Mohammadsadegh Abolhasani - u1416052
HW3 CS6190 - Spring 23

Practice

1. All images have been attached and can be viewed in the following pages.

   (a) The normalization constant is reported to be 1.1688. You can see the density below:

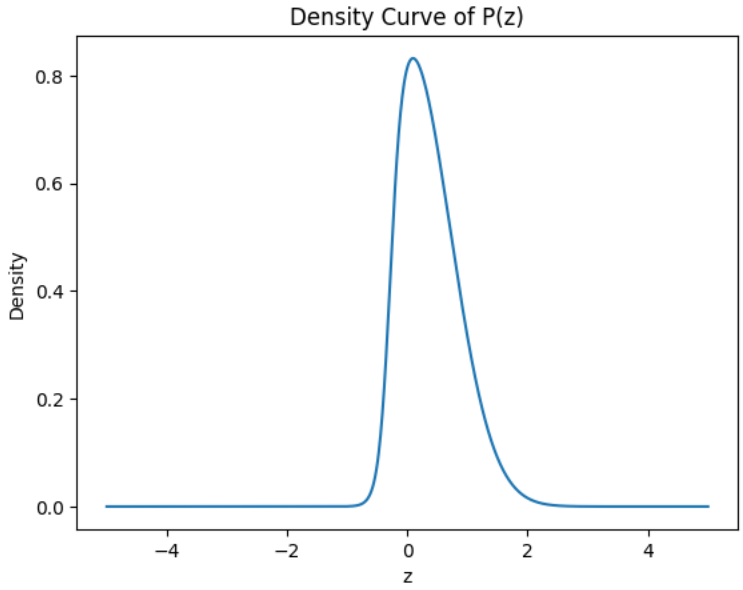


Figure 1: density curve of $p(z)$

   (b) Reported mean is 0.0947 and reported variance is 0.2591.

   (c) Reported mean is 0.0805 and reported variance is 0.2231.

   (d) All images have been attached and can be viewed in the following pages. We can see that the graph of density of LVI is more close to the ground truth than the Laplace approximation. One of the reasons can be the constant upates that will lead LVI to achive better and optimized answers.

2.

   (a) The reported model accuracy on test data set is 0.956 and the reported average predictive likelihood is 0.932.

   (b) The reported model accuracy on test data set is 0.956 and the reported average predictive likelihood is 0.920.

   (c) The reported model accuracy on test data set is 0.956 and the reported average predictive likelihood is 0.933.

   (d) The reported model accuracy on test data set is 0.838 and the reported average predictive likelihood is 0.830.

   (e) We can see that the accuracy and average predictive likelihood were virtually the same on off all the models. Although, the LVI without fully factorised posterior has the best average predictive

