CSE 6363 Machine Learning

Homework/Project III

*Submitted by*

Mohammed Ghouse Parvez Mohammed Akmal

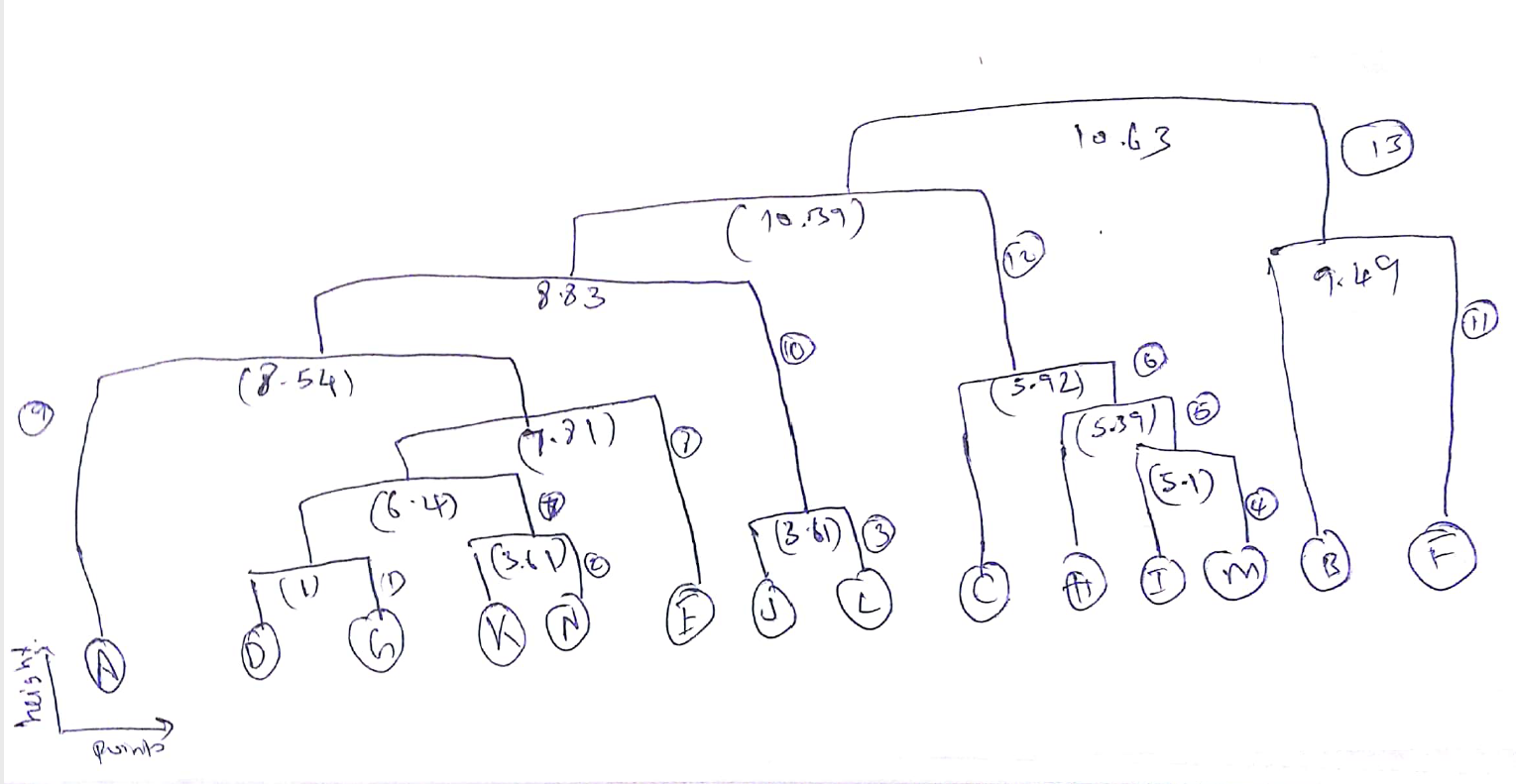
1001625800

**1a)**

Lets first name each point in the data set using alphabets as follows:

1. 170,57,32
2. 192,95,28
3. 150,45,30
4. 170,65,29
5. 175,78,35
6. 185,90,32
7. 170,65,28
8. 155,48,31
9. 160,55,30
10. 182,80,30
11. 175,69,28
12. 180,80,27
13. 160,50,31
14. 175,72,30

The tree below is the single linkage cluster hierarchy tree.The numbers in the () represents the distance between the clusters and the rounded number indicates the order of merge operations.

