2025 Data Analysis and Mining - Homework 03

Several questions are given below and you are asked to complete the questions according to the corresponding requirements.

Q1: Clustering Algorithm Implementation and Analysis

Objective:

Please implement the following algorithms using Python libraries (e.g., scikit-learn, NumPy, or SciPy) on the provided datasets, and conduct a comparative analysis of their time complexities: 1) Hierarchical Clustering, 2) Spectral Clustering, 3) K-Means, 4) Expectation-Maximization (EM), 5) DBSCAN.

Provided Datasets:

a） Iris

b） Wine Quality (white)

[Optional] Q2: Paper Reading

Objective:

Please read this paper *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*.

Submission Requirements

It is strictly required that the assignment be submitted as a ZIP file named *studentID_name_hw3.zip*. The file should include the following contents:

- **Code:** It is recommended that you implement your solution using python. Submit a code folder (create subfolders by question ID if applicable) containing the complete implementation of the tasks, and accompanied by a README file outlining your thought process and execution steps.
- **Report Document:** Write a report in PDF format, using an appropriate length to describe how you completed the task.
- **Submission Deadline:** 2025/04/07, 23:59:59.

Note:

- The report can be written in Chinese or English.
- Please submit your file to Canvas.
- Please email late submissions to 1461715017@qq.com.

If you have any questions, please contact the teaching assistant, Shijie Luo (QQ: 1461715017).