Report Analysis HOMEWORK ASSIGNMENT 6

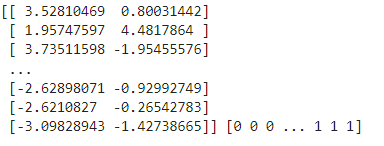
Requirement for the report: - results for each step, including a comparison of the several models - description of the clustering method - description of the quality metrics for the models.

1. We generate synthetic data for analysis and making a model.

Input data

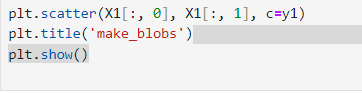


Output

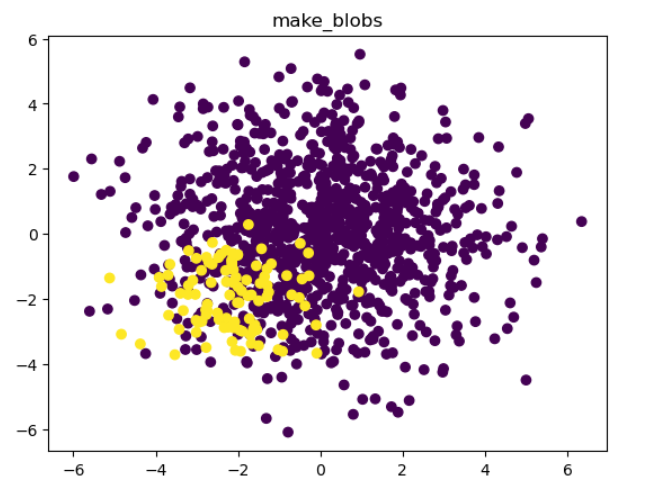


1. visualize the input data on the Graph.

Input data.

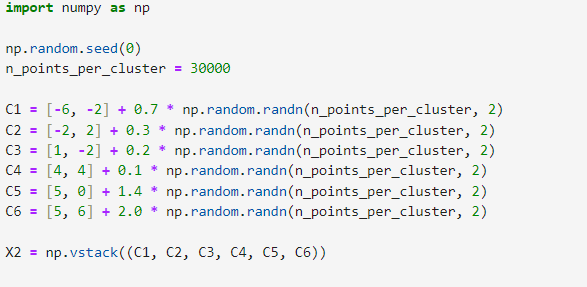


Output.



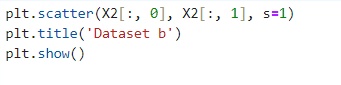
1. For dataset b we will perform as it is step.

Input data

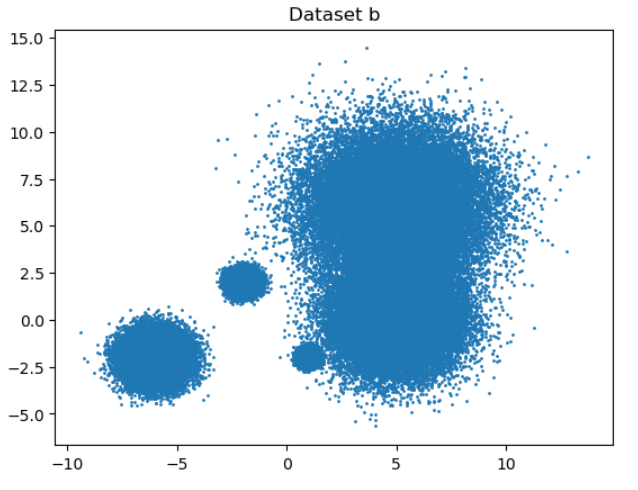


1. Visualize the dataset b

Input data.

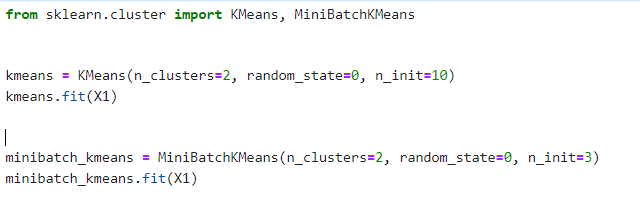


Output



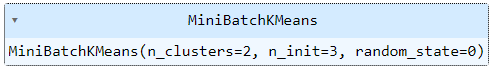
1. Create a clustering model for K-means using: cluster.KMeans and cluster.MiniBatchKMeans

