Gaussian Mixture Models (GMM)

Pham Van Hong
11201638

Problem 3:

Pros:

- KMeans and GMM are both clustering models

Cons:

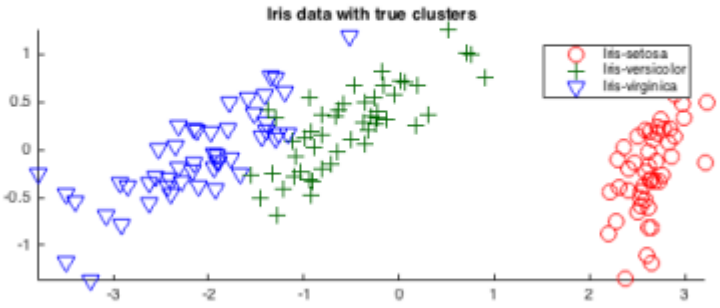
- KMeans group data points using distance form the cluster centroid, GMM uses probalilistic assigment of data point to clusters
- KMeans is prototype – based clustering while GMM generates density – based clusters.
- The performnce of GMM is better than that of KMeans.
- K-Means requires much less time to discover and group the workloads into required number of clusters than required by GMM for corresponding number of Gaussian components.

Problem 4:

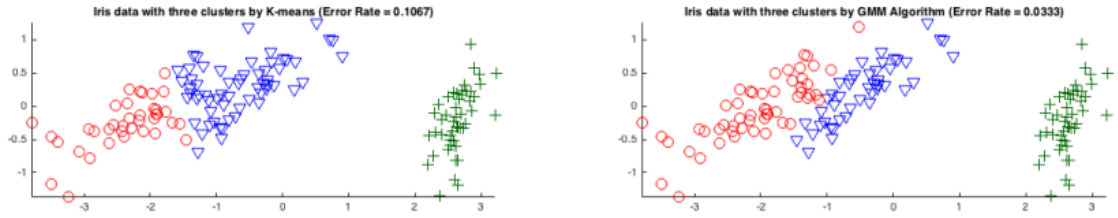
KMeans Vs. GMM

Example 1:

We have Iris Data with the True Labels



After apply GMM on the dataset:

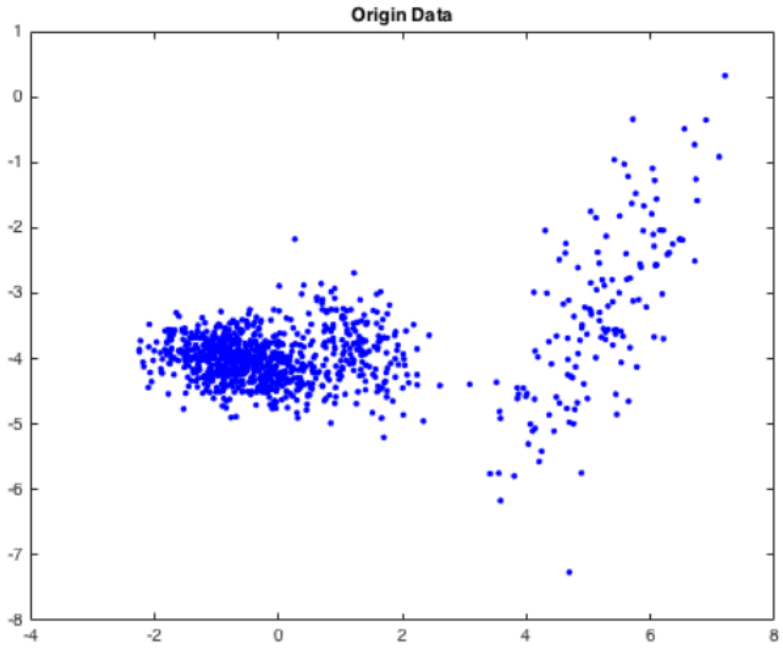


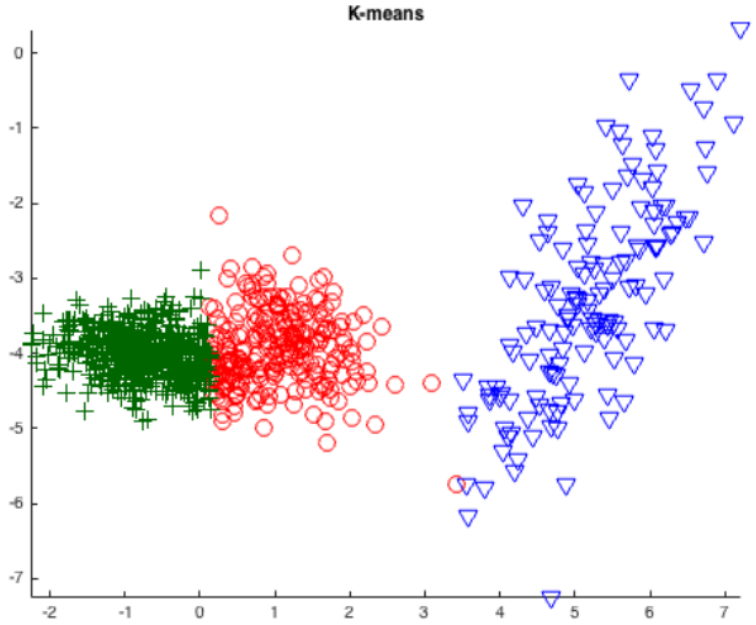
K-means vs. GMM on Iris Data

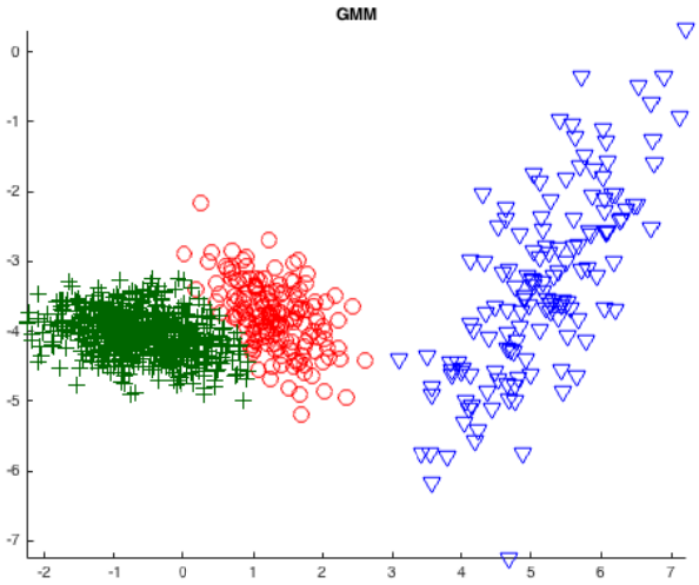
The performance of GMM is better thn that of K-mens. The three clusters in GMM plot are closer to the original ones. The error rate of GMM is smaller than K-means, so it better than.

Example 2:

K-means and GMM on a Toy Dataset


Origin Data
1
0
-1
-2
-3
-4
-5
-6
-7
-8
-4
-2
0
2
4
6
8


K-means
0
-1
-2
-3
-4
-5
-6
-7
-2
-1
0
1
2
3
4
5
6
7


GMM
0
-1
-2
-3
-4
-5
-6
-7
-2
-1
0
1
2
3
4
5
6
7