Subject: 19CSE305

Lab Session: 08-10

Notes:

1. Please read the assignment notes carefully and comply with the guidelines provided.
2. Code should be checked into GitHub and the report to TurnItIn. Once done, please submit your assignments in Teams.
3. Code non-availability in GitHub shall be marked as zero.
4. Any content copy (statements, figures, codes etc.) from anywhere shall attract a penalty of 10 marks. If you obtain content from anywhere for illustration purposes, please cite the source to avoid penalty.
5. Snapshot / screenshot of code and results not allowed in the report. You may copy content from your own code & results and add it to the report.
6. Provide data, code snippets or illustrations to support your answer, as applicable.

Please use the data associated with your own project.

Refer:

- **https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html**
- **https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html**
- **https://scikit-learn.org/stable/auto_examples/cluster/plot_agglomerative_dendrogram.html**
- **https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SequentialFeatureSelector.html#**

Main Section (Mandatory):

A1. Merge your train & test sets and remove the class labels. Use k-means algorithm with k = 3 or 5 (based on your dataset) to form the clusters.

A2. Determine the ideal k value for your dataset. Determine the clusters for a range of k ϵ [1,31]. Use elbow method to determine the ideal value of k based on average Euclidean distance from cluster center.

A3. Use Agglomerative Clustering for hierarchical clustering of your data. Plot the dendrogram to visualize the clusters.

A4. Perform sequential feature selection algorithms for arriving at the optimal number of features needed for your dataset.

A4. Perform PCA on your data set. Observe the variances along the principal components. Identify the features needed to capture 95% of data variance. Let's say the number of features needed is ***K***.

A5. Transform your data with the initial ***K*** PC's. Calculate the accuracy using the transformed dataset.

Report Assignment:

1. Update the report according to your weekly progress and keep submitting it every week.