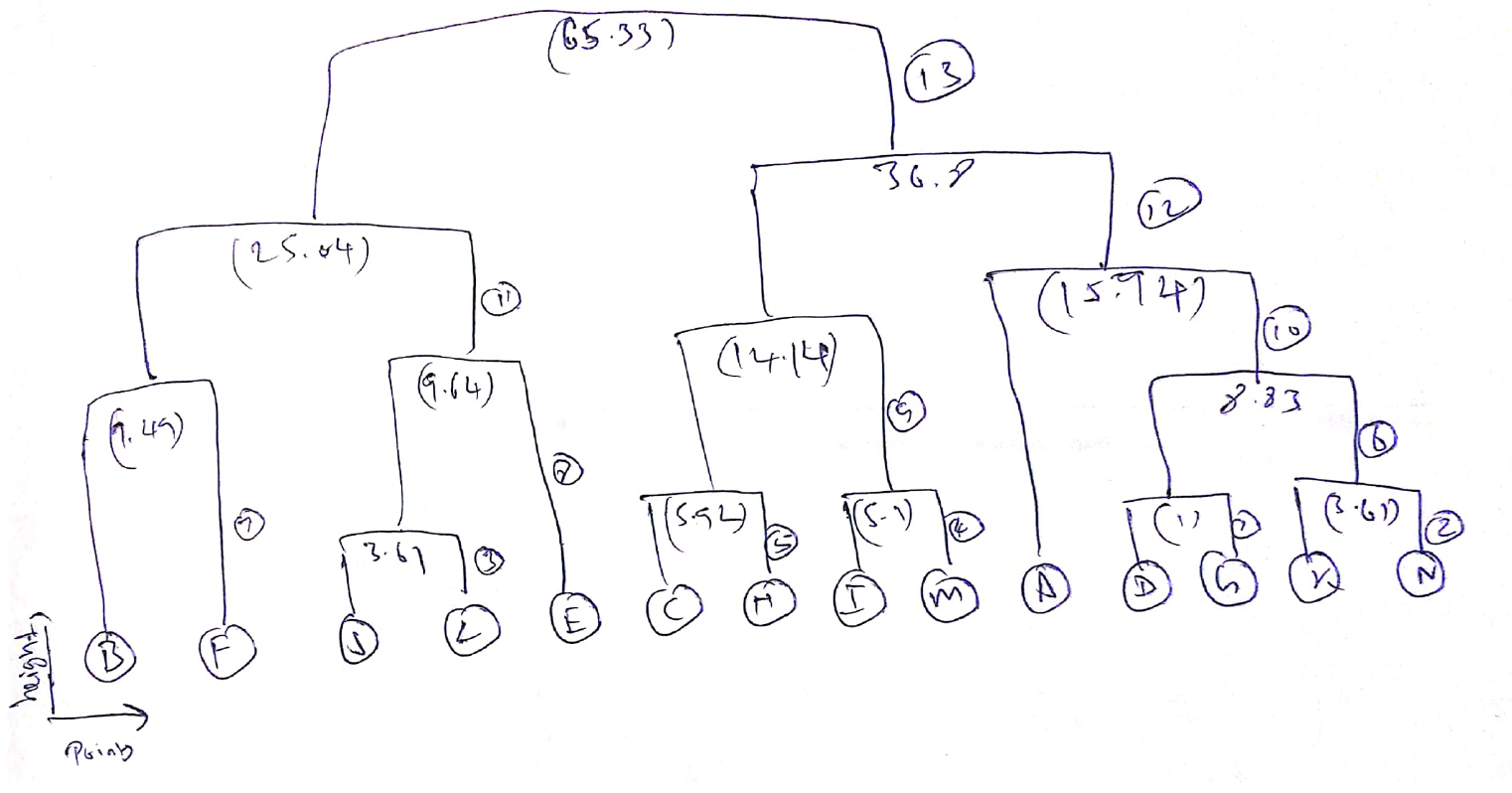


**1b)**

Lets first name each point in the data set using alphabets as follows:

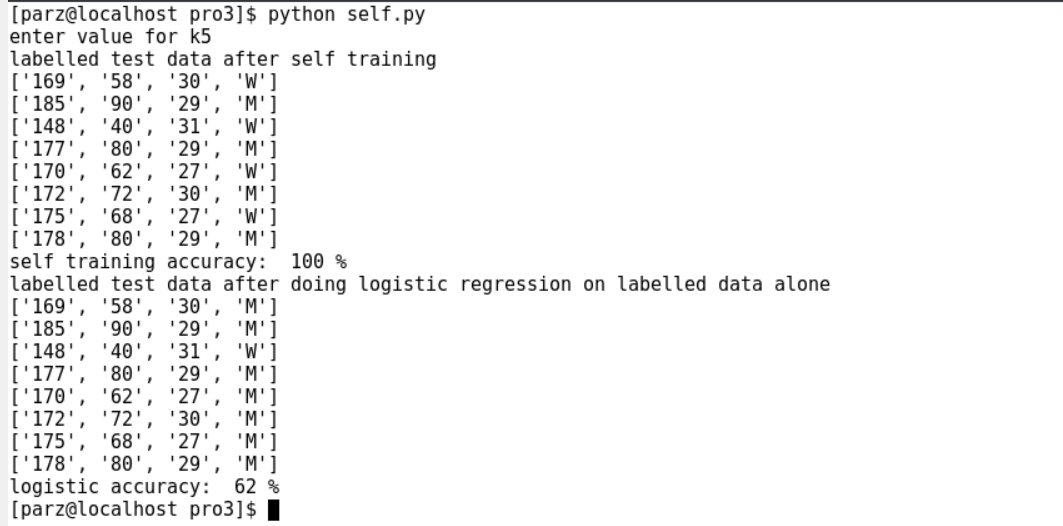
1. 170,57,32
2. 192,95,28
3. 150,45,30
4. 170,65,29
5. 175,78,35
6. 185,90,32
7. 170,65,28
8. 155,48,31
9. 160,55,30
10. 182,80,30
11. 175,69,28
12. 180,80,27
13. 160,50,31
14. 175,72,30

The tree below is the Maximum(complete) linkage cluster hierarchy tree.The numbers in the () represents the distance between the clusters and the rounded number indicates the order of merge operations.



**2a)** A self-training system using a logistic regression classifier was implemented and the code is in the codes folder.

**2b)**The following is the result obtained from executing the program with K=5:



The accuracy obtained by implementing the self training (100 %)algorithm is much higher than the simple logistic regression accuracy (62%). The accuracy of simple logistic regression is 62 which is really low, this may be due to the small dataset used for training. Since the self training algorithm uses the unlabeled data in its algorithm it can learn from more data when compared to logistic regression.