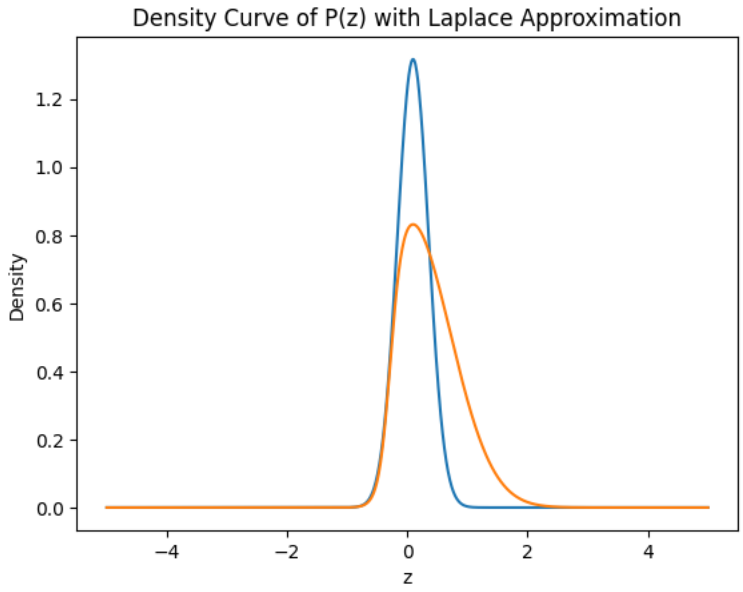


Figure 2: density curve of $p(z)$ with Laplace approximation in blue

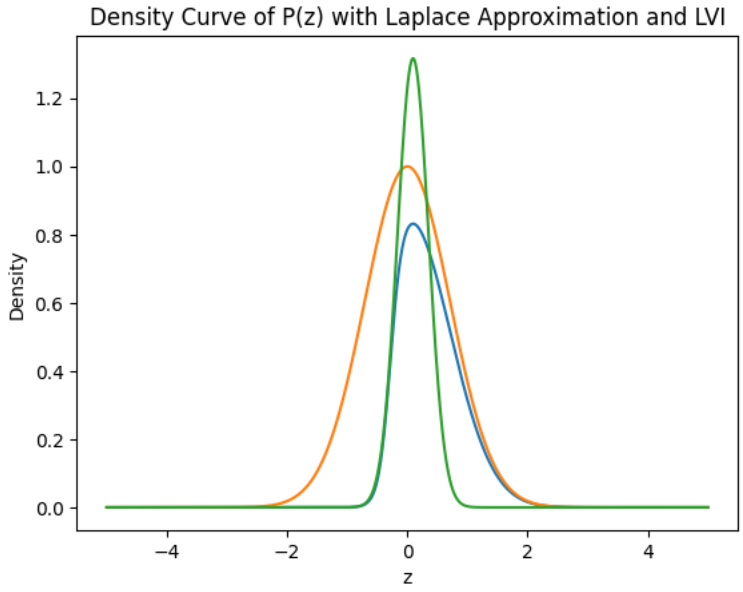


Figure 3: density curve of $p(z)$ with LVI in orange

```
m_n :  [-2.45714652 -1.44747126 -1.56431965 -0.6961338 ]
S_n :  [[0.06076195 0.02992161 0.03792227 0.00866902]
 [0.02992161 0.03182664 0.03120047 0.01631676]
 [0.03792227 0.03120047 0.03464284 0.0131994 ]
 [0.00866902 0.01631676 0.0131994  0.02080629]]
```

Figure 4: Posterior distribution for standard Laplace approximation part

```
m_n :  [-2.45714652 -1.44747126 -1.56431965 -0.6961338 ]
S_n :  [[0.06076195 0.         0.         0.        ]
 [0.         0.03182664 0.         0.        ]
 [0.         0.         0.03464284 0.        ]
 [0.         0.         0.         0.02080629]]
```

Figure 5: Posterior distribution for standard Laplace approximation part (diagonal Hessian)

```
m_n :  [[-2.49081204]
 [-1.47123318]
 [-1.5889216 ]
 [-0.70870699]]
S_n :  [[0.06076195 0.         0.         0.        ]
 [0.         0.03182664 0.         0.        ]
 [0.         0.         0.03464284 0.        ]
 [0.         0.         0.         0.02080629]]
```

Figure 6: Posterior distribution for Implementing variational logistic regression

```
m_n :  [[-2.49081204]
 [-1.47123318]
 [-1.5889216 ]
 [-0.70870699]]
S_n :  [[0.06076195 0.         0.         0.        ]
 [0.         0.03182664 0.         0.        ]
 [0.         0.         0.03464284 0.        ]
 [0.         0.         0.         0.02080629]]
```

