EECE5644 - HW2

Shijin Wang

Question1 Part 1:

The class priors are P(L=0) = 0.9 and P(L=1) = 0.1

The parameters of the class-conditional Gaussian pdfs are:

$$\mathbf{m}_0 = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \quad \mathbf{C}_0 = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 2 \end{bmatrix} \quad \mathbf{m}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad \mathbf{C}_1 = \begin{bmatrix} 2 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

Determine the classifier that achieves the minimum probability of error using the knowledge of the true pdf:

$$\frac{P(X|L=1)}{P(X|L=0)} \underset{D=0}{\overset{D=1}{\gtrless}} \frac{P(L=0)(\lambda_{10}-\lambda_{00})}{P(L=1)(\lambda_{01}-\lambda_{11})} = \frac{0.9}{0.1}$$

$$\frac{g_1(x|m_1, C_1)}{g_0(x|m_0, C_0)} \underset{D=0}{\overset{D=1}{\gtrless}} 9$$

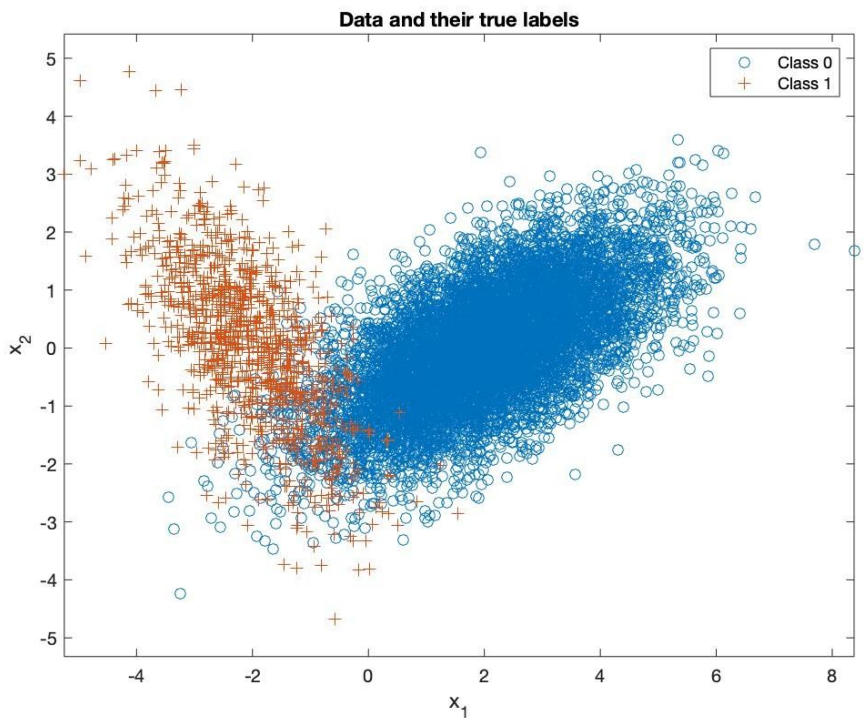

Figure 1. Data and their true labels for the validation

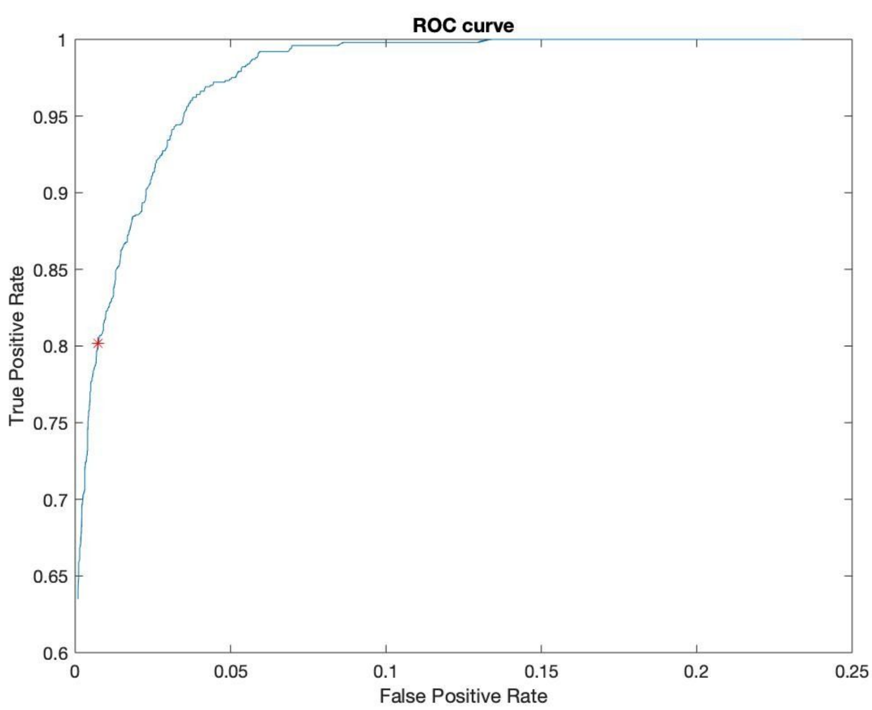

Figure 2. ROC curve and the location of the min-P(error)

