**Optional Project-3**

Q1) Steps I followed in order to implement the Hierarchical Clustering are as given below:

1. Imported the Dataset
2. Treated Every Data Point as an Individual Cluster
3. Found the distance matrix
   1. Find the least distance between two clusters
   2. Combine the obtained clusters
   3. Distance Matrix is changed by :
      1. If Complete\_Linkage:
         1. Maximum distance of both cluster is assigned to the distance matrix for the distance between the combined cluster and other clusters
      2. If Single\_Linkage:
         1. Minimum distance of both cluster is assigned to the distance matrix for the distance between the combined cluster and other clusters
4. Repeat Step 3 till we get one cluster.
5. In order to get the given number of clusters we select that number of row from the last in the output numpy array which we get from the HAC function in the code file.
6. For accuracy we first assign each cluster the labels where the maximum of the data points in the cluster have the same label is considered to be the label of the cluster.
7. We compare the labels with the true labels and find the each cluster accuracy
8. For weighted accuracy we find the product of every cluster and its weight ratio and then sum it up.

**Results:**

As we can see from above obtained results that weighted accuracy for total number of 8 clusters is highest amongst both single linkage clustering and complete linkage clustering but comparing both we can say that single linkage clustering slightly edges onto the complete linkage clustering as both of the clustering techniques have weighted accuracy of 55.83% and 53.33% respectively

**Conclusion**

### Accuracy of single linkage > Accuracy of complete linkage

Q2) Steps I followed in order to achieve self-trained KNN Model:

1. Imported The DataSet
2. Split the the dataset where Ds contains the first 10 elements of each material and remaining in Du which is unlabelled
3. Implementation of self learning:
   1. Perform KNN on given data points using Ds as training dataset
   2. Update Ds append the given test data points with their predictions
   3. Delete the data points from Du.
   4. Repeat the steps a,b,c till Du has no points
4. Implement Normal KNN
5. Compared all the different accuracies we get when we insert different numbers of data points on every iteration.

**Results**

After performing self-training KNN model and Normal KNN model we find that the accuracy on the given particular dataset we can say that when we add 3 and 4 data points the accuracy slightly edges to Normal KNN as both have accuracies of 55.55 and 56.66 respectively compared to Normal KNN accuracy of 53.33 From above experiment when we add 5 points in every iteration we get less accuracy than the accuracy of Normal KNN.

**Conclusion**

Thus we can say that whenever we add less number of points we get more accuracy then normal KNN as the model is trained by itself and is gaining knowledge in every iteration that's why the lesser the data point in inserted in every iteration the more will be the accuracy