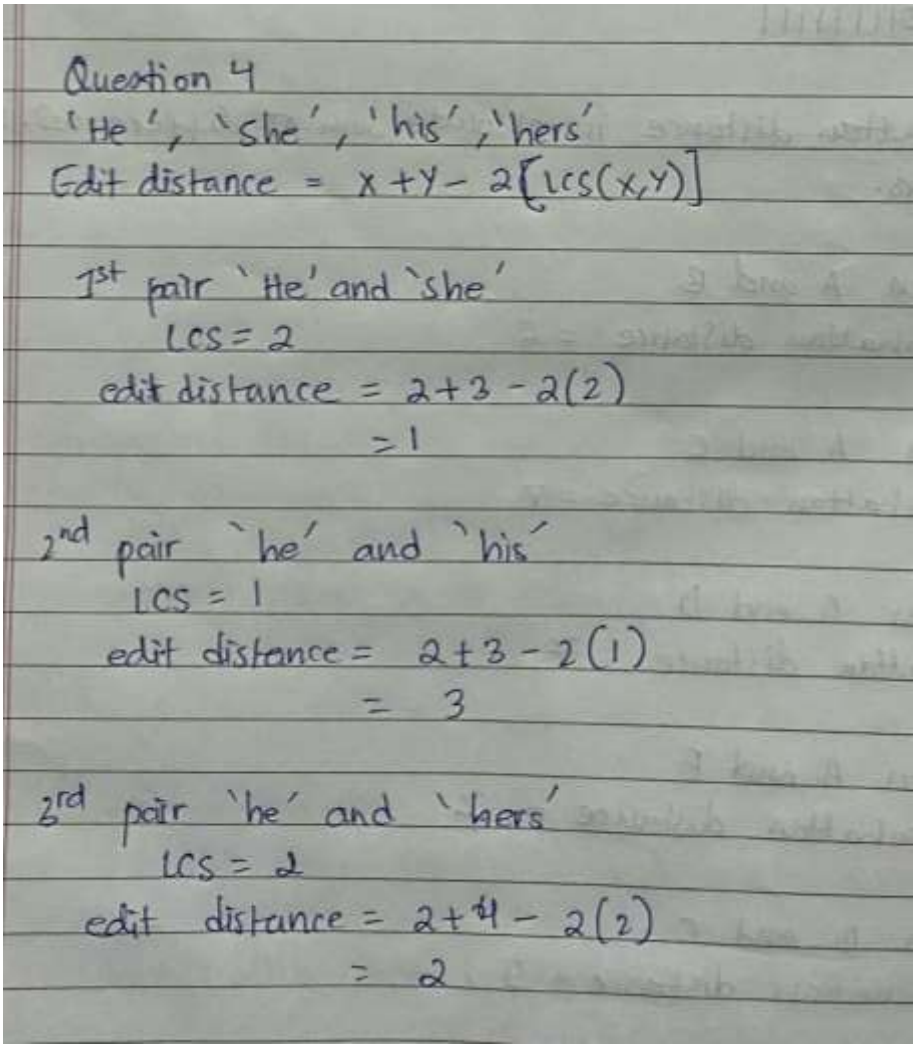
between C and E
Manhattan Distance = 5

between D and 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



4th pair 'She' and 'his'

LCS = 1

edit distance = 3 + 3 − 2(1)

= 4

5th pair 'she' and 'hers'

LCS = 2

edit distance = 3 + 4 − 2(2)

= 3

6th pair 'his' and 'hers'

LCS = 2

edit distance = 3 + 4 − 2(2)

= 3