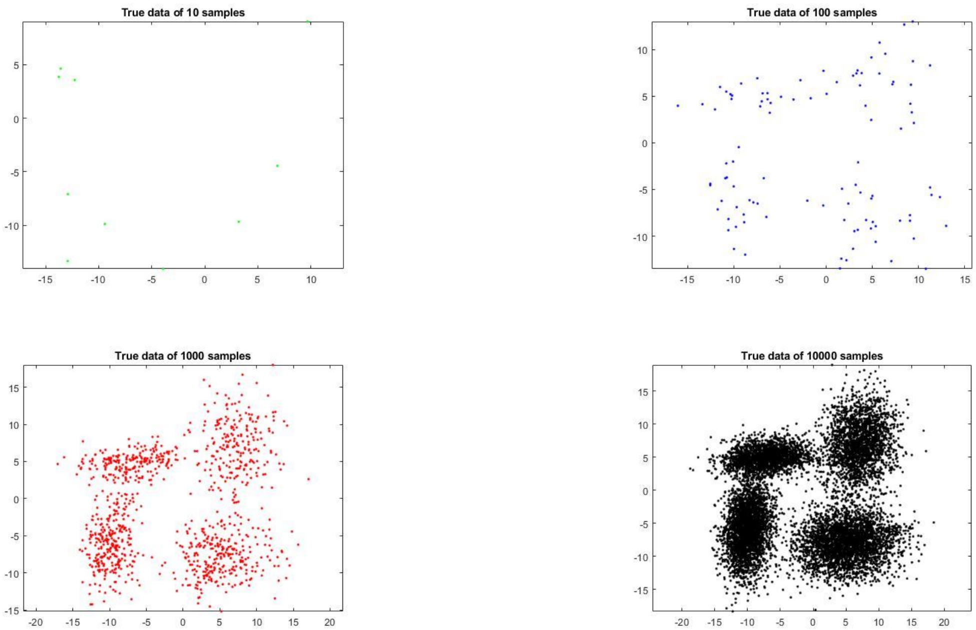
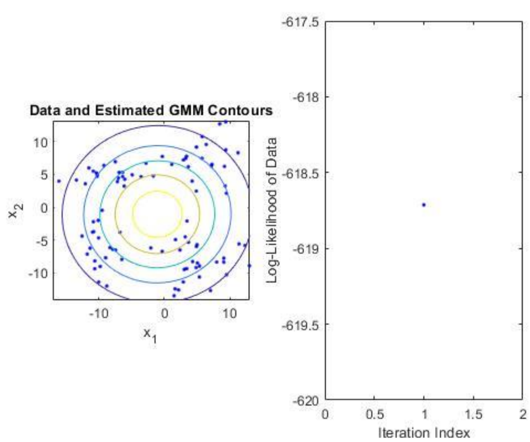
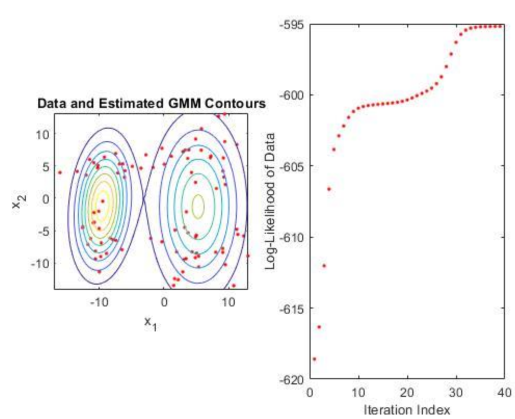
EECE5644
Yu Shun Lin
ID 001459022 Homework 03

1. Conduct GMM model using 10-fold cross-validation procedure



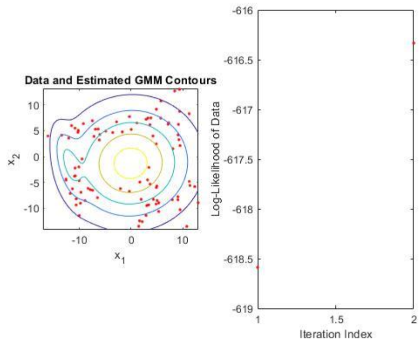
These plots show the true data of 10, 100, 1000, 10000 sample of the parameter
Alpha = [0.3, 0.2, 0.2, 0.3]
Mean = [-10 -7 7 5;-6 5 7 -8];
Sigma(:,:,1) = [3 1;1 12];
Sigma (:,:,2) = [12 1;1 2];
Sigma (:,:,3) = [7 1;1 13];
Sigma (:,:,4) = [15 1;1 7];
Apply the 10-fold to each sample, that the 10-sample case is not enough for 10-fold cross validation because there are not enough sample for picking the initial EM algorithm. Thus, dismiss the 10-sample case.