Figure 7: Posterior distribution for Implementing standard mean-field update

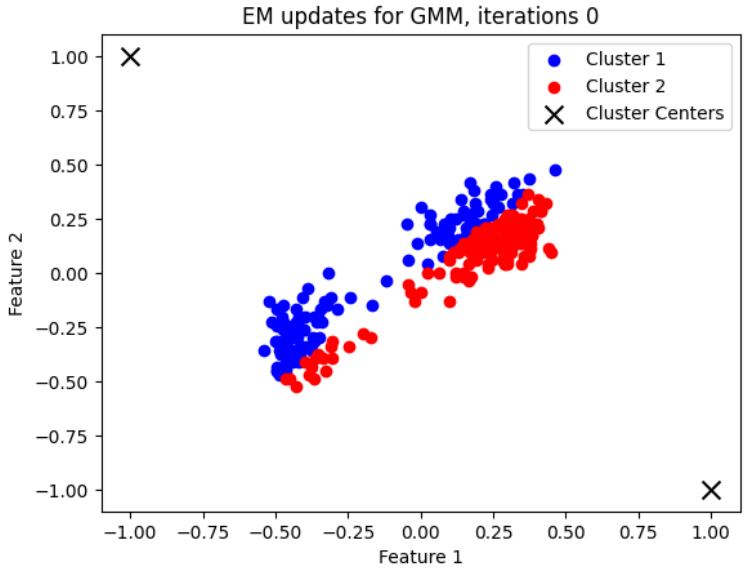
likelihood. For the factorized posteriors, I expected the accuracy to be the same as others, and I believe my code still has problems, but I think it should have gotten a better result compared to all of them, when our only concern is the accuracy and average predictive likelihood.

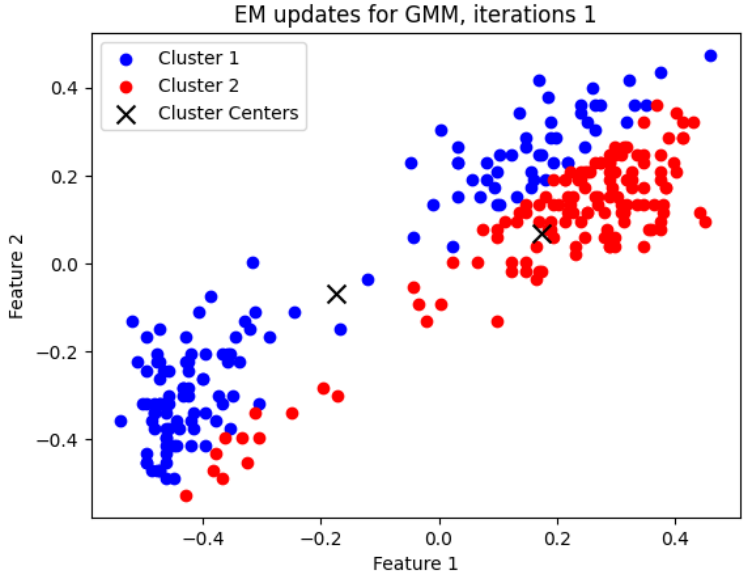
3. All images have been attached and can be viewed in the following pages.

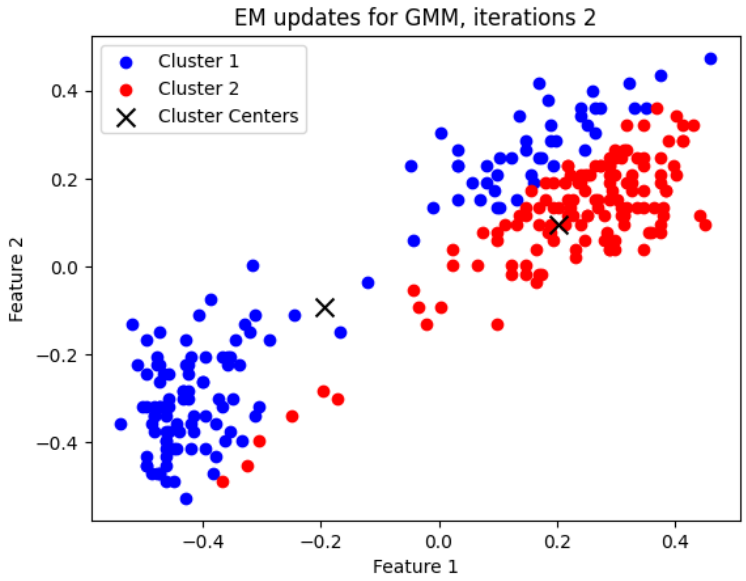
   (a) Below you can see the 1, 2, 5, 100 iterations for the data. In step 0, the data and cluster centers are shown before any change and transformation.