Estimated min-P(error):

```
"Min Perror is "    "0.0264"

"Best threshold is "    "11.9971"

"tp is "    "0.80199"

"fp is "    "0.0072262"

"perror"    "0.0267"
```

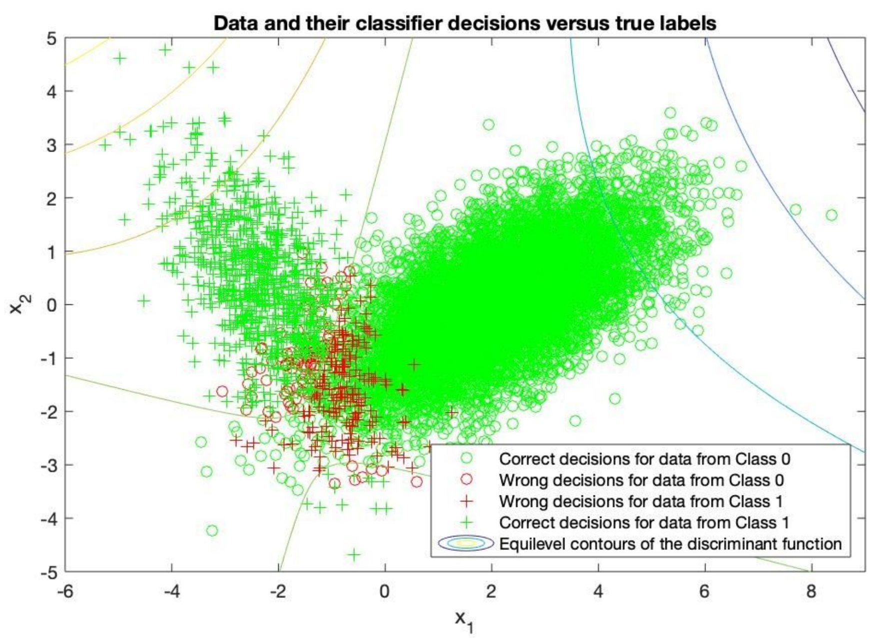

Figure 3. Decision boundary overlaid on the validation dataset

Question 1 Part 2:

How to use a class-label-posterior approximation to classify a sample:

We can use MAP: $\underset{\partial}{\text{argmax}}\, p(\partial | D) = \underset{\partial}{\text{argmax}} \frac{p(D|\partial)p(\partial)}{p(D)}$

