
# ES 615 Nature Inspired Computing
# Fall 2021-22

Instructor Prof. Nithin George

## **Assignment -5** Grading using Hierarchical Clustering

Name - Devanshu Thakar

Roll no. – 18110174

**Problem** : Given names and final marks (Out of 100) of students of a course at IITGN. Using a hierarchical clustering approach, assign grades to these students. Assume that the Instructor has decided to allot only A, B, C and D grades to students. How will the overall grading change if the academic office makes it mandatory to have at least ONE F grade in a course of more than 20 students? Plot the dendrogram.

**Steps in Hierarchical Clustering algorithm :**

1. Initially all the *N* points are in their own cluster. There are *N* cluster total

2. Generate the distance matrix with these *N* clusters.

3. Combine those two cluster which have minimum distance into a single cluster. Append this entry into the *linkage_matrix*.

4. Update the distance matrix – compute the distance of newly formed combined cluster with the remaining clusters into the distance matrix.

5. Repeat this until only one cluster is remaining.

The type of hierarchical clustering used in this problem was **maximum linkage** type. In other words when we need to compute distance between cluster *A* and with a cluster formed by combining *B* and *C* i.e. (*B*,*C*) we choose amongst the maximum of distance between *A* and *B* and *A* and *C*.

$$\text{Distance}(A, (B,C)) = \max(\text{Distance}(A, B), \text{Distance}(A,C))$$

The python program hierarchical_clustering.py was coded implementing the hierarchical clustering algorithm. The scipy.cluster.hierarchy library was used to plot the dendrogram. The dendrogram from the code is shown below. The number 0,1,2…25 indicates student A,B, ….Z.

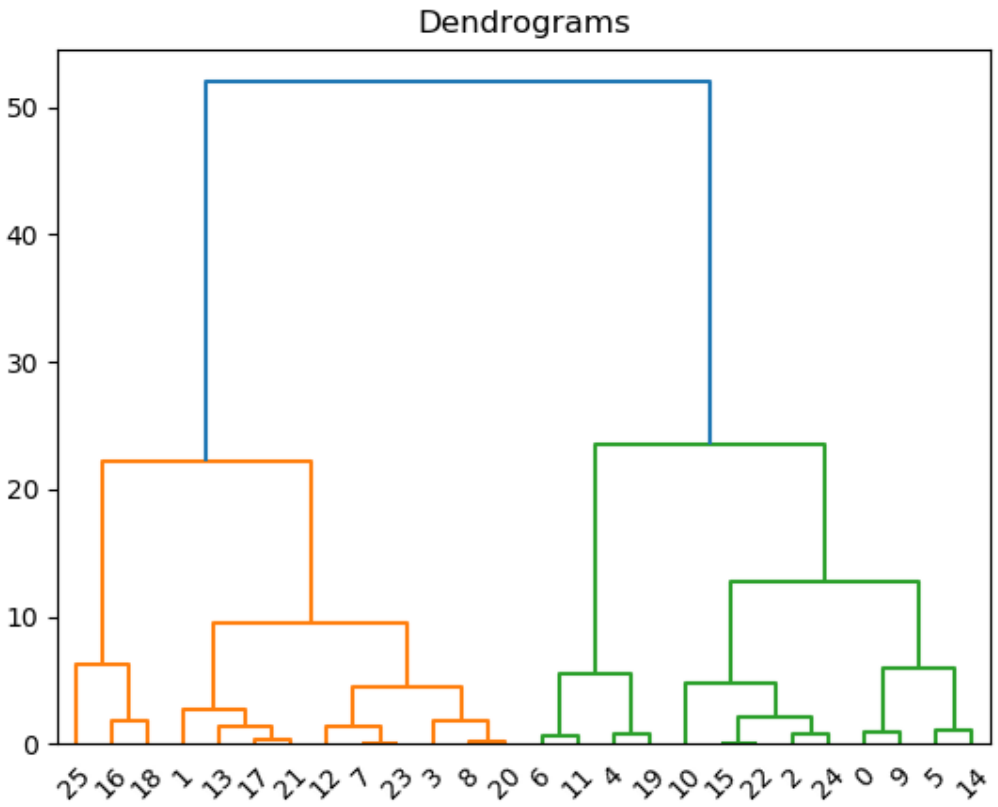


Figure 1 Plot of dendrogram

To assign grades of A, B, C and D we can draw a horizontal line which intersects four vertical lines one such line drawn is shown in Figure 2.

