PATTERN RECOGNITION ASSIGNMENT-2
REPORT
Bayes Classifier with Gaussian Mixture Model

9th October 2016

GROUP - 3

NEHA MUTHIYAN	**SIDDHANT KUMAR**
B14113	**B14133**
muthiyan_neha@students.iitmandi.ac.in	*siddhant_kumar@students.iitmandi.ac.in*

in the guidance of
Dr. A.D Dileep
(Asst. Professor)
School of Computing and Electrical Engineering
Indian Institute of Technology, Mandi

Introduction

Gausian Mixture Model:

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm. Unlike Bayes classification model, in GMM every class is assumed to have multiple gaussian clusters.

Let class C_i have K gaussian clusters then the probability of **x** given class C_i is,

$$p(\mathbf{x} \mid C_i) = \Sigma_{k=1:K} \ \Pi_k \ p_k(\mathbf{x} \mid \boldsymbol{\mu}_{ik}, \Sigma_{ik})$$

Here,
Π_k : mixture coefficient of k^{th} cluster
$p_k(\mathbf{x} \mid \boldsymbol{\mu}_{ik}, \Sigma_{ik}) = N(\mathbf{x} \mid \boldsymbol{\mu}_{ik}, \Sigma_{ik})$

K means clustering:

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through certain number of clusters (assume k clusters) fixed apriori. The choice of K is purely experimental.

The decision boundary in K means clustering depends on the distance measure used. Incase of squared eucledian distance the decision is linear whereas for Mahalanobis distance the decision boundary is hyperquadratic.

Table of Contents

DECLARATION OF AUTHORSHIP

We hereby certify that the report is entirely our own original work except where otherwise stated.
Our work can be found at the following Github repository -
https://github.com/saytosid/PR

Name – Neha Muthiyan (B14113)
Siddhant Kumar (B14133)

Date – 9th September 2016

Acknowledgements

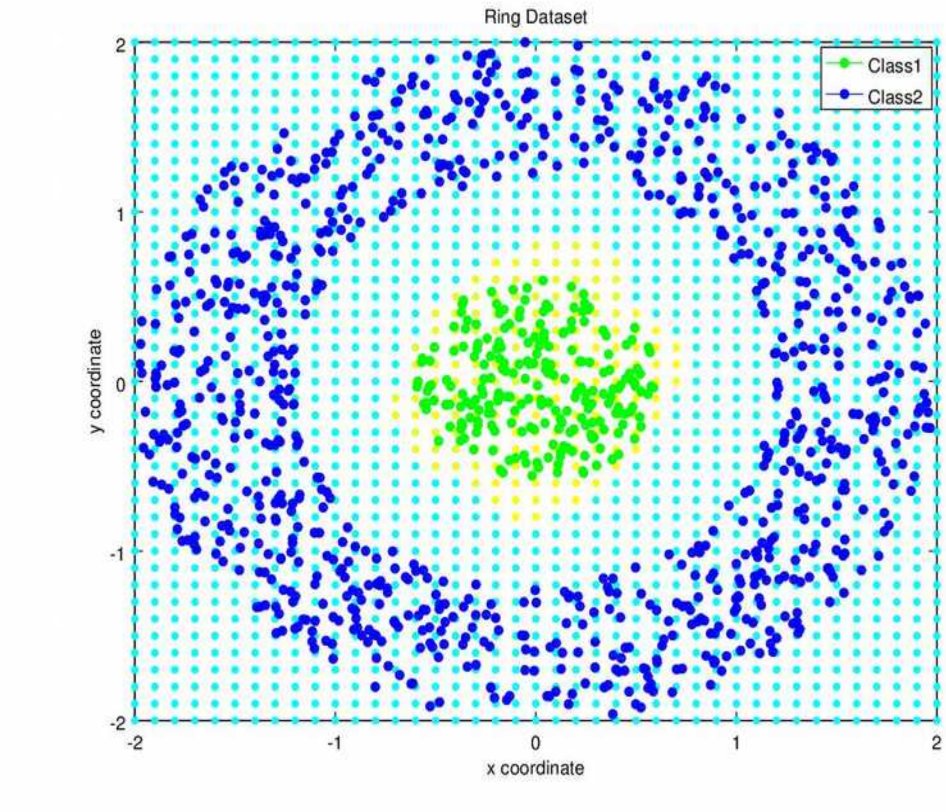
Every project is successful largely due to the effort of number of people who have always given their valuable advice or lent a helping hand. We sincerely appreciate the inspiration, support and guidance of all those people who have been instrumental in the completion of this assignment.