$= \underset{\partial}{\text{argmax}} \ln p(D|\partial)) + \ln p(\partial)$

Then we can estimate mean and variance from this likelihood function

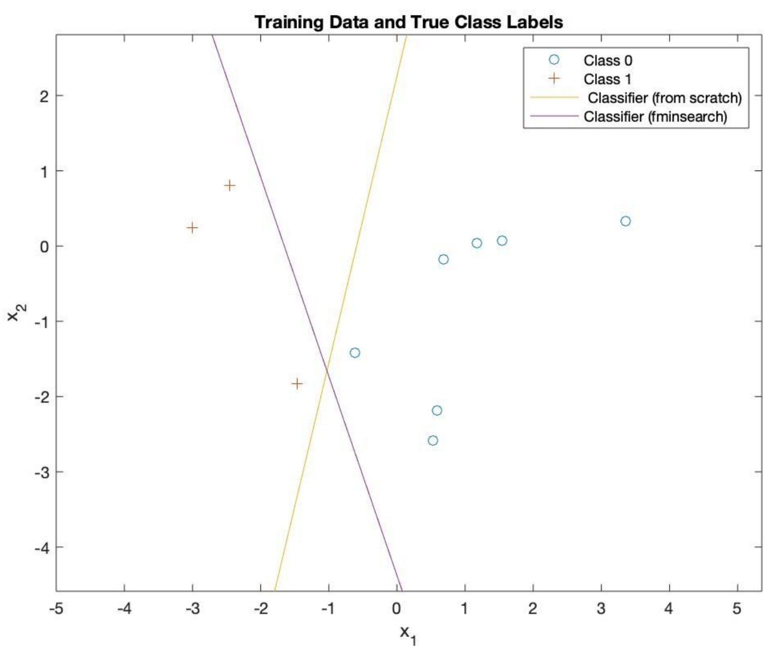

Figure 4. Decision boundary for D=10 Training Data

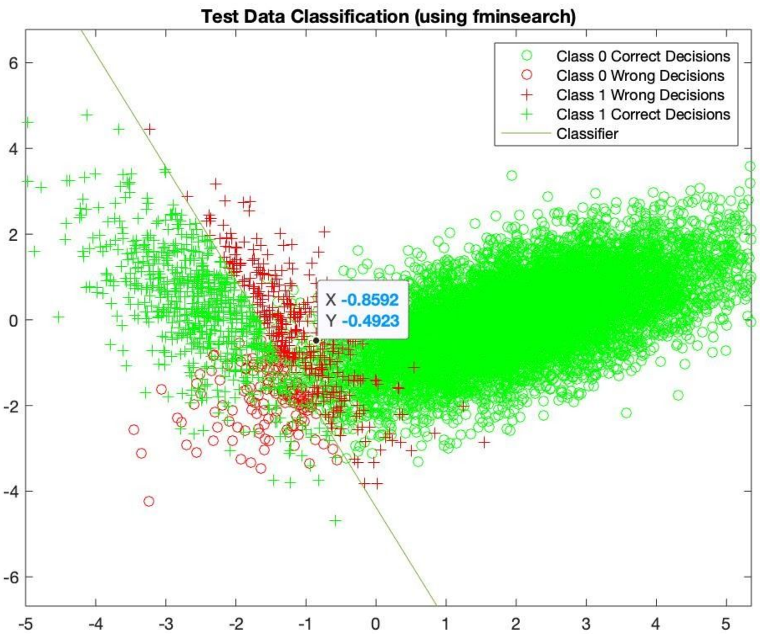

Figure 5. Decision boundary for D=10 on the validation dataset

Probability of error:

Total error (classifier using fminsearch): 4.63%

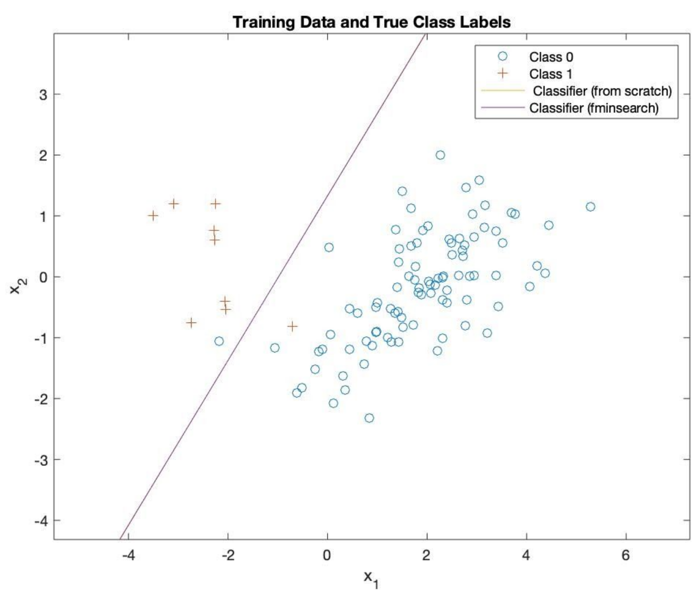

Figure 6. Decision boundary for D=100 Training Data

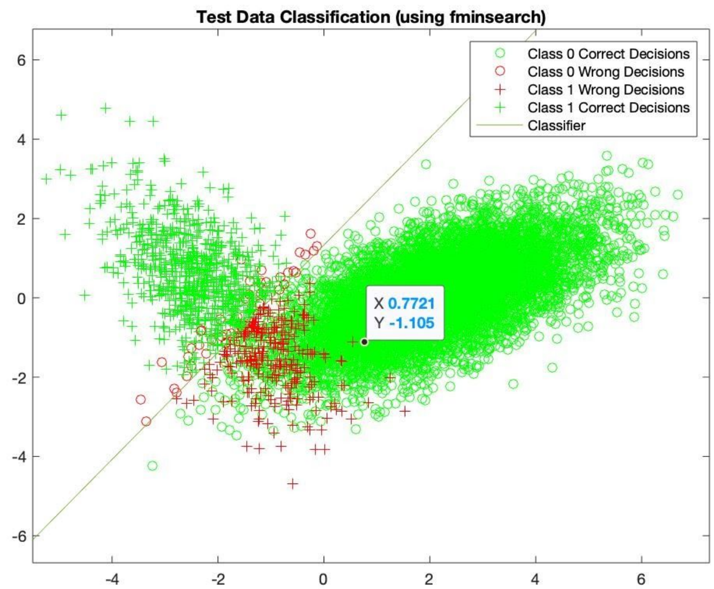

Figure 7. Decision boundary for D=100 on the validation dataset

Probability of error:

