SSE 2024 Data Analysis and Mining

Assignment 3: Hierarchical Clustering Algorithm

Assignment Requirements

We have introduced the hierarchical clustering in the class. The hierarchical clustering can be categorized into agglomerative and divisive. Please choose one kind to implement by your own, especially pay attention to the time complexity of your implementation.

Please test your implementation on the datasets from UC Irvine Machine Learning Repository(https://archive.ics.uci.edu/) and document a report. Experimental settings and running times should be reported. As we have mentioned in the class, the distance calculation of two clusters is a key part of the algorithm. You can also discuss about it through the experiments.

The evaluation will depend on the implementation (e.g., efficiency), the report and your opinion/analysis/observation from the experiments.

Submission Requirements

- **Submission Format:** Submit the assignment as a ZIP file named *studentID_name_hw3.zip*, where the name should be in Chinese. The file should include the following contents.
- **Code**: Submit a code file containing the complete implementation of the hierarchical clustering algorithm, and accompanied by a README file outlining your thought process and execution steps.
- **Report Document**: Write a report with PDF format describing your algorithm design, especially how you consider and deal with time complexity and space complexity. The report should also include a dendrogram of the algorithm's results and a brief analysis of the results and two complexities.
- **Submission Date:** 2024/04/07, 23:59:59.
- **Note:** Please submit to Canvas.