Fall 2022 5710 Machine Learning: Assignment 6

Programming elements:
Hierarchical Clustering

In class programming:

1) (Provide only mathematical solutions for this question) Six points with the following attributes are given, calculate and find out clustering representations and dendrogram using Single, complete, and average link proximity function in hierarchical clustering technique.

point	x coordinate	y coordinate
p1	0.4005	0.5306
p2	0.2148	0.3854
p3	0.3457	0.3156
p4	0.2652	0.1875
p5	0.0789	0.4139
p6	0.4548	0.3022

Table : X-Y coordinates of six points.

	p1	p2	p3	p4	p5	p6
p1	0.0000	0.2357	0.2218	0.3688	0.3421	0.2347
p2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
p3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
p4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
p5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
p6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points

2) Use CC_GENERAL.csv given in the folder and apply:
   a) Preprocess the data by removing the categorical column and filling the missing values.
   b) Apply StandardScaler() and normalize() functions to scale and normalize raw input data.
   c) Use PCA with K=2 to reduce the input dimensions to two features.
   d) Apply Agglomerative Clustering with k=2,3,4 and 5 on reduced features and visualize result for each k value using scatter plot.
   e) Evaluate different variations using Silhouette Scores and Visualize results with a bar chart.

Submission guideline:

- Make a proper pdf report containing all steps and results for Q1 and screenshots for Q2 with brief description of solution and results.
- Provide brief description with each screenshot to elaborate the output results.
- Push your solutions and code to github and provide link in your report.
- Also make 1-3 mins video demonstrating your solution and results (Exemption of video submission if demonstrated in class).

- **Note:** *Do not just provide link, otherwise would be marked 0. Cheating, plagiarism, disruptive behavior, and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy.*