Total error (classifier using fminsearch): 3.15%

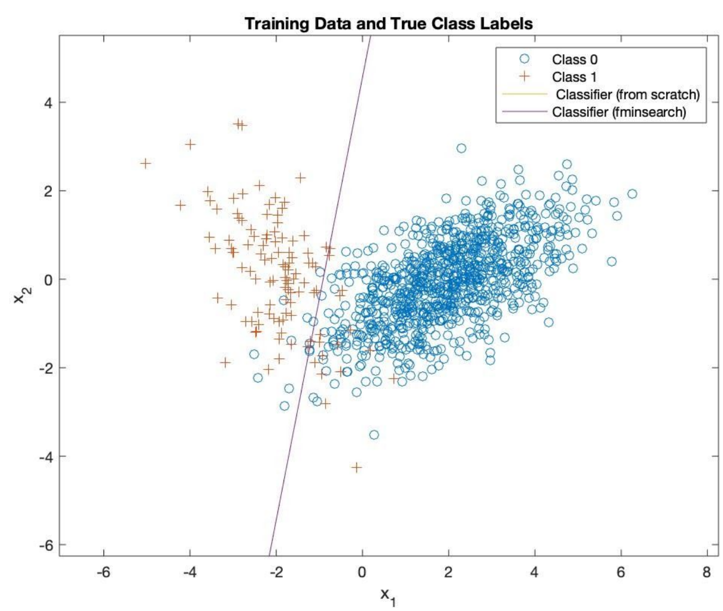

Figure 8. Decision boundary for D=1000 Training Data

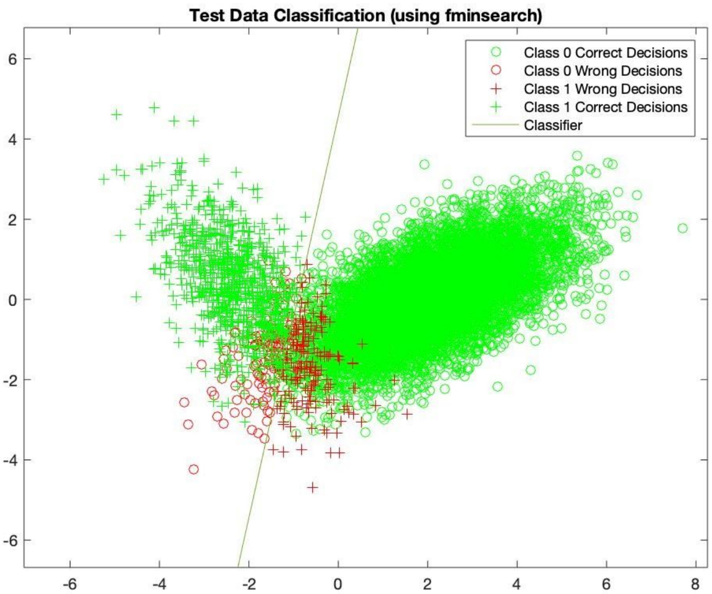

Figure 9. Decision boundary for D=1000 on the validation dataset

Probability of error:

Total error (classifier using fminsearch): 2.90%

Question 1 Part3:

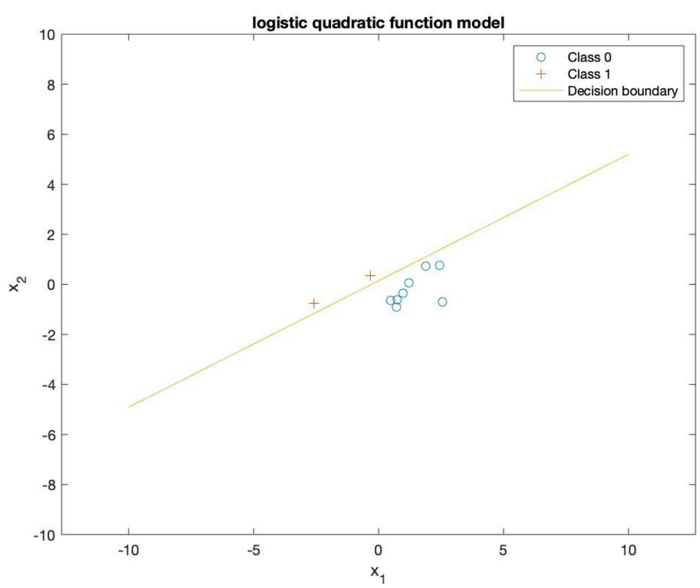

Figure 10. Decision boundary for D=10 on the validation dataset

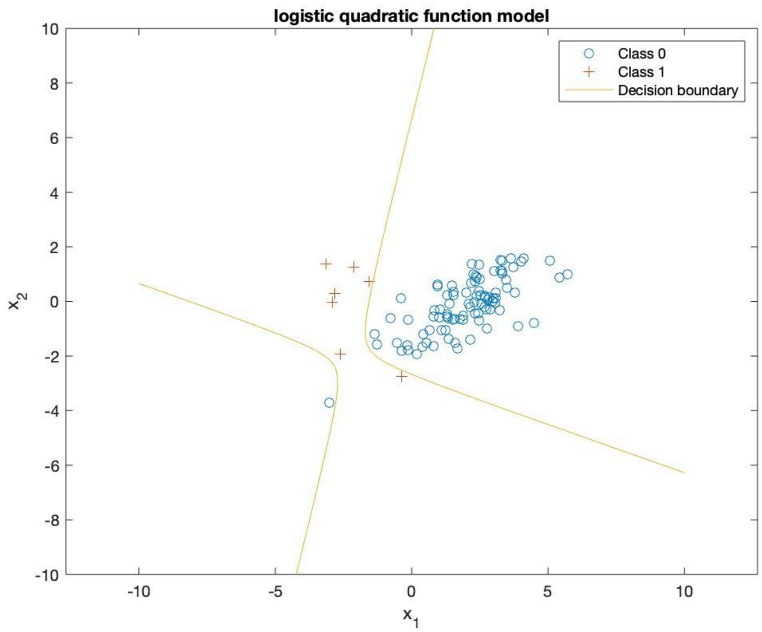

Figure 11. Decision boundary for D=100 on the validation dataset

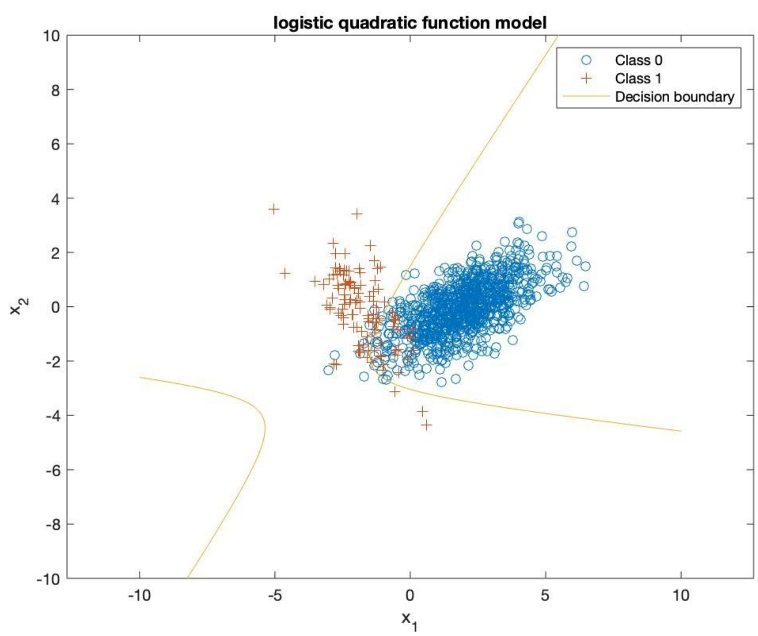

Figure 12. Decision boundary for D=1000 on the validation dataset

Question2:

Determine the MAP estimate:

$w = [a.\ b.\ c.\ d]^T$

$$w = \arg\max_{w} p(w)\, p(D|w) = \arg\max_{w} p(w) \prod_{i=1}^{N} p(D_i|w)$$

$$= \arg\max_{w} \frac{1}{4\pi^2\gamma^4} \exp\left(\frac{-a^2+b^2+c^2+d^2}{2\gamma^2}\right) \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-ax^3-bx^2-cx-d)^2}{2\sigma^2}\right)$$

$$= \arg\max \exp\left(-\frac{a^2+b^2+c^2+d^2}{2\gamma^2}\right) \prod_{i=1}^{N} \exp\left(-\frac{(y_i-ax_i^3-bx_i^2-cx_i-d)^2}{2\sigma^2}\right)$$

$$= \arg\min \frac{a^2+b^2+c^2+d^2}{2\gamma^2} + \sum_{i=1}^{N} \frac{(y_i-ax_i^3-bx_i^3-cx_i-d)^2}{2\alpha^2}$$

$$= \arg\min \frac{w^Tw}{2\gamma^2} + \sum_{i=1}^{N} \frac{(x_i^Tw-y_i)^2}{2\alpha^2}$$

$$= \arg\min (Xw-y)^T(Xw-y) + \frac{\alpha^2}{\gamma^2} w^Tw$$

Apply regression:

$$L(w) = (Xw-y)^T(Xw-y) + \frac{\alpha^2}{\gamma^2} w^Tw$$

$$= w^TX^TXw - y^TXw - w^TX^Ty + y^Ty + \frac{\alpha^2}{\gamma^2} w^Tw$$

$$\Rightarrow \frac{\partial L(w)}{\partial w} = 2X^TXw - X^Ty - X^Ty + \frac{2\alpha^2}{\gamma^2} w$$

