Pramod Hegde                                                    pxh170330@utdalls.edu

## ASSIGNMENT 3
## EM ALGORITHM

**Initial setting:**

Initial means and variance were selected randomly.
1) Initial Mean: [[25.87791226 24.19551948 4.74058781]]
2) Initial Variance: [[7.28042205 7.28042205 7.28042205]]

It took 9 iterations to converge.
Number of iterations: 9
With the highest log likelihood of -9584.40855420643
Log likelihood list: [-24422.58895704028, -19025.234151545843,
-18082.85128263917, -14810.766252835705, -11542.57288518026,
-9742.031291395982, -9584.480954409542, -9584.408590459887,
-9584.40855420643]

**Initialization strategy used:**
Initial K Gaussian means by randomly selecting K initial data points and selecting
the initial K variances randomly which is multiple of overall data variance.
Table 1.1 Parameters for different initialization(K=3)

| SN | Initial Mean | Initial Variance | Final Mean | Final Variance | Loglikelihood | Iterations |
|---|---|---|---|---|---|---|
| 1 | [24.73527387, 15.40178278, 15.05798473] | [2.69092665, 2.69092665, 2.69092665] | [25.49940413, 15.45688924, 5.51203542] | [0.99875855, 0.96765949, 1.03078067] | -9584.40855 | 15 |
| 2 | [4.36735471, 26.10750806, 23.91932875] | [2.99862498, 2.99862498, 2.99862498] | [ 5.51203542, 25.49940413, 15.45688924] | [1.03078067, 0.99875855, 0.96765949] | -9584.40855 | 7 |
| 3 | [ 8.05613952, 5.03300826, 23.89400827] | [8.10780226, 8.10780226, 8.10780226] | [15.45688924, 5.51203542, 25.49940413] | [0.96765949, 1.03078067, 0.99875855] | -9584.40855 | 9 |

Table 1.1 Parameters for different initialization(K=3)

| SN | Initial mean | Initial variance | Final mean | Final variance | loglikelihood | iterations |
|----|--------------|------------------|------------|----------------|---------------|------------|
| 1 | [26.01410786 16.6350372 14.22529361 26.40312823] | [4.74517866 4.74517866 4.74517866 4.74517866] | [25.46542571 15.45688924 5.51203542 8.5773826] | [0.93055071 0.96765949 1.03078067 1.38451662] | -9504.3932 | 80 |
| 2 | [25.80282293 4.73193109 24.91775332 24.57384755] | [6.03090481 6.03090481 6.03090481 6.03090481] | [2.54994041 5.51203542 9.64310789 1.54568893] | [9.98758273 1.03078058 9.30055251 9.67659257] | -9563.5572 | 26 |
| 3 | [26.31454591 6.54516764 27.09746192 5.01390968] | [9.54985087 9.54985087 9.54985087 9.54985087] | [16.02286035 15.06276693 25.49940413 5.51203542] | [0.67059413 0.80140099 0.99875855 1.03078068] | -9503.6618 | 53 |

Table 1.2 Parameters for different initialization(K=4)

**Observations:**

The EM algorithm is run on different initialization parameters. For less number of clusters (K=3 Table 1.1) algorithm converges with less number of iteration as compared to large number of clusters (K=4 Table 1.2).

**Next Step (Variance = 1):**

Initial variance is [1. 1. 1.]. It remains constant as final variance is [1. 1. 1.]. The mean values are as following.

| SN | Initial mean | Final mean | Loglikelihood | Iteration |
|----|--------------|------------|---------------|-----------|
| 1 | [6.67052548 6.23213941 6.25556869] | [25.49940413 5.51203541 15.45688923] | -9587.3653 | 26 |
| 2 | [2.78281581 5.93774529 26.08622784] | [ 5.51203541 15.45688923 25.49940413] | -9587.3738 | 6 |
| 3 | [15.68066654 25.87349446 6.47415114] | [15.45688923 25.49940413 5.51203541] | -9587.3738 | 3 |

Table 1.3 Parameter for different initialization(K=3)

**Which is better?**
Consider SN 1 from table 1.3
It took 26 iterations to converge with the initial variance of [1. 1. 1.]

For the same mean value: [6.67052548 6.23213941 6.25556869]. considering random variance value: [7.67788792 7.67788792 7.67788792]. It takes 39 iterations to converge. (following output)

(Note: To get this result you can uncomment the code and comment the random value generator line)

Initial Mean: [[6.67052548 6.23213941 6.25556869]]
Initial Variance: [[7.67788792 7.67788792 7.67788792]]
Final Mean: [[25.49940413 5.51203542 15.45688924]]
Final Variance 0: [[0.99875855 1.03078067 0.96765949]]
Number of iterations: 39

So, Initializing variance to 1. And updating only means in M-step converges faster because difference between initial value and true value is high (initial - true).