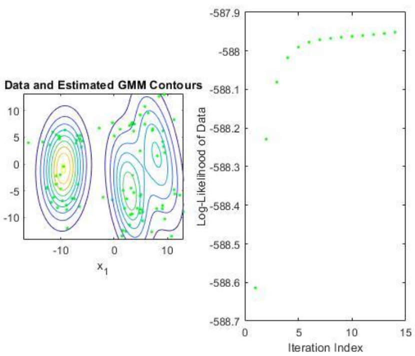


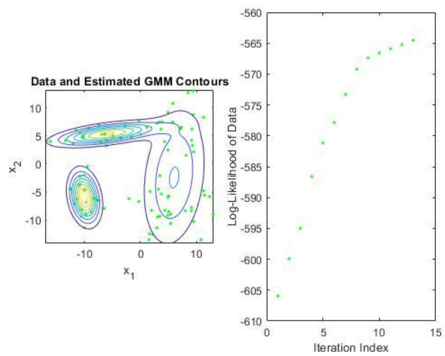
These two plots shows that using 1, 2 as the number of Gaussian training component.
In the order 1, only one point is picked for the EM algorithm. Thus the likelihood will set up immediately to determine the new mean and sigma value.

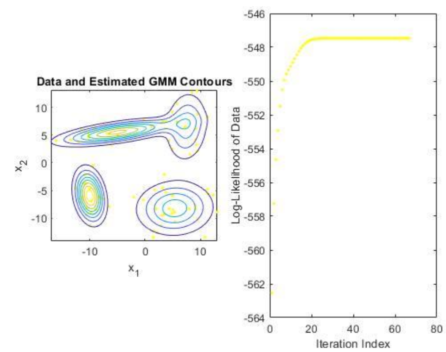
In second plot, it shows that assuming two components of the Gaussian training.



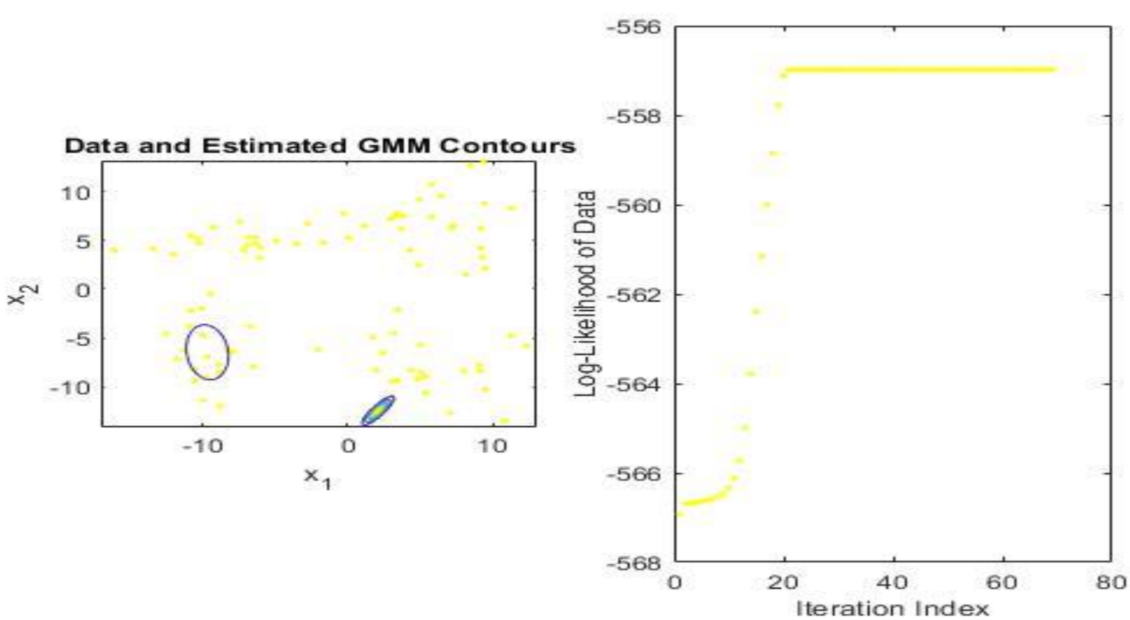
Here shows the case that in the beginning of the order 2 case in one of the 10-fold training. It will finally convergence to two gaussian contour.