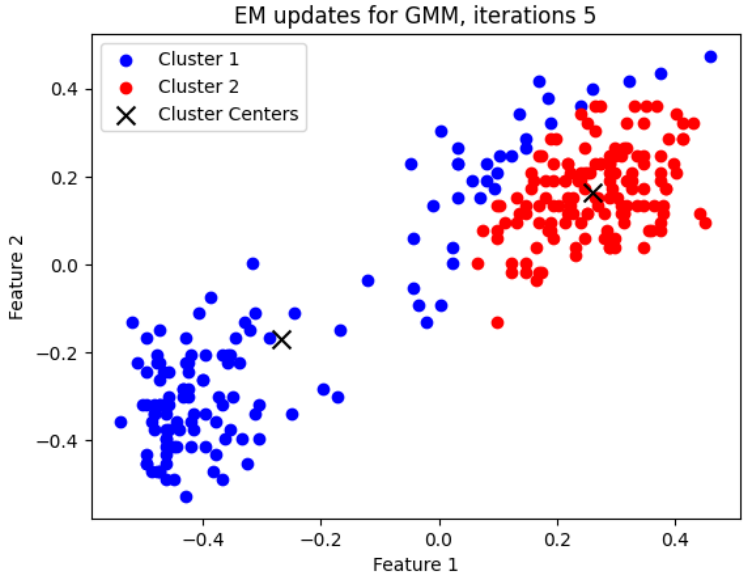
   (b) Below you can see the 1, 2, 5, 100 iterations for the data. In step 0, the data and cluster centers are shown before any change and transformation.

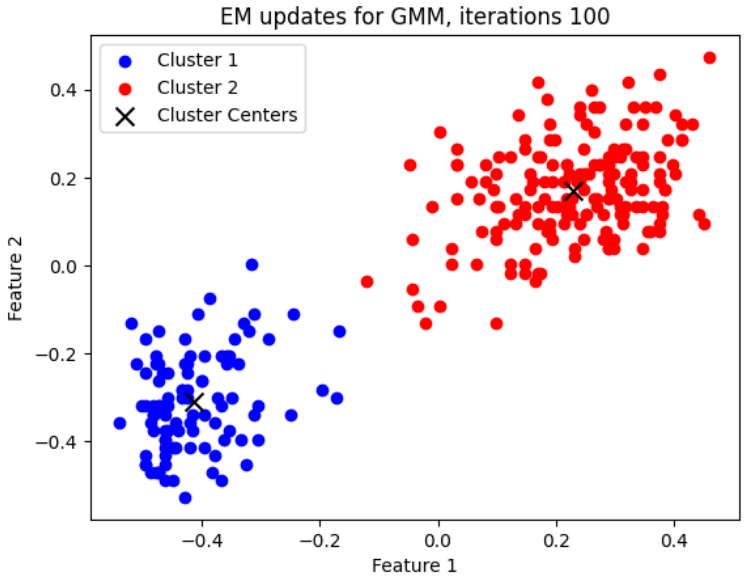
   (c) We can see that having initial cluster centers that are actually close to the real cluster centers will lead the model to converge more quickly. The second model that has cluster centers close to its real cluster centers quickly converged. That is because it took less iterations to shift the centers more close to where they should be.

