Name: Aditya Borde

-----EM Algorithm-----

Following parameters are noted when EM converges.

**Parameters - EM for General GMM:**

**\*\***Parameters arenoted on the result having highest log-likelihood.

**For K =1**, Iterations = 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Initial Parameters | | Final Parameters | |
| Mean | Variance | Mean | Variance |
| 1 | 16.7037 | 69.0083 | 15.4816 | 67.5149 |

**For K =2**, Iterations = 42

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Initial Parameters | | Final Parameters | |
| Mean | Variance | Mean | Variance |
| 1 | 25.8608 | 175.2417 | 20.3207 | 27.9433 |
| 2 | 4.3160 | 192.1862 | 5.4869 | 0.9861 |

**For K =3,** Iterations = 323

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Initial Parameters | | Final Parameters | |
| Mean | Variance | Mean | Variance |
| 1 | 6.8958 | 141.231199 | 5.509279 | 1.030257 |
| 2 | 7.3815 | 133.127659 | 15.449160 | 0.967115 |
| 3 | 25.7610 | 173.180150 | 25.486654 | 0.998096 |

**Parameters - EM for known variance GMM:**

**\*\***Parameters arenoted on the result having highest log-likelihood.

**For K =1**, Iterations = 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Initial Parameters | | Final Parameters | |
| Mean | Variance | Mean | Variance |
| 1 | 14.9735 | 1.0 | 15.4816 | 1.0 |

**For K =2**, Iterations = 17

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Initial Parameters | | Final Parameters | |
| Mean | Variance | Mean | Variance |
| 1 | 5.8638 | 1.0 | 10.4628 | 1.0 |
| 2 | 25.6739 | 1.0 | 25.46063 | 1.0 |

**For K =3,** Iterations = 7

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Initial Parameters | | Final Parameters | |
| Mean | Variance | Mean | Variance |
| 1 | 26.61713 | 1.0 | 25.4866 | 1.0 |
| 2 | 14.86803 | 1.0 | 15.4491 | 1.0 |
| 3 | 14.46524 | 1.0 | 5.5092 | 1.0 |

**Heuristics for Initialization:**

For initialization of EM parameters, two approaches are used and results are observed.

1. Mean = Selecting K random data points as mean for K different clusters.

Alpha = K different random values summing to 1.

Then, doing EM steps.

1. Set Wik such that and where is sum of weights of points in a given component.

*All the above readings are noted using Initialization Strategy 1). But, results are observed on both the strategies.*

**Performance Dependency on Heuristic Selection:**

For both the initialization heuristics, convergence parameters are same. But, if we use strategy 2) for initialization, number of iterations required are less in comparison with the strategy 1) for convergence. Therefore, Strategy 2) is more efficient with respect to performance, also it doesn’t affect the final parameters that we get at the time of convergence.

**Initialization Strategy:**

EM for General and EM for known variance GMMs are implemented in the code using Heuristic 1) and 2) as mentioned above. All results are noted in the report for Heuristic 1) but observed for both the heuristics.

**Better Approach:**

EM for General GMM will be better approach than the EM for known variance GMM.

***Explanation:*** In EM for known variance, we are keeping the variance same throughout the algorithm. Which ensures the distribution of the values from the mean is same at every step. For equally distributed values from mean, algorithm converges showing unpredictable behavior as maximization step is keeping its variance constant resulting all values are normally distributed over the range. Therefore, Even though the data distribution is not equally spread from mean and we are enforcing this by keeping variance =1, convergence do not assure consistency of parameters in GMM showing improper clustering of data points.

Whereas, in case of EM for general GMM, the fact that value distribution is calculated at each step and precisely data is not distributed normally from the mean in most of the cases. Algorithm takes proper steps to converge over all of its values. Hence, convergence gives consistent parameters making data points into proper clustering.

**Conclusions:**

1) Depending on the initialization heuristics, convergence time changes. So, choosing good initialization heuristic will improve the performance. Moreover, it won’t affect the result at the time of convergence for EM for general GMM.

2) Initialization heuristic won’t change the convergence parameters.

3) EM for known variance shows unpredictable nature in the convergence time, as our assumption that all values are known to be distributed equally from the mean at each step, but the General GMM which doesn’t assume any variance in the any stage clusters data properly for the GMM. Clustering using known variance will have bottle neck of inconsistent clustering of data points.

4) Sometimes EM for Known Variance GMM fail to converge or takes so many iterations to converge. Clustering would be difficult at those situations.

5) EM for known variance strategy can show same results of final parameters as that of EM for general GMM. As EM for known variance GMM is staying at the same maxima that EM for general GMM stays.