Systems and Biomedical Engineering
SBE 4032 - Advanced Topics in Medical Informatics (2)
Spring 2024

Assignment 4

For this assignment, you're required to implement from scratch the **Agglomerative Hierarchical Clustering**.

1- Refer to the steps described in the lecture and the tutorial to implement a **function** that performs **Agglomerative Hierarchical Clustering.**
   Your function should take as input the data, number of clusters and linkage method and return the clusters.
   - You should implement at least 2 linkage methods: Single-Link and Ward's Distance.
2- Load the Iris dataset. For visualization purposes, use only the following two features as your data for this assignment: "sepal length" and "petal length".
3- You're required to apply your function to the described data using both **Single-Link and Ward's Distance** methods. For each method, plot the clusters produced by your function for three numbers of clusters: 3 clusters, 4 clusters and 5 clusters. (i.e. 3 plots for each linkage method, each is a scatter plot of the data with color mapping for the produced clusters).
4- Compare the plots of both linkage methods in terms of sensitivity to outliers and the shape of formed clusters.

Instructions

1- The assignment is **individual**. Sharing code is not accepted. Cheating will result in a zero grade.
2- Add comments that describe your code.
3- Your conclusions and comments on the results can be included in markdown cells within your code.

4- **Late submissions** will be penalized.

5- **Submission** will be through providing me with a link to your code on Google Colab.
   The link should be submitted and turned in on Google classroom.
   Please remember to provide me with access to the code as an Editor.

6- Editing your code after the submission deadline is not accepted.