Set $\frac{\partial L(w)}{\partial w} = 0$

So that $w = (X^TX + \frac{\alpha^2}{\gamma^2} I)^{-1} X^Ty$

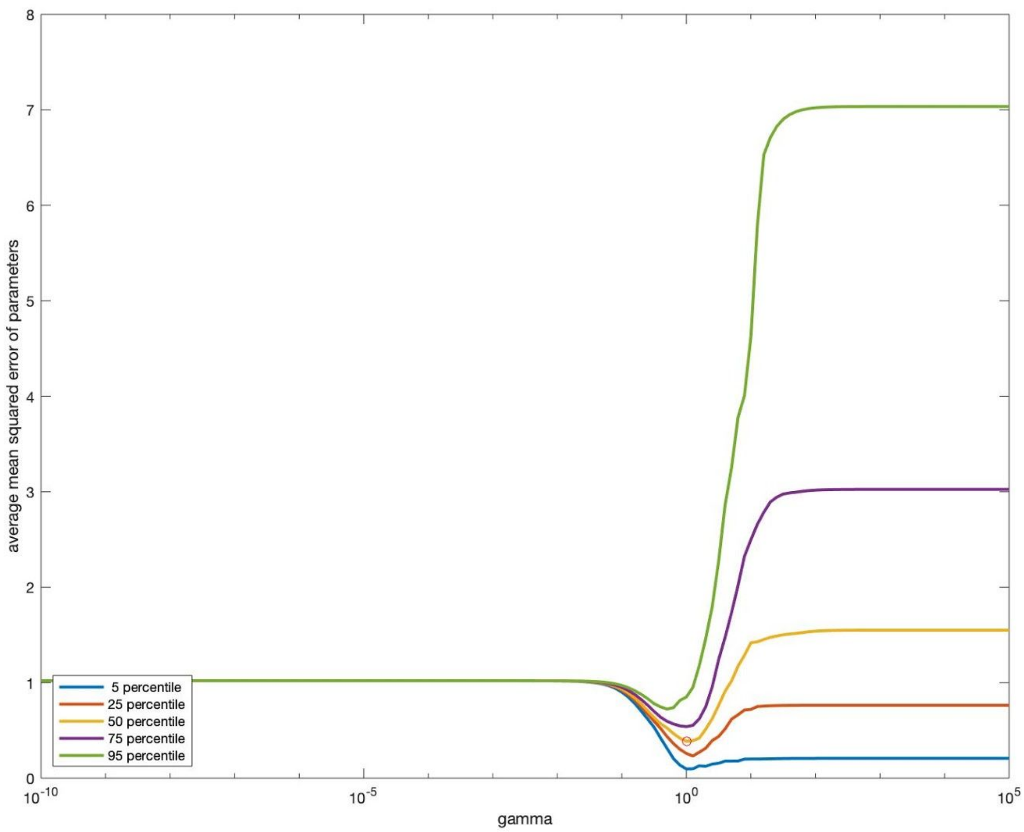

Figure 13. MAP estimator for each value of γ in a single plot

From the plot, we can see that as the gamma increases, the squared error decreased at first. However, after gamma = 10, the squared error increased significantly when gamma goes up.

Question 3:

GMM components:

$$u_1 = \begin{bmatrix} -8 \\ -8 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 24 & 1 \\ 1 & 8 \end{bmatrix} \quad \alpha_1 = 0.2$$

$$u_2 = \begin{bmatrix} -8 \\ 8 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 25 & 3 \\ 3 & 9 \end{bmatrix} \quad \alpha_2 = 0.25$$

$$u_3 = \begin{bmatrix} 8 \\ 8 \end{bmatrix} \quad \Sigma_3 = \begin{bmatrix} 14 & -5 \\ -5 & 16 \end{bmatrix} \quad \alpha_3 = 0.3$$

$$u_4 = \begin{bmatrix} 8 \\ -8 \end{bmatrix} \quad \Sigma_4 = \begin{bmatrix} 6 & 1 \\ 1 & 20 \end{bmatrix} \quad \alpha_4 = 0.35$$

In order to perform Cross-validation. I first split the dataset to 10 block. Then, in each iteration. I use 1/10 as the Validation set. the rest of the data will be the training set. Each cross validation pick current number of gaussian components for the number of random initializations.

Average log-likelihood of testing data for 6 components:

sample \ components	1	2	3	4	5	6
10	-7.94	-10.14	-12.2	-36.6	-110.4	-110.4
100	-71.15	-69.4	-70.48	-68.76	-69.9	-71.66
1000	-722.8	-706.4	-692.9	-685	-686.3	-685.6

For 10 samples, the selection is 1 component

For 100 samples, the selection is 4 component.

For 1000 samples. the selection is 4 component.