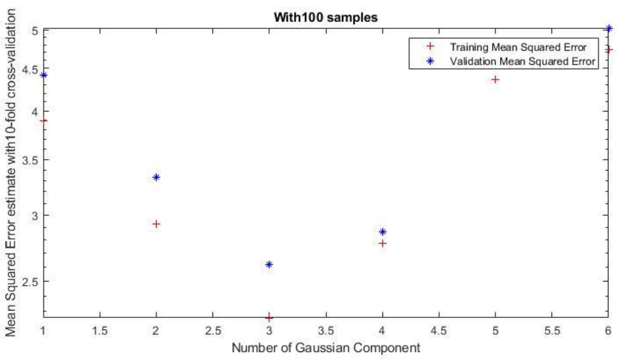


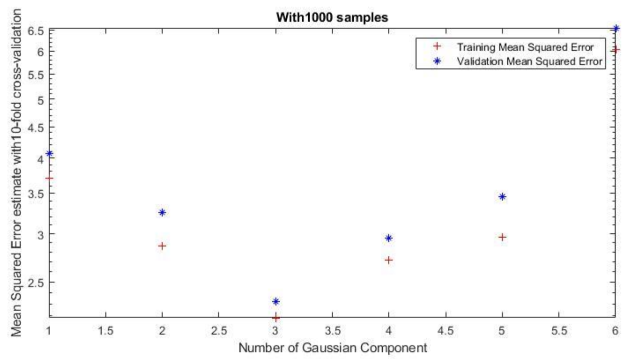
Same, we can observe the processing of order 3 and order 4. However, the processing will not stop for the convergence sometimes. The following plot will show the bad case.

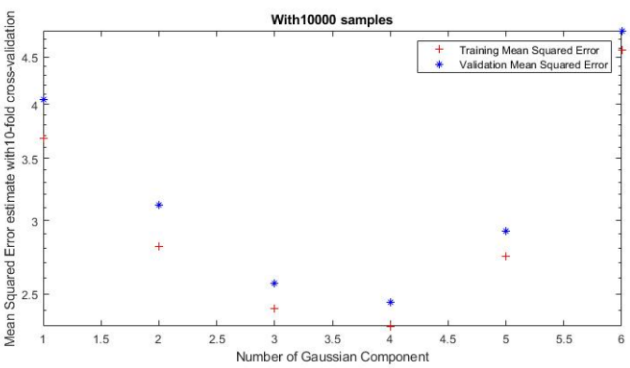


This plot shows the bad case of the processing will not terminate. The EM algorithm pick the numbers of the point as the assuming components. However, the picking point will gather up sometimes. This means the initial point is picked up from the same true GMM group. Thus, the processing will not find the best parameter sometimes. Thus, setting a threshold for determine the bad choosing case to continue the next block training.

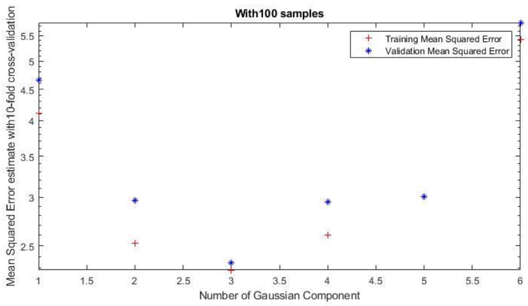
By the observation and the testing mentioned before, we could calculate the mean squared error of each number of Gaussian guessing. The following plots show the result of the mean squared error of different samples and different order.

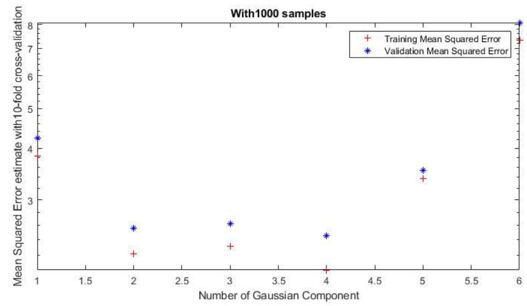


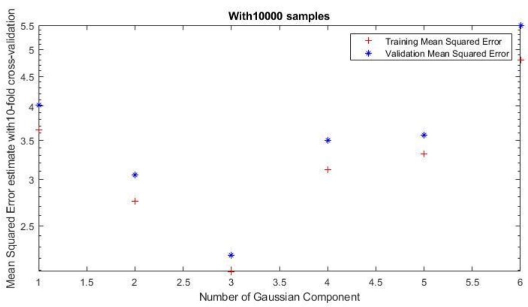




According to the plots showed above. When we have 100 samples. The best guessing is setting Gaussian component number as 3. 1000 samples case is the same. 10000 sample case is determined as 4 is the best number of the component. We can observe that the training result of error is decreasing when the samples number increase. Finally, we determine the best order will be around 3 to 4 because the true number of Gaussian components is 4. The result will not be more than 4 is because more parameter has determined, more error will be detected when the assuming number is large than true number.