Input data





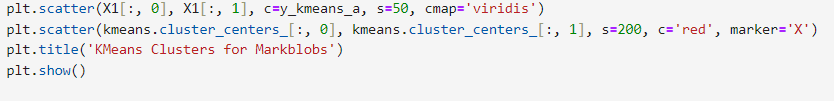
Output



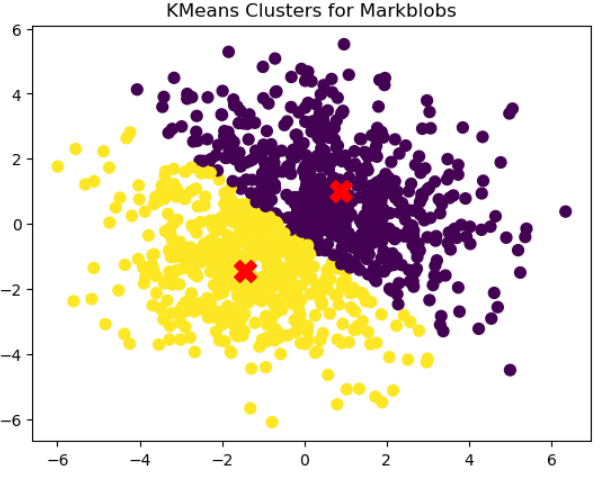


1. Visualise through Graph.

Input data.

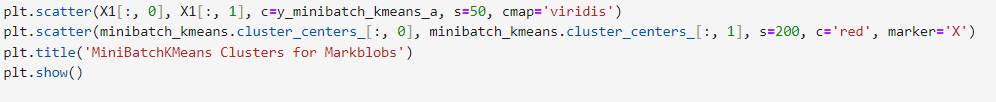


Output

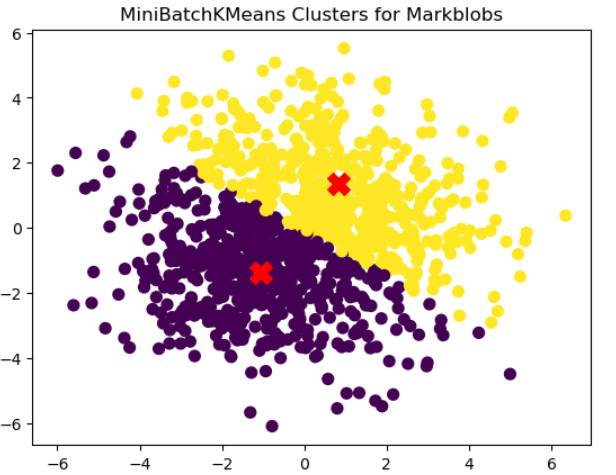


1. Using MiniBatchKMeans

Input Data

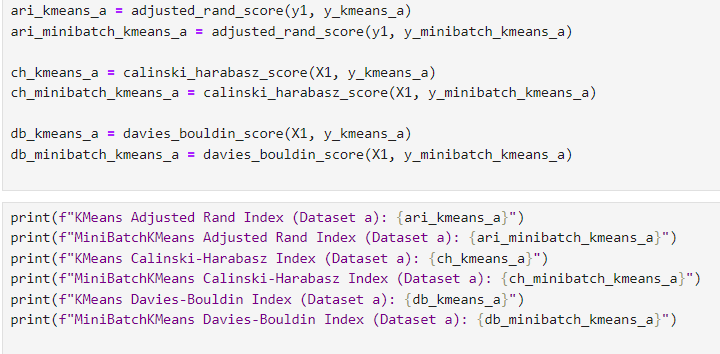


Output

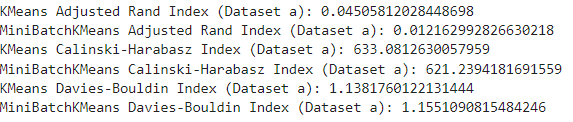


1. For each alternative model use quality metric from sklearn.metrics: Adjusted Rand Index, Calinski-Harabasz Index, Davies-Bouldin index.

Input data

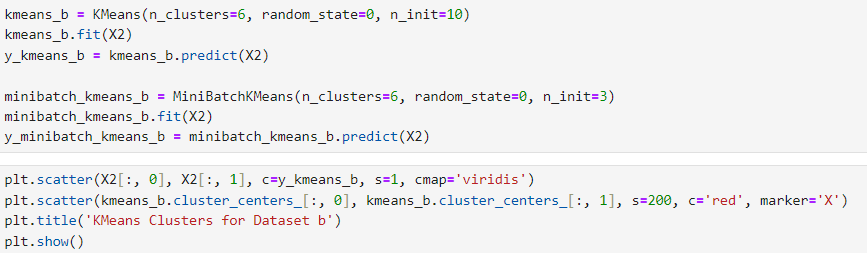


Output.

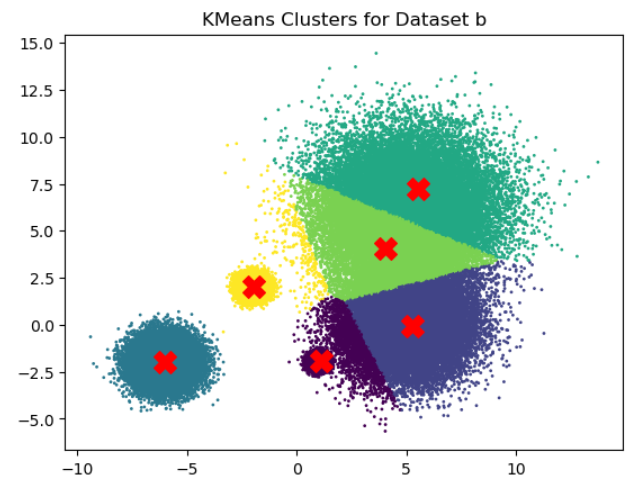


For dataset b we will perform the same step.