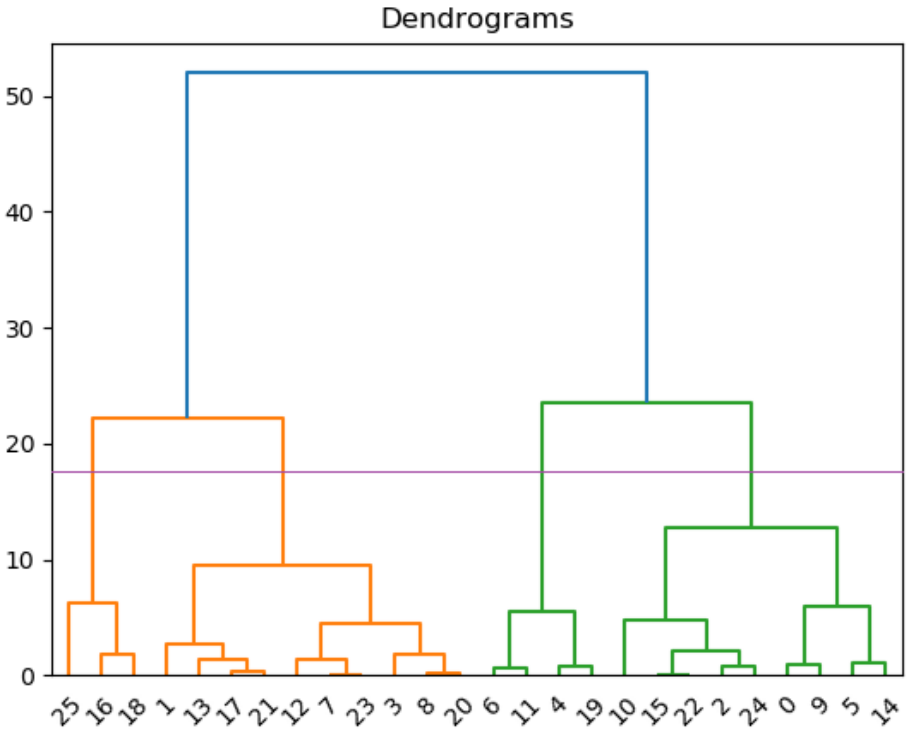


Figure 2 Plot of dendrogram divided into four clusters

The avergae marks in each cluster was computed and depending on that the grade was assigned to a particular cluster.

The list of students with students with grades obtained from the code is as follows :

Grade A : G L E T

Grade B : K P W C Y A J F O

Grade C : B N R V M H X D I U

Grade D : Z Q S

If the academic office mandates to have at least ONE F grade, then we can divide the cluster with grade – D into two clusters, the one having higher average of marks can be assigned grade-D and the other having lower average marks can be assigned grade F. The dendrogram for the same is shown below :

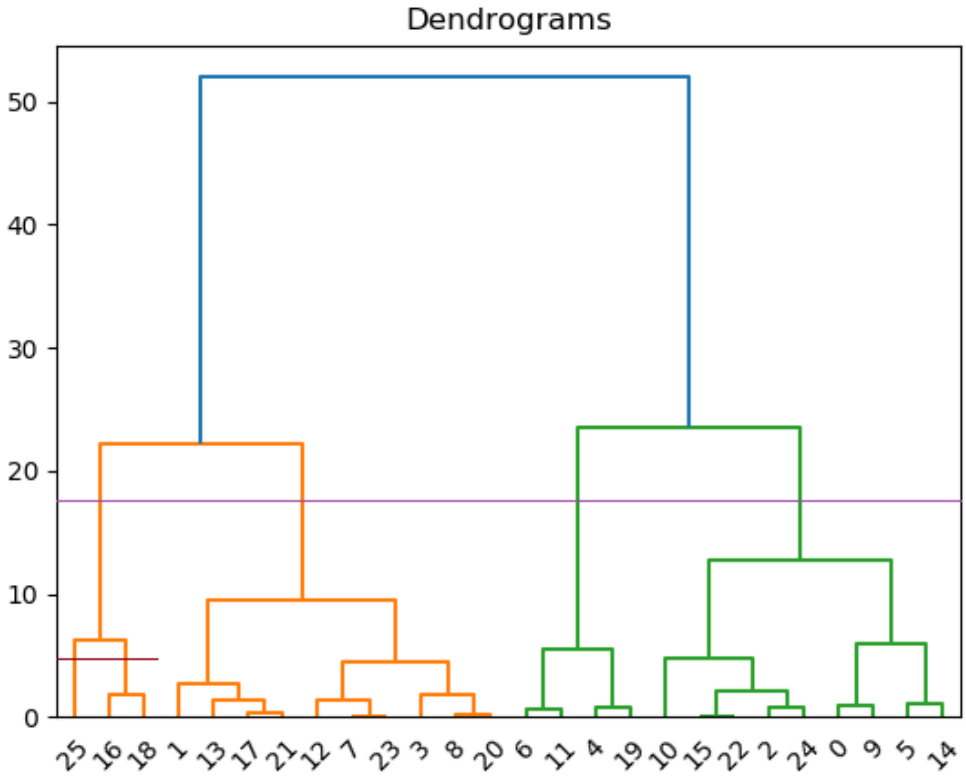


Figure 3 Plot of dendrogram with clusters of grade A,B,C, D and F

The following change in the overall grading will be seen if at least one F grade is to be given.

Final Grades if atlead ONE F grade is to be given.

Grade A : G L E T

Grade B : K P W C Y A J F O

Grade C : B N R V M H X D I U

Grade D : Q S

Grade F : Z

References :

1. https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html