We are also grateful to our advisor Dr. A.D Dileep for his provision of expertise and technical support. His superior knowledge and experience played a great role in the completion of this assignment.

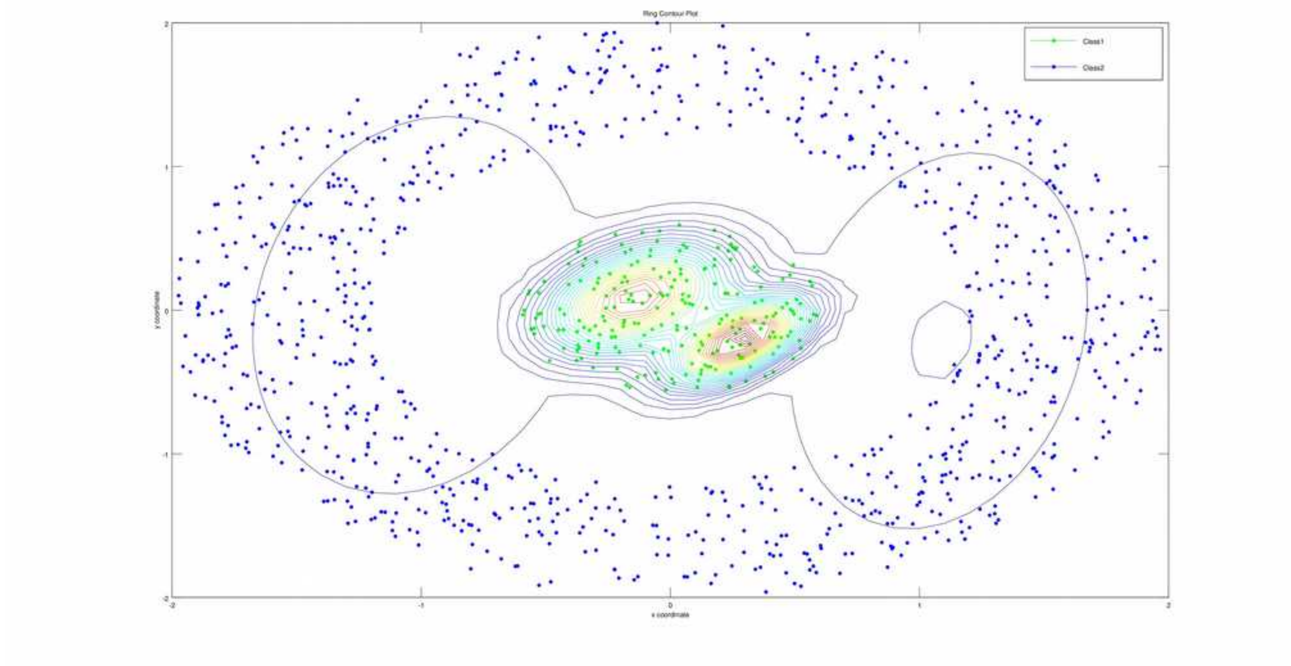
Experimental Observations

3.1 Ring dataset

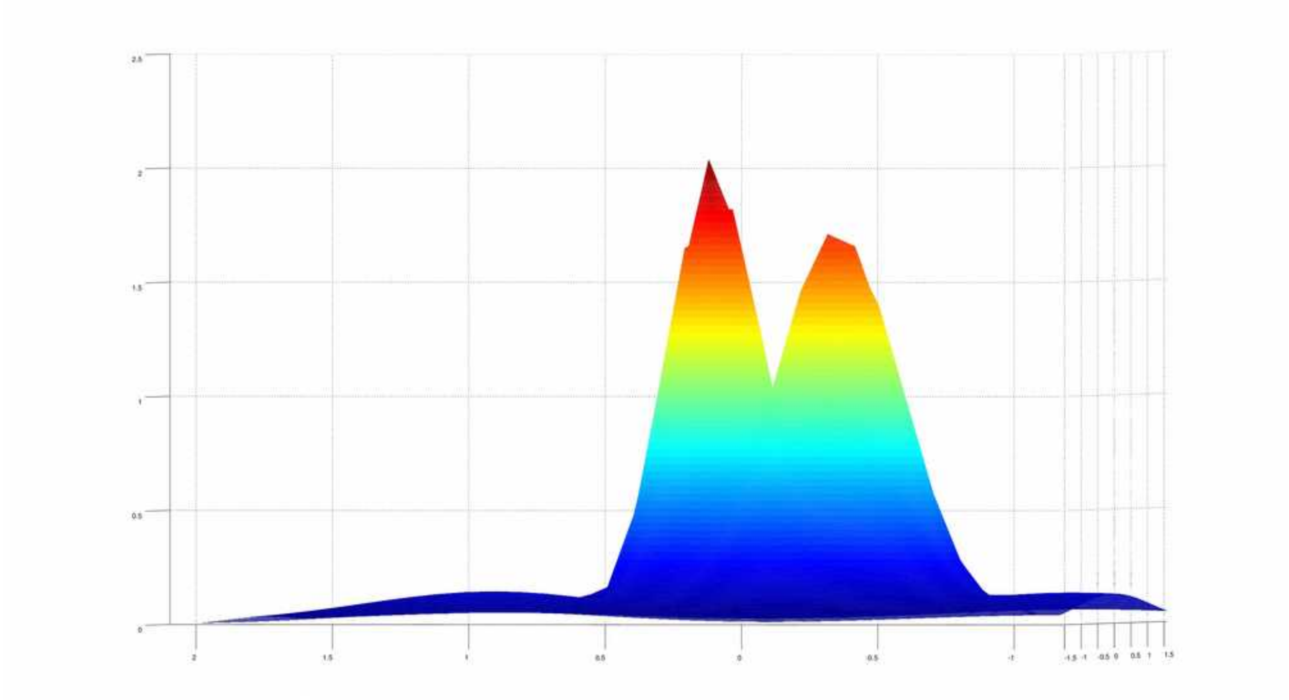
3.1.1 2 clusters



Contour Plot (K = 2):



3D contour plot (K = 2):



K = 2

Classification Accuracy (%)	100%
Precision for Class1 Precision for Class2	1 1
Mean Precision	1
Recall for Class1 Recall for Class2	1 1
Mean Recall	1
F-measure for Class1 F-measure for Class2	1 1
Mean F-measure	1