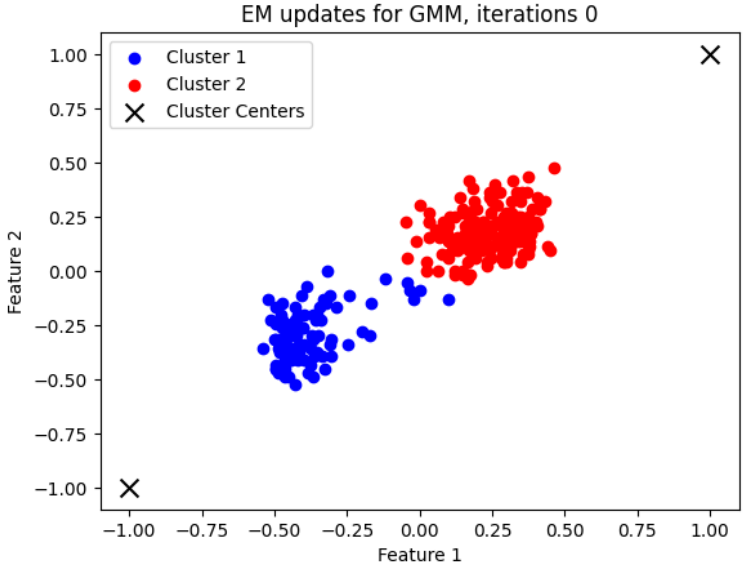

EM updates for GMM, iterations 0
Cluster 1
Cluster 2
Cluster Centers
Feature 1
Feature 2
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

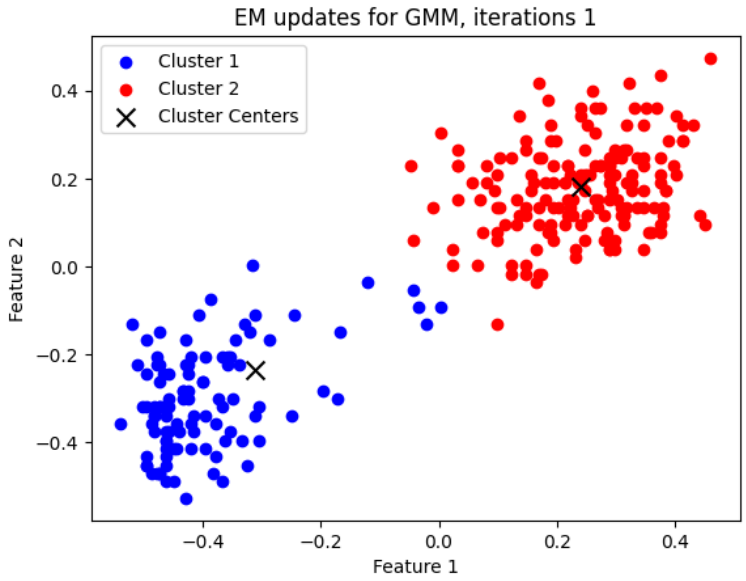

EM updates for GMM, iterations 1
Cluster 1
Cluster 2
Cluster Centers
Feature 1
Feature 2
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4

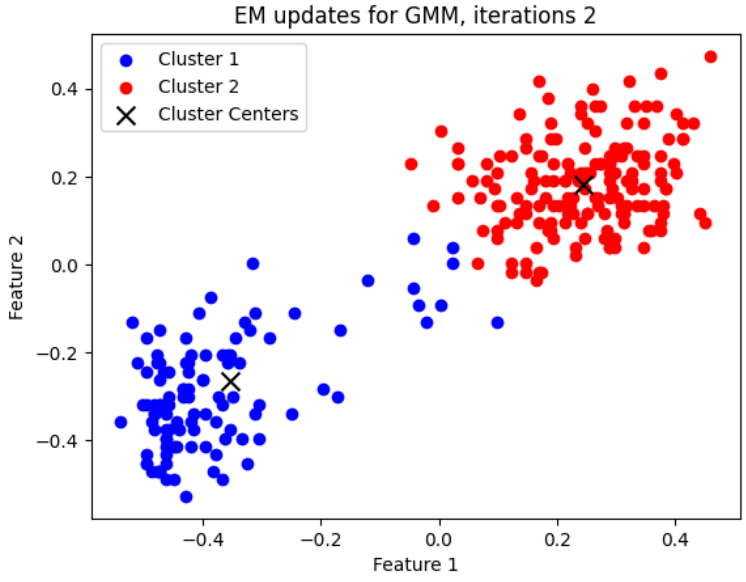

EM updates for GMM, iterations 2
Cluster 1
Cluster 2
Cluster Centers
Feature 2
Feature 1
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4

