V2V2V2V2 V2V2V2V2 V2V2V2V2 V2V2V2V2 V2V2V2V2 V2V2

National University of Computer & Emerging Sciences
FAST-Karachi Campus
CS4051- Information Retrieval
Quiz#4 (Double Weightage)
Dated: May 09, 2024 Marks: 20
Time: 20 min.
Std-ID: _____Sol_____

Question No. 1

Consider the following examples for the task of text classification

Dataset	**DocID**	**Features- Words in documents**	**Class Fruit=Yes/No**
Training set	1	Orange, Orange, Lemon	No
	2	Orange, Red, Blue	No
	3	Apricot, Apple, Mango	Yes
	4	Apple, Banana , Orange	Yes
	5	Apple, Orange, Melon	Yes
Test set	6	Orange, Mango, Melon	?
	7	Orange, Red, Lemon	?

a. Build a vector space representation of these documents using Boolean representation. Using Rocchios algorithm find the two central mass vectors for classes $\mu_{(Fruit=Yes)}$ and $\mu_{(Fruit=No)}$

Dictionary = {Apple, Apricot, Banana, Blue, Lemon, Mango, Melon, Orange, Red}

$\mu_{(Fruit=Yes)}$ = **< 1, 1/3,2/3, 0, 0, 1/3, 1.3. 2/3, 0>**

$\mu_{(Fruit=No)}$ = **< 0, 0, 0, 1/2, 1/2, 0, 0, 1, 1/2 >**

b. Classify the two test instances using the central mass vectors for the 2 classes in part a.

Dist (D6, $\mu_{(Fruit=Yes)}$) < Dist(D6, $\mu_{(Fruit=No)}$) => D6 belong to class Fruit=Yes

Dist (D7, $\mu_{(Fruit=Yes)}$) < Dist(D7, $\mu_{(Fruit=No)}$) => D7 belong to class Fruit=Yes

V2V2V2V2 V2V2V2V2 V2V2V2V2 V2V2V2V2 V2V2V2V2 V2V2

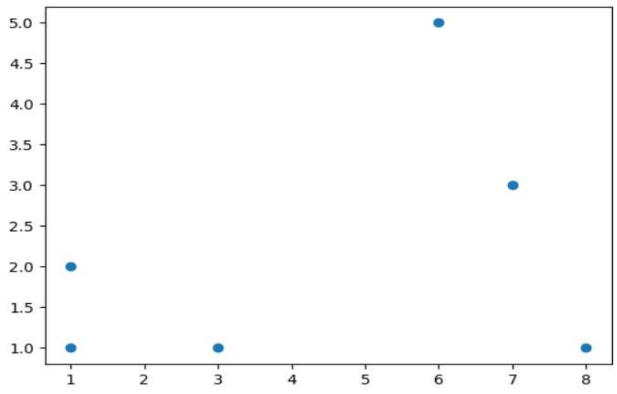
Question No. 2

Consider the following data points in 2D space.

P1(3,1), P2(1,2), P3(1,1), P4(8,1), P5(7,3), P6(6,5), P7(6,7)

Apply Bottom-Up version of Hierarchical Agglomerative Clustering (HAC) for the given data points, using Single Link approach for merging. Show each step of the algorithm

Consider the following plot in2 D space with the given points:



For Bottom Up Hierarchal clustering, Point P3 and P2 are more similar and will be merged first.

Second Point P1 will be merged with the cluster C1(P2, P3) to produced C2(P1, C1).

Thirdly, we will merge P6 and P7 to produced C3(P6, P7) It will merge later P5 to produced C4(C3, P5). Similarly, it will merge the P4 to it to produced C5(C4, P4) and in the final merge C2 and C5 will be merged.