Report Clustering Methods

Homework 1

NYCU

This report is present a comprehensive analysis of clustering algorithms applied to a variety of open-source datasets. The objective is to evaluate the performance and applicability of different clustering techniques on datasets that exhibit a range of complexities and structural characteristics. This evaluation is conducted using nine distinct datasets: Complex9, Curves2, Diamond9, Disk-5000n, Gausians1, Hypercube, Impossible, Disk-4000n, and Mopsi-Joensuu, and Sizes5 sourced from the Clustering Benchmark database.

The clustering methods that evaluate selected for this study include Mean Shift, DBSCAN, HDBSCAN, Agglomerative, K-means and OPTICS. These methods were chosen based on methods that we have studied, distinct methodological approaches and their potential for revealing intricate patterns in complex datasets. Each method will be thoroughly tested across all datasets, with a focus on tuning hyper-parameters to optimize performance on clustering pattern.

In the experiment, the dataset was displayed on a 2D graph as shown in Figure 1, arranged in the following order: compex9, curves2, diamond9, disk5000n, gaussians1, hypercube, impossible, disk-4000n, mopsi-joensuu, and sized5. Each dataset exhibits its own unique pattern and presents specific extraction challenges. Consequently, a method that is suitable for one dataset might not necessarily be effective for another. Additionally, the performance of clustering depends on the precise tuning of each hyper-parameter.

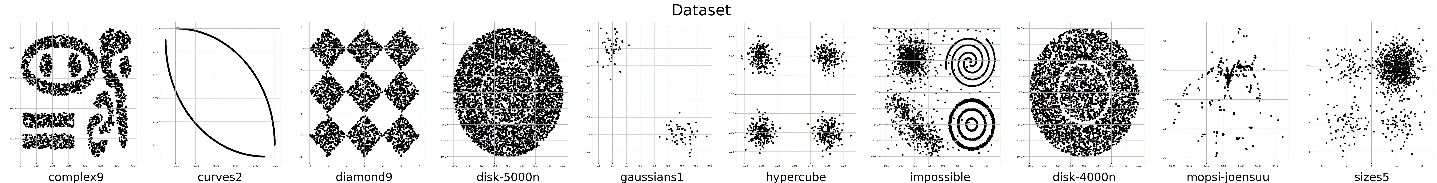


Figure Dataset Collections

After the testing all methods and tuning the hyper-parameter, It’s got the results of the clustering that is shown on Figure 2. On this results, it shows that each result has their own special pattern and different from each other. So this shows that each method will be useful according to the clustering problem to be solved.