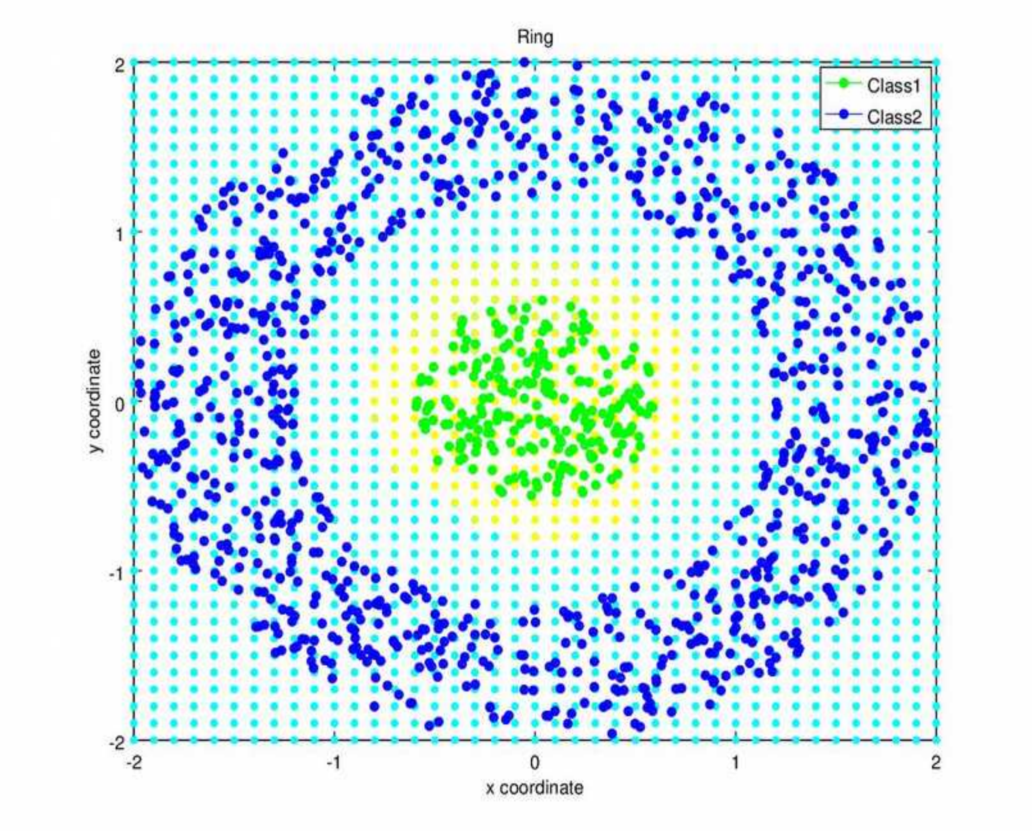
Confusion Matrix :

$$C = \begin{matrix} 75 & 0 \\ 0 & 300 \end{matrix}$$

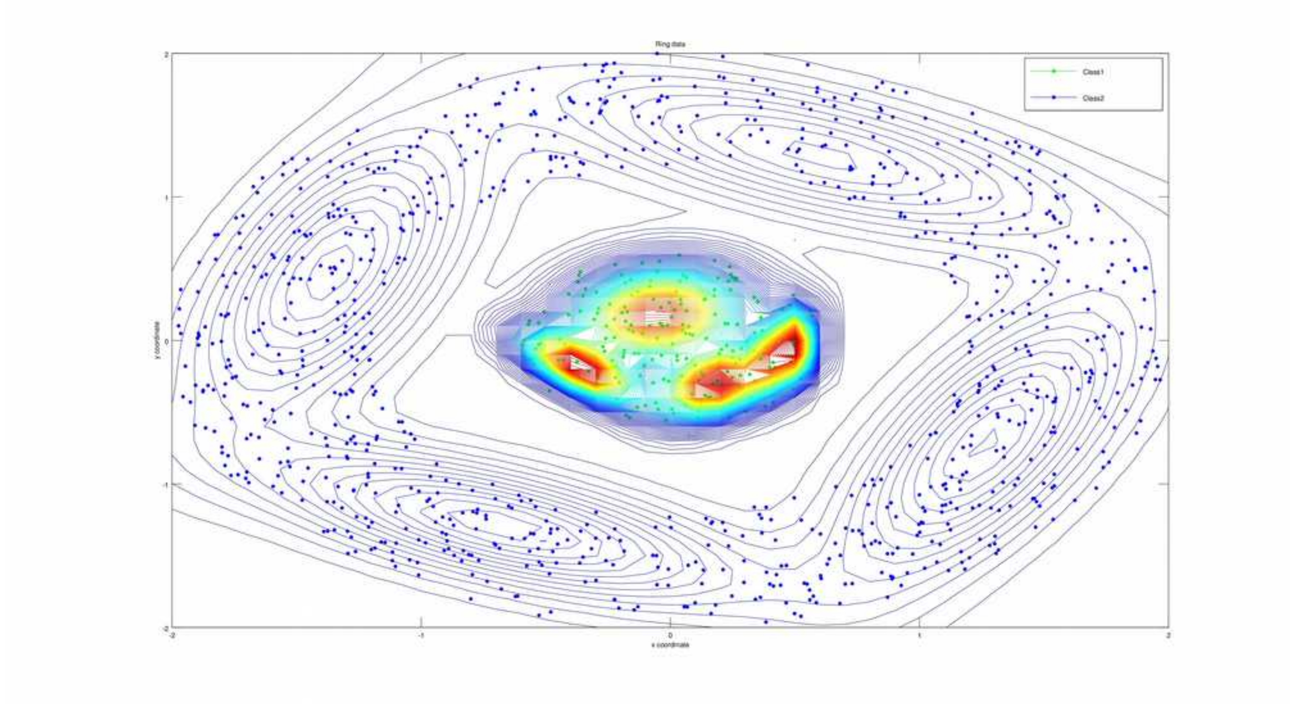
Observations :