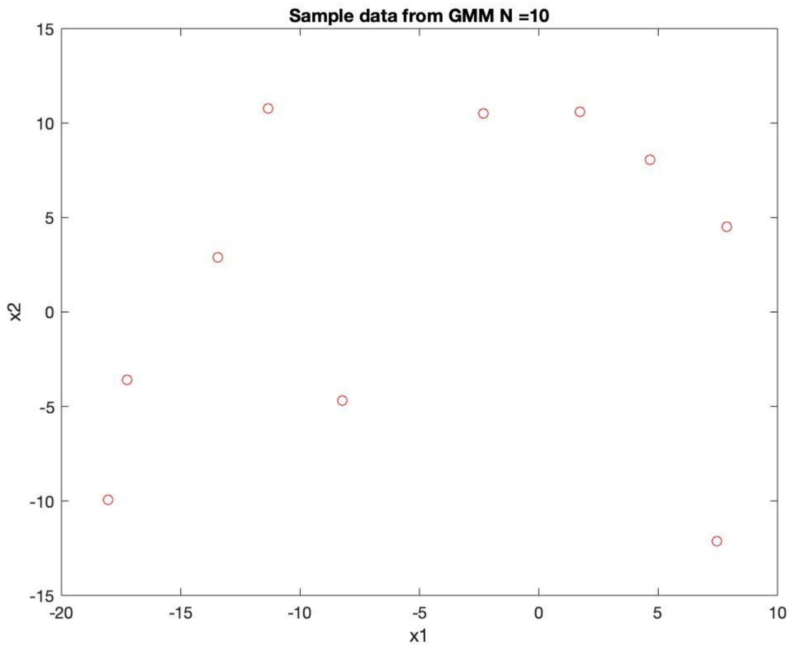

Figure 14. Sample data generated by GMM when N = 10

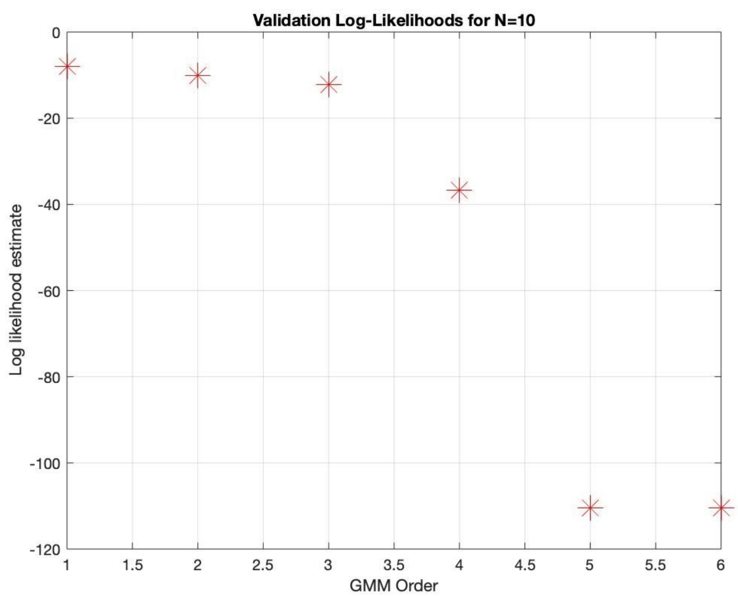

Figure 15. Validation log-likelihoods when N = 10 using cross-validation

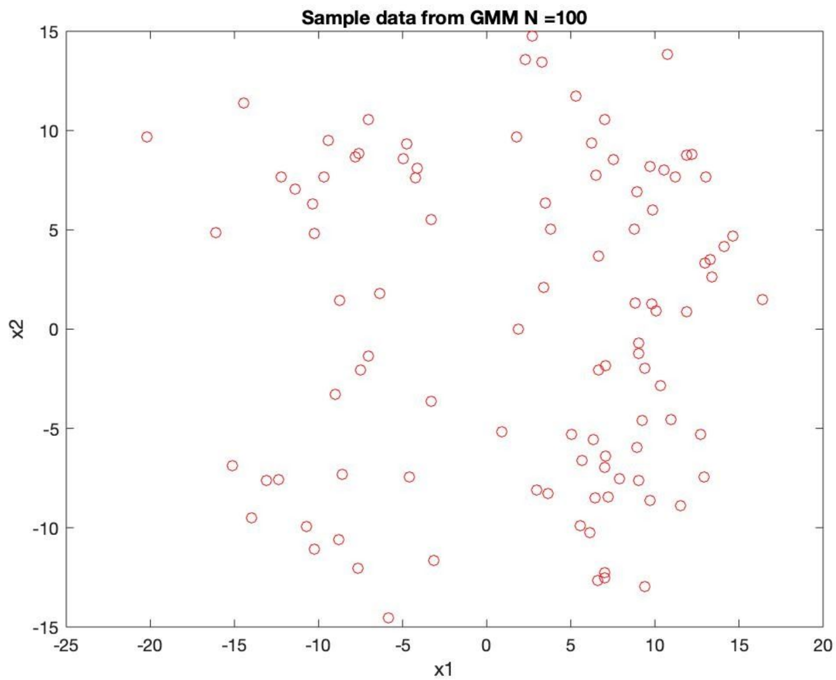

Figure 16. Sample data generated by GMM when N = 100

