CS 422

DATA MINING

HOME WORK – 7

SUBMITTED BY :

SACHIN KRISHNA MURTHY

CWID : A20354077

<u>Exercise 7.2.6:</u> Consider the space of strings with edit distance as the distance measure. Give an example of a set of strings such that if we choose the clustroid by minimizing the sum of the distances to the other points we get one point as the clustroid, but if we choose the clustroid by minimizing the maximum distance to the other points, another point becomes the clustroid.

Solution :

- The solution for this problem is practically not possible.
- If we have to get a solution for this problem, we need to know that there can be no single set of two string with min distance or max distance, but their sum can be different.
- We have to choose strings in such a way that if distance between string1 and string2 is the least among the whole set, the distance between these two strings and the rest of the strings should be very high, making the sum of distance for string1 and string2 higher.
- Also the distance between the other strings should be more than the least but a smaller value.

Consider the below example for finding distance:

	ghkl	hnm	jok
kghi	4	5	5
jok	5	6	
hnm	5		

	Max	Sum
kghi	5	14
jok	6	16
hnm	6	18
ghkl	5	14

- In this case there are two cases “kghi” and “ghkl” with least max value of 5.
- The same ones have least value for sum i,e “kghi” and “ghkl” has the sum of 14
- It is not possible to get different ones for these two cases.