- The accuracy obtained in GMM is better than in unimodal gaussian model.
- In unimodal case, the decision surfaces obtained were linear and elliptical depending on the covariance matrices of classes.
- The contours for outer class are lower than those for inner class because the variance for the outer class is very high due to which the peaks are lower in height.
- The contours of inner class define the decision boundary at the points of intersection with the contours of outer class.
- The nonlinear decision boundary in the graph is an envelope of the gaussian components.

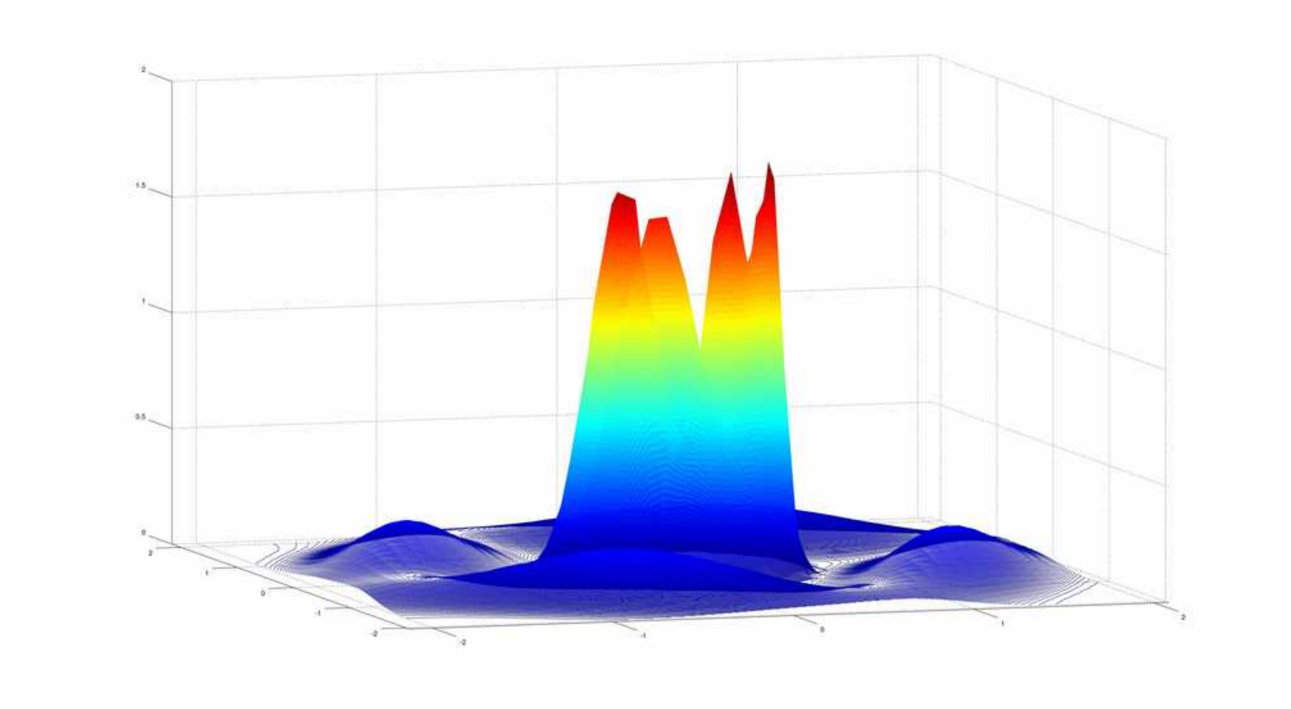
3.1.2 4 clusters



Contour Plot (K = 4)



3D Contour Plot:



For K = 4

Classification Accuracy (%)	100%
Precision for Class1 Precision for Class2	1 1
Mean Precision	1
Recall for Class1 Recall for Class2	1 1
Mean Recall	1
F-measure for Class1 F-measure for Class2	1 1
Mean F-measure	1

Confusion Matrix :

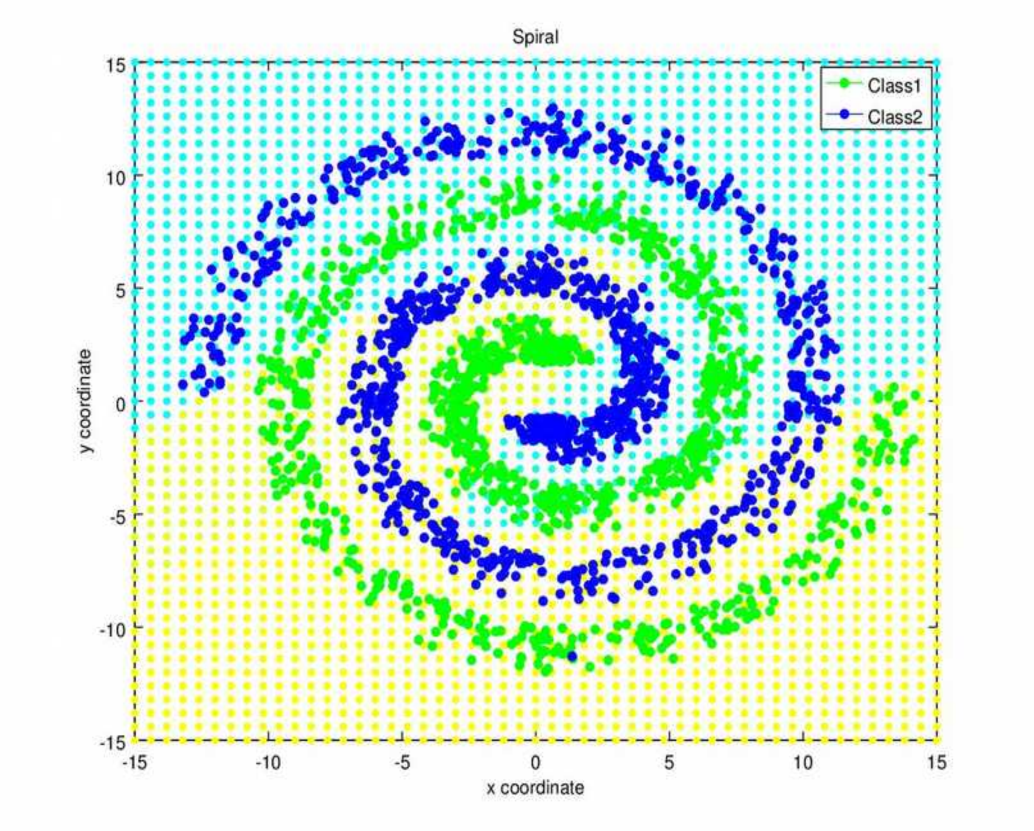
$$C = \begin{matrix} 75 & 0 \\ 0 & 300 \end{matrix}$$

