HIERARCHICAL CLUSTERING ASSIGNMENT

Problem Statement

This assignment aims to utilize Hierarchical Clustering algorithms to analyze given datasets, extracting meaningful insights from their underlying structures. By exploring the datasets and employing Hierarchical Clustering techniques, students are expected to categorize the data into distinct clusters. The assignment emphasizes parameter tuning for optimal clustering results and requires interpretation of the clustering outcomes to derive valuable insights.

Guidelines

1. Foundational Knowledge
   - Understand the principles of clustering and how data can be segmented into clusters.
   - Familiarize yourself with Hierarchical Clustering algorithms and comprehend their principles and advantages.
   - Recognize different linkage methods (e.g., Ward, complete, average) and their impact on Hierarchical Clustering.

2. Data Exploration
   - Analyze the dataset's structure and characteristics using various exploratory techniques such as scatter plots, boxplots, heatmaps, etc.
   - Gain insights into the dataset's attributes to guide the clustering process.

3. Preprocessing and Parameter Selection
   - Standardize features if required, as Hierarchical Clustering can be sensitive to feature scales.
   - Select appropriate linkage methods and distance metrics based on data exploration.

4. Hierarchical Clustering
   - Implement Hierarchical Clustering algorithms using chosen parameters and methods.
   - Adjust linkage methods and other parameters to optimize clustering efficiency.

5. Cluster Analysis
   - Analyze resulting clusters to understand their attributes and characteristics.
   - Evaluate unclustered data points to derive conclusions about the dataset.
   - Compare findings with initial exploratory analysis to reinforce insights.

Step-by-Step Approach to Hierarchical Clustering

1. Setup and Data Preparation
   - Import necessary libraries: pandas, matplotlib, and Scikit-Learn.
   - Load the dataset for clustering.
   - Preprocess the data, ensuring standardized features if necessary.

2. Hierarchical Clustering Parameters
   - Choose appropriate linkage methods (e.g., Ward, complete, average).
   - Define distance metrics suitable for the dataset.

3. Performing Hierarchical Clustering
   - Initialize the Hierarchical Clustering model with selected parameters.
   - Apply the model on the prepared data.

4. Result Analysis
   - Examine cluster labels and interpret the clusters formed.
   - Visualize clusters to differentiate them using appropriate markers/colors.

5. Evaluation and Iteration
   - Compare Hierarchical Clustering results to any available actual clusters.
   - Adjust parameters (linkage methods, distance metrics) if clustering results are unsatisfactory.

6. Interpretation and Conclusion
   - Understand cluster patterns and distinctions.
   - Decide on handling noise or outliers, if any.

Links to Datasets for the Assignment

- Customer Clustering
https://www.kaggle.com/datasets/dev0914sharma/customer-clustering?select=segmentation+data.csv
- Credit Card Dataset for Clustering
https://www.kaggle.com/datasets/arjunbhasin2013/ccdata/data
- Mall Customer Segmentation Data
https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python