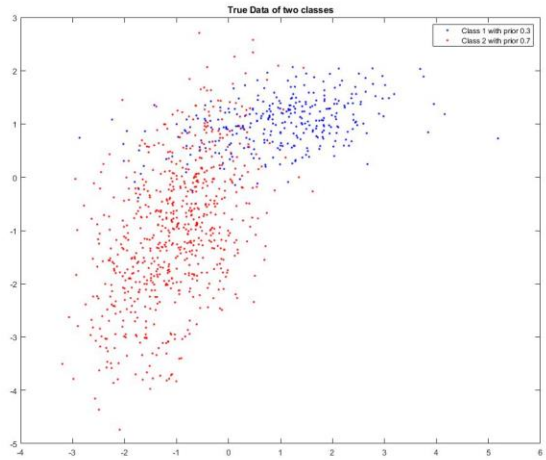
Another training case: this case shows 3 is the best when 100 samples, 4 is the best when 1000 samples, 3 is the best when 10000 sample. Although it has a different result. All the error estimation will decrease when the amount of sample increase.

Now we try to discreate the data sets more far away, which means the overlapping will be more fewer.

The mean squared error result is as follow.

EECE5644
Yu Shun Lin
ID 001459022 Homework 03

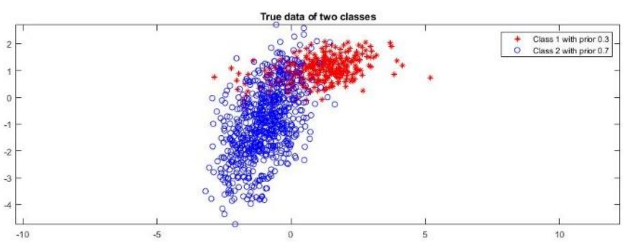
2. Training for the classifier with MAP. Fisher LDA classifier and Logistic classifier with 999 2 dimensional samples.



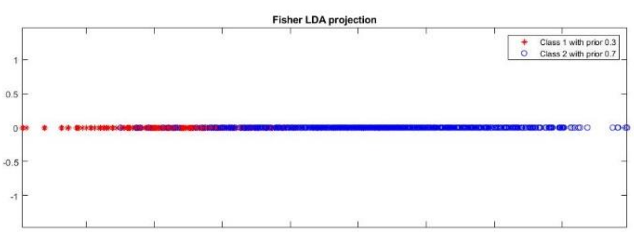
The plot shows the true data of two data sets with 2 diremption, different mean value and different sigma value. The covariance matrices are no diagonal with distinct eigenvalues (which means two set have different direction.)

mu = [1 -1;1 -1];
Sigma(:,:,1) = 0.1*[15 2;3 2];
Sigma(:,:,2) = 0.1*[6 7;3 17];
classPriors = [0.3,0.7];

Fisher LDA classifier



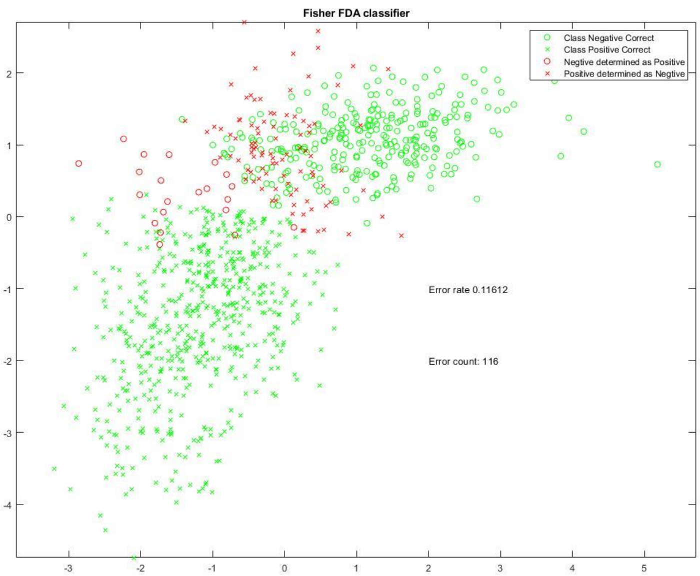
These plots show the fisher LDA projection for the classifier. Blue spots show the positive data, and the red spots shows the negative data.