Observations :

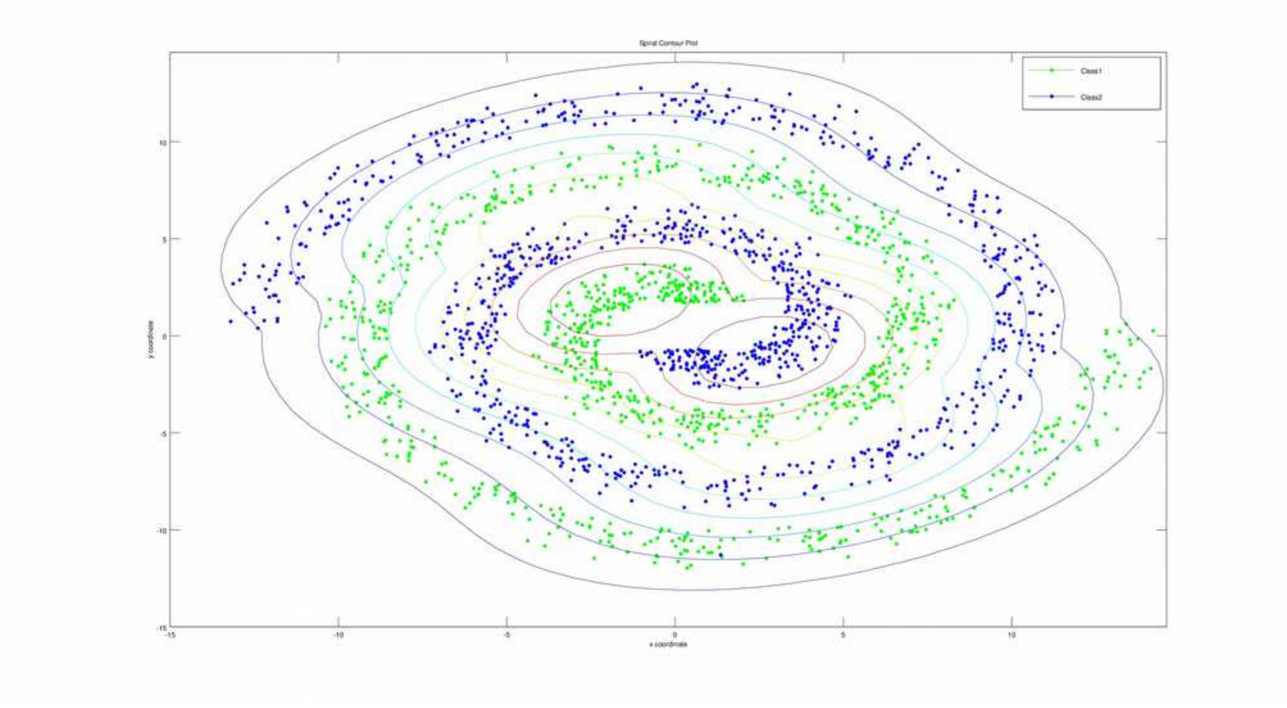
- In this case we have total of 8 gaussian distributions(4 per class) and 4 inner class gaussians define the decision boundary at their point of intersection with outer class contours.
- 100% accuracy is obtained for both values of K.
- The packing of data is more precise for K = 4 as compared to K = 2

3.2 Spiral dataset

