

SSE 2024 Data Analysis and Mining

Assignment 4: Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Assignment Requirements

In this assignment, you are required to implement the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm by yourself. DBSCAN is a density-based clustering algorithm that can identify clusters of varying shapes in the presence of noise, by grouping together closely packed points and marking low-density regions as outliers. This algorithm primarily relies on two parameters: *eps* (neighborhood radius) and *minPts* (minimum number of points).

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

Key Steps

1. Understand and implement the DBSCAN algorithm by your own.
2. Choose an appropriate dataset for testing.
3. Adjust parameters (*eps* and *minPts*) and analyze how different settings affect the clustering outcome.
4. Submit a report that includes a detailed description of the algorithm implementation, the rationale behind the chosen dataset, results of parameter tuning, and a visualization of the clustering outcomes.

Submission Requirements

- **Submission Format:** Submit the assignment as a ZIP file named *studentID_name_hw4.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 the results of parameter tuning, and a visualization of the clustering outcomes.
- **Submission Date:** 2024/05/08, 23:59:59.
- **Note:** Please submit to [Canvas](#).