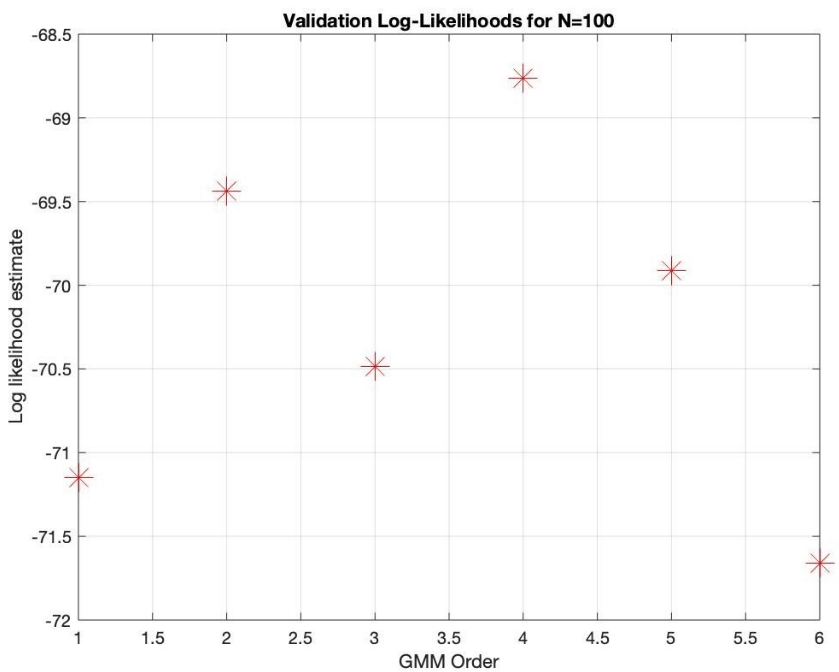

Figure 17. Validation log-likelihoods when N = 100 using cross-validation

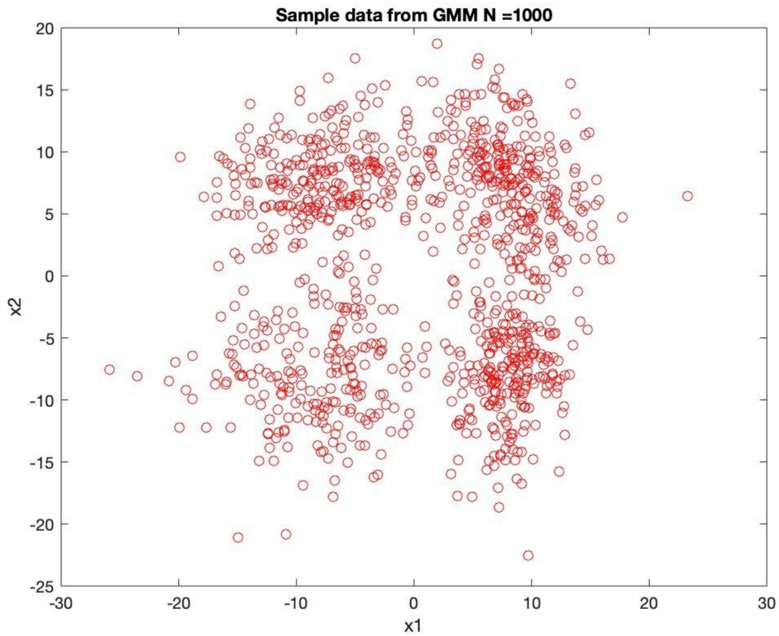

Figure 18. Sample data generated by GMM when N = 1000

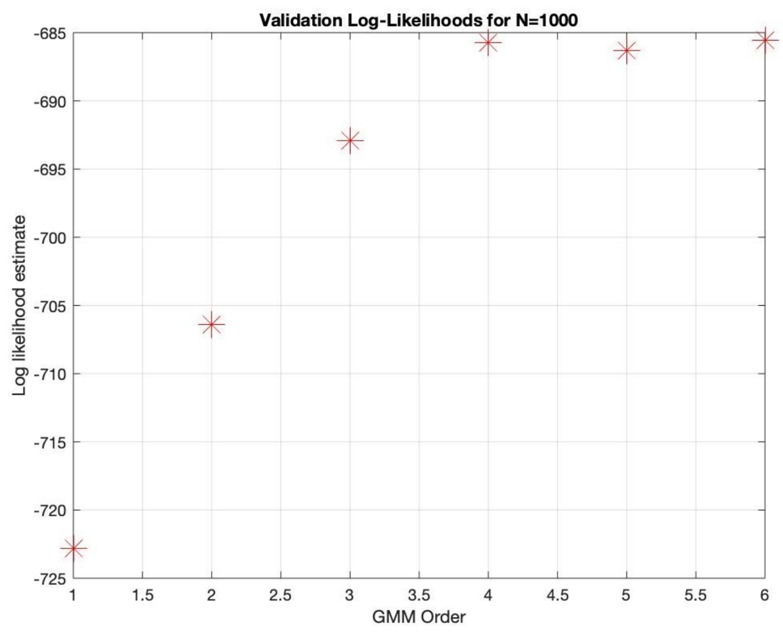

Figure 19. Validation log-likelihoods when N = 1000 using cross-validation