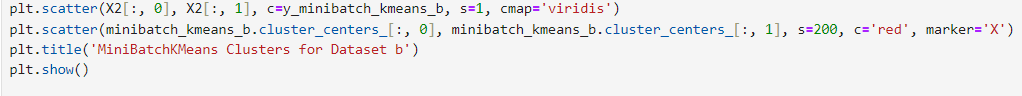
Input Data.



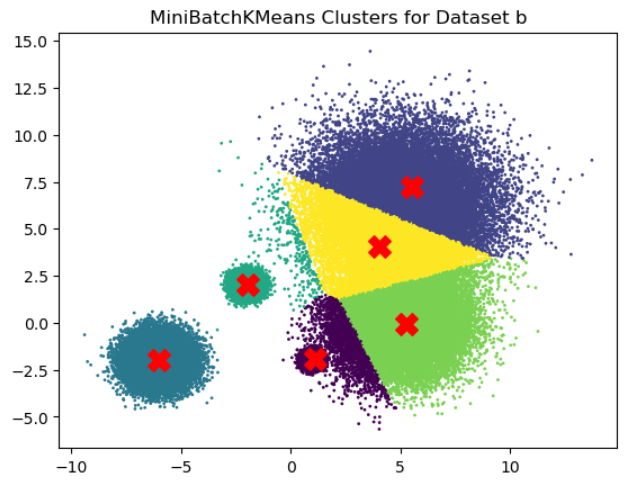
Output.



Input data.

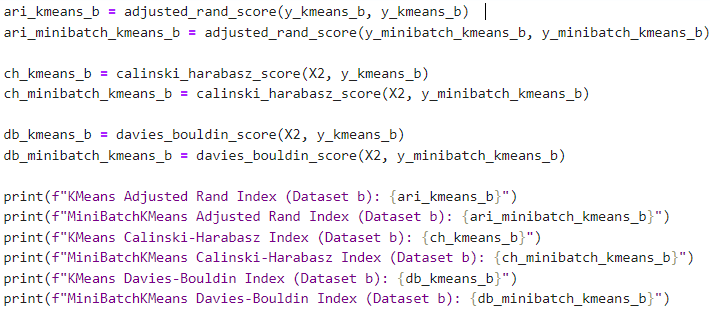


Output

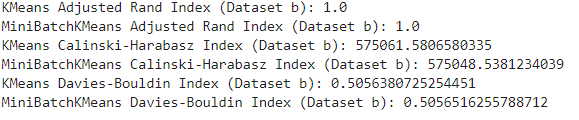


Metrics Analysis.

Input Data.



Output



Description of quality metrics and comparision of the model and analysis.

Dataset a

|  |  |  |  |
| --- | --- | --- | --- |
| Model | ARI | CH | DB |
| K Means | 0.04 | 633.0 | 1.13 |
| MiniBatch KMeans | 0.012 | 621.0 | 1.155 |

**KMeans ARI = 0.04**: Indicates a low level of similarity between the true labels and the predicted labels.

**MiniBatchKMeans ARI = 0.012**: Even lower similarity between true and predicted labels compared to KMeans.

Calinski-Harabasz Index (CH)

**KMeans CH = 633.0**: Higher value indicates well-defined clusters.

**MiniBatchKMeans CH = 621.0**: Slightly lower than KMeans, indicating slightly less well-defined clusters.

**Davies-Bouldin Index (DB)**

KMeans has a slightly lower DB score, suggesting better separation between clusters compared to MiniBatchKMeans.

KMeans slightly outperforms MiniBatchKMeans in terms of ARI, CH, and DB scores, indicating better clustering quality for Dataset a.

Dataset b

|  |  |  |  |
| --- | --- | --- | --- |
| Model | ARI | CH | DB |
| K Means | 1.0 | 575061.58065 | 0.50563 |
| MiniBatchKMeans | 1.0 | 575048.5381 | 0.5065 |

**KMeans vs. MiniBatchKMeans**: Both models perform almost identically, with KMeans slightly outperforming MiniBatchKMeans in terms of CH **Calinski-Harabasz Index (CH)and DB scores**.KMeans slightly outperform than MiniBatchKMeans.

**Adjusted Rand Index (ARI)** perfact in both the cases.

**MiniBatchKMeans DB = 0.5065**: Slightly higher than KMeans, but still low, indicating good separation between clusters.

**Overall for both the dataset KMeans outperform than the MiniBatchKMeans also this indicates for smaller dataset KMeans is better**.