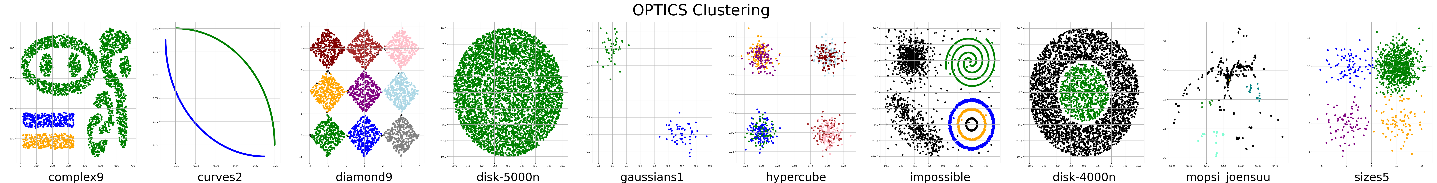
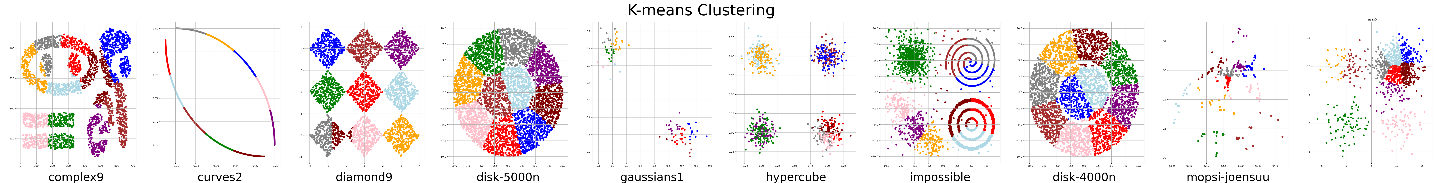
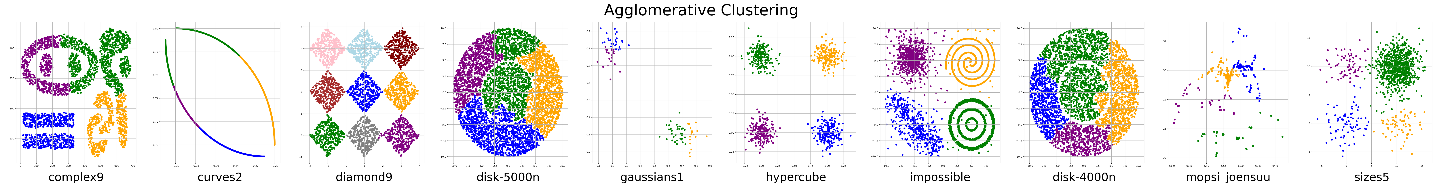
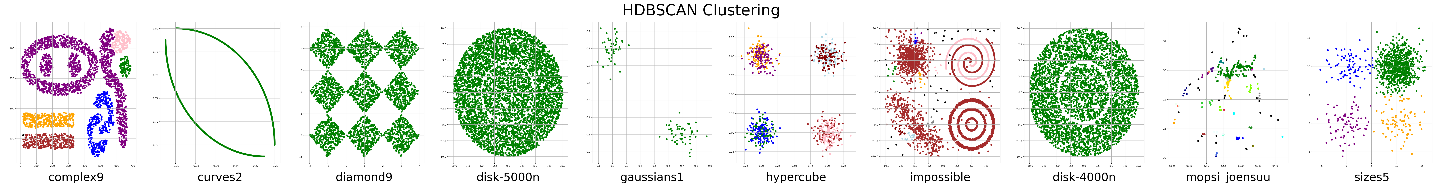
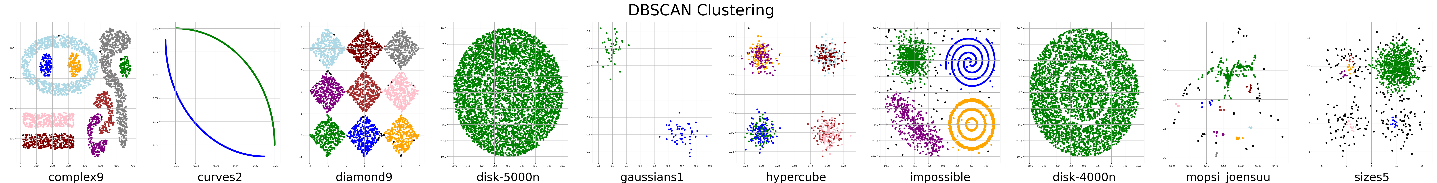
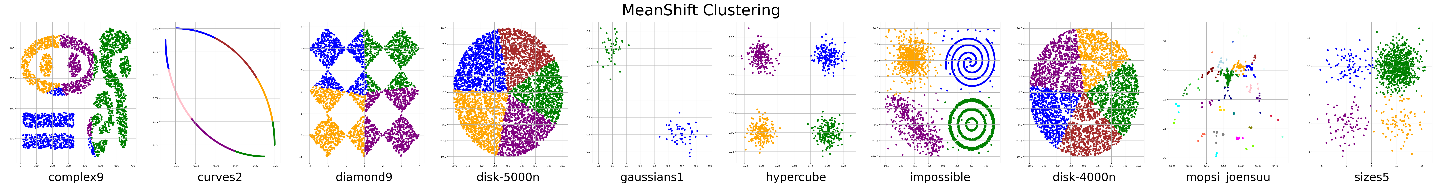


Figure Results of Clustering Methods

MeanShift Clustering

DBSCAN Clustering

HDBSCAN Clustering

Agglomerative Clustering

K-means Clustering

OPTICS Clustering