Spots lay on LDA line on both side negative and positive. This could provide for making the decision. Red spot lays > 0 means negative data is decided as positive. On the other hand, blue spots lays < 0 means positive data is decided as negative.

EECE5644
Yu Shun Lin
ID 001459022
Homework 03

Corresponding to the indices of true data, the following plot show the fisher LDA classifier result.



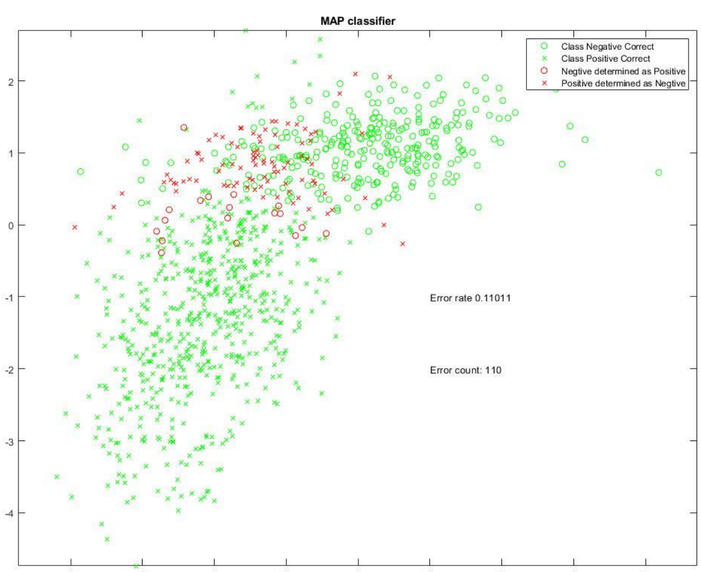
Result show 116 error count in 999 data.

Error rate is 0.11612 = 11.612%

MAP classifier

With the Maximum a prior method, get the discriminant score from both negative and positive data sets. Choose the bigger discriminant score to make the decision if it is negative or positive.

The following shows the MAP classifier result



Result show 110 error count in 999 data

Error rate is 0.11011 = 11.011%

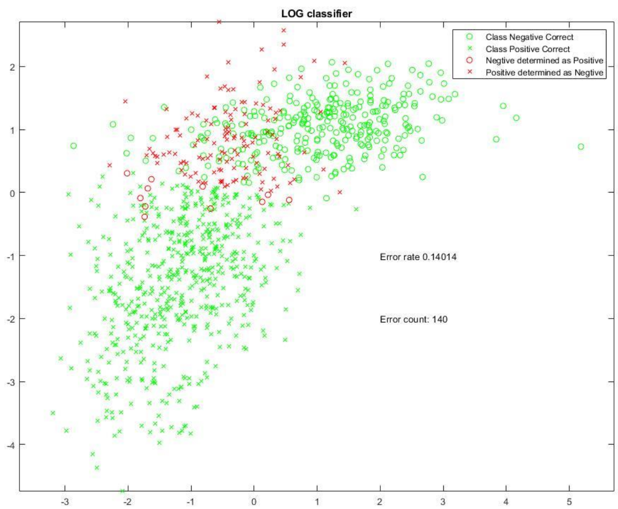
Logistic Linear classifier

Train the parameters of a logistic function y(x) = 1/(1 + exp (w T x+b)), model the positive likelihood that y(x) ≈ P(Label = +|x) and 1−y(x) ≈ P(Label = −|x) as negative. To find the best parameter of w and b. Train the data sets with $argmin(-\frac{1}{N})\sum_1^N \ln(\frac{1}{1+e^{w^T x2+b}})$, we dismiss the $\sum_1^N \ln(x)$ because the former equation is enough to train the model. Suppose the mean bias = 0 to train the data.

EECE5644
Yu Shun Lin
ID 001459022 Homework 03

Using fminsearch in matlab, initial w = LDA's w as the starting point to train model.
The result comes with w = [-28.0441;-172.6168]/numbers of positive data.

The following plot show the result of Logistic classifier



Result show 140 error counts in 999 data.

Error rate is 0.14014 = 14.014%

Reference

https://ml-cheatsheet.readthedocs.io/en/latest/linear_regression.html#multiple-linear-regression-predict
https://medium.com/datadriveninvestor/k-fold-cross-validation-6b8518070833
Richard O Duda, Peter E Heart, David G Stork, Pattern Classification, Ch5.8, Ch3.2

Code resource

Problem1 : https://github.com/MakiseYuki/EECE5644-Machine-Learning/tree/master/Homework%203/Problem1
Problem 2 : https://github.com/MakiseYuki/EECE5644-Machine-Learning/tree/master/Homework%203/Problem2