**CMPE 544: Pattern Recognition (Fall 2020)**

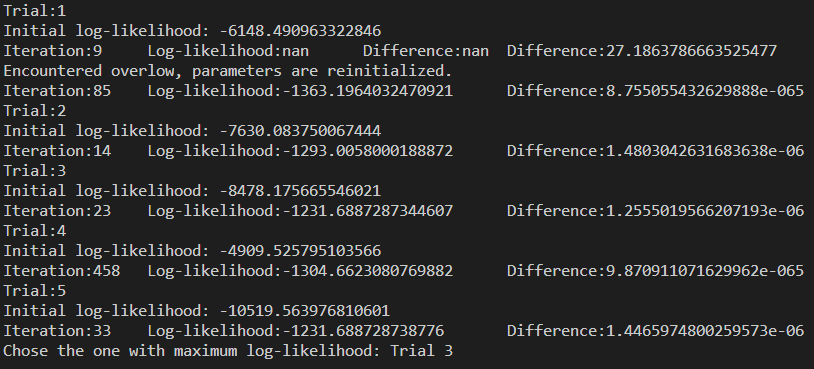
**Homework2 Report – Yaşar Alim Türkmen – 2019700123**

**Implementation Details**

The program has a main function called *run().* The function takes a number n as a trial number and runs EM algorithm for n times and takes the solution with the maximum log-likelihood as the optimal one. Based on that solution, it also assigns clusters to the points and plots them. To make the program more readable, a class called GM (Gaussian Mixture) is used. This object both stores the data and means and covariance matrices for Gaussian distributions. EM algorithm is also implemented under this class.

In each trial, *run()* function calls *EM()* function which follows all of EM steps one by one. First, it creates an instance of GM object which initializes random means (between minimum and maximum values of features) and covariance matrices and equal mixing coefficient values for each 3 clusters. After initialization, E-step is processed (*E\_step()*). Conditional probabilities of distributions are computed for the given data. These probabilities are stored in GM instance as *expectation\_matrix*. Then, in M-step, mixing coefficient values, means and covariance matrices of clusters are updated respectively(*update\_coeffs(), update\_means(), update\_covars()*). Finally, log-likelihood is calculated for iterated parameters in *is\_converged()* function. If the difference of two successive iterations’ log-likelihood values is less than the threshold, which is set to 5.10-5, EM algorithm stops and returns GM object and last log-likelihood value. Sometimes, it is possible to encounter overflow or singularity problem for covariance matrix. Therefore, when it occurs, the parameters of GM object are reinitialized.

Based on returned last log-likelihood values, optimal solution is chosen:

The maximum log-likelihood reached by the algorithm is **-1231.68883**.

The parameters estimated by EM algorithm and the plot are given below:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean | Covariance Matrix | Mixing Coefficient |
| Cluster1 (Red) |  |  | 0.333 |
| Cluster2 (Blue) |  |  | 0.333 |
| Cluster3 (Green) |  |  | 0.333 |