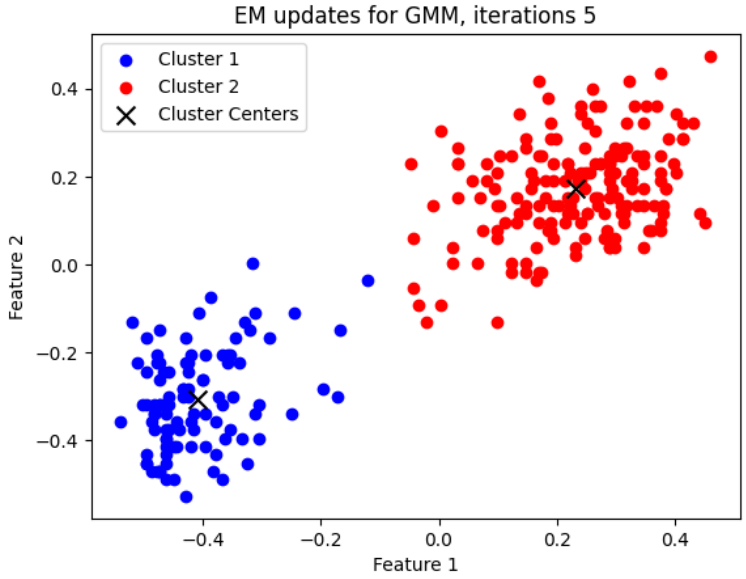

EM updates for GMM, iterations 5
Cluster 1
Cluster 2
Cluster Centers
Feature 2
Feature 1
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4

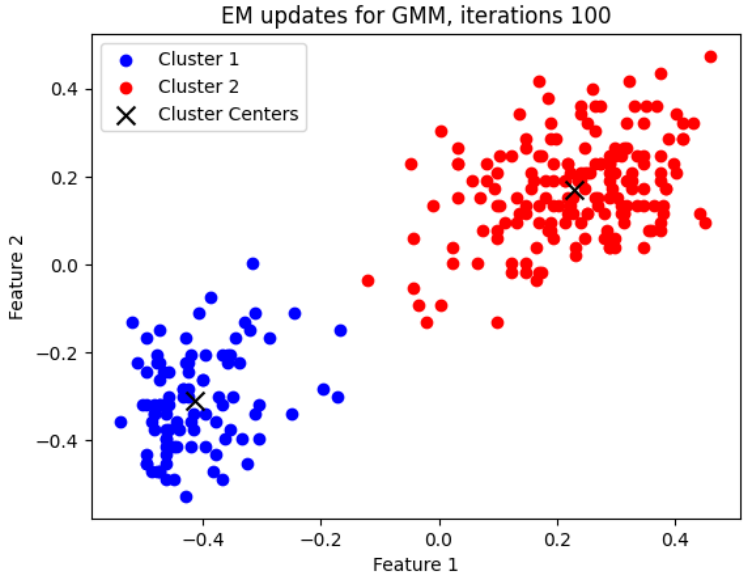

EM updates for GMM, iterations 100
Cluster 1
Cluster 2
Cluster Centers
Feature 2
Feature 1
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4


EM updates for GMM, iterations 0
Cluster 1
Cluster 2
Cluster Centers
Feature 2
Feature 1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00


EM updates for GMM, iterations 1
Cluster 1
Cluster 2
Cluster Centers
Feature 2
Feature 1
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4


EM updates for GMM, iterations 2
Cluster 1
Cluster 2
Cluster Centers
Feature 2
Feature 1
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4


EM updates for GMM, iterations 5
Cluster 1
Cluster 2
Cluster Centers
Feature 2
Feature 1
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4


EM updates for GMM, iterations 100
Cluster 1
Cluster 2
Cluster Centers
Feature 2
Feature 1
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4