**Clustering Techniques used:**

1. **K-Means:** It uses K-different points in the data that are randomly initialized and called “Centroids.” Then it assigns every data point to the nearest centroid and adjusts the centroid to the average of all of the points assigned to it iteratively.
2. **Fuzzy Clustering:** It is a soft clustering approach, where each data point is assigned a likelihood or probability score to belong to that cluster. A number of clusters c and degree of fuzziness m is selected, and Partition Matrix is initialized. Then the centroids are calculated, and the partition matrix is updated.
3. **BIRCH:** Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) is a clustering algorithm that generates a small and compact summary of the large dataset that retains as much information as possible. This smaller summary is then clustered instead of clustering the larger dataset.
4. **Gaussian Mixed Model Clustering:** It uses variance to determine the width of the bell-shaped Gaussian curve and hence the distribution of the data in the clusters. It’s a soft clustering technique.

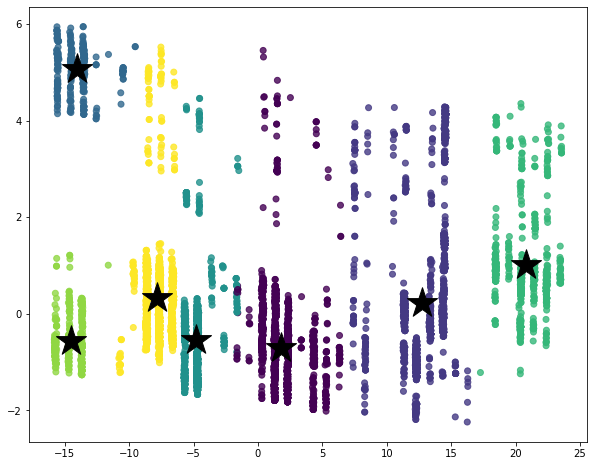
**(1,2) The centroid of each cluster for every model and Visualization of the clusters.**

Number of points used to represent the clusters: 1000

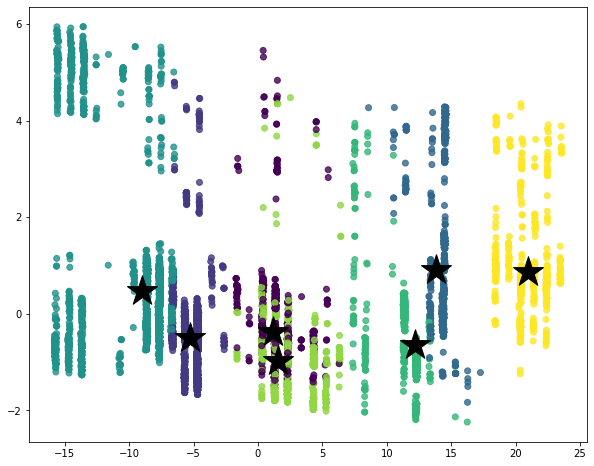
Different clusters are shown in different colors.

The bold black “\*” symbolises the centroids of the respective clusters.

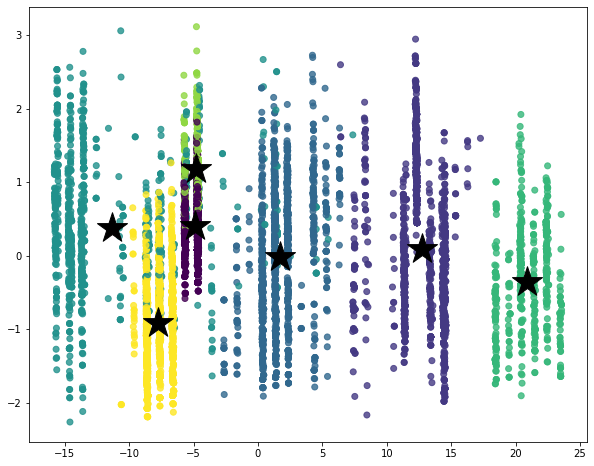
1. **Clusters and centroids using K-means Clustering.**



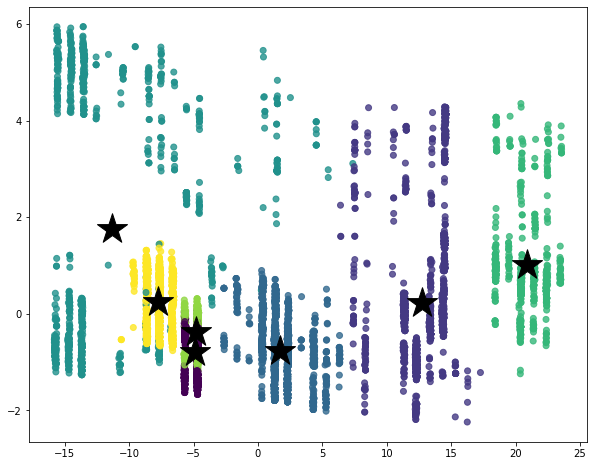
1. **Clusters and centroids using Fuzzy Clustering.**



1. **Clusters and centroids using BIRCH Clustering.**



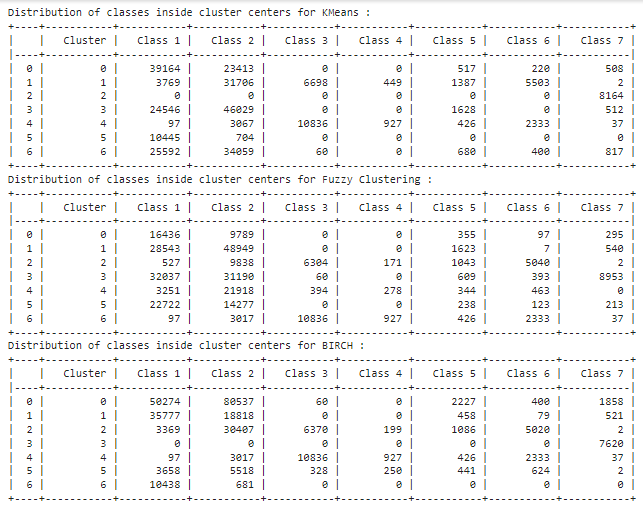
1. **Clusters and centroids using Gaussian Clustering.**

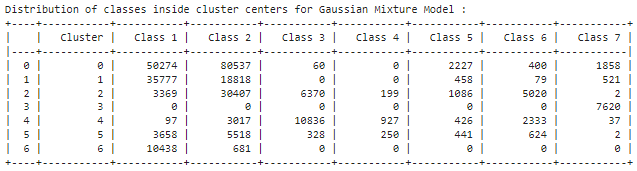


**3. Compare your cluster distribution with the true label count.**

In each technique, we have 7 clusters, having Ids- 0,1,2,3,4,5 and 6. In each cluster, we have data points that have original labels 1,2,3,4,5,6,7. So we calculate that in the cluster “c,” how many data points of class “i” belong, where 0 <= c <=6 and 1<= i <=7.

Output:





**4. Compare the cluster formation of the gaussian based method with the other three clustering**

**methods and report your observations on the results.**

1. With Gaussian Clusters, we say more pure clusters as compared to other clustering distributions, for example, Cluster 6 has 0 data points of class 3,4,5,6, and 7, Cluster 3 has 0 data points of classes 1,2,3,4,5 and 6, etc.
2. The F1-score of the Gaussian Mixture Method is lower. Gaussian Method is a probabilistic model and others are non-probabilistic models and hence Gaussian may not perform best for all the datasets. It depends on the variance and hence depends highly on the dataset.

**Used validation data to test the models, Balanced F1-Score:**

With KMeans, 0.6322

With Fuzzy C - Means, 0.5717

With Birch, 0.6342

With Gaussian Mixture Method, 0.4094

Inference.py is working perfectly to the best of our knowledge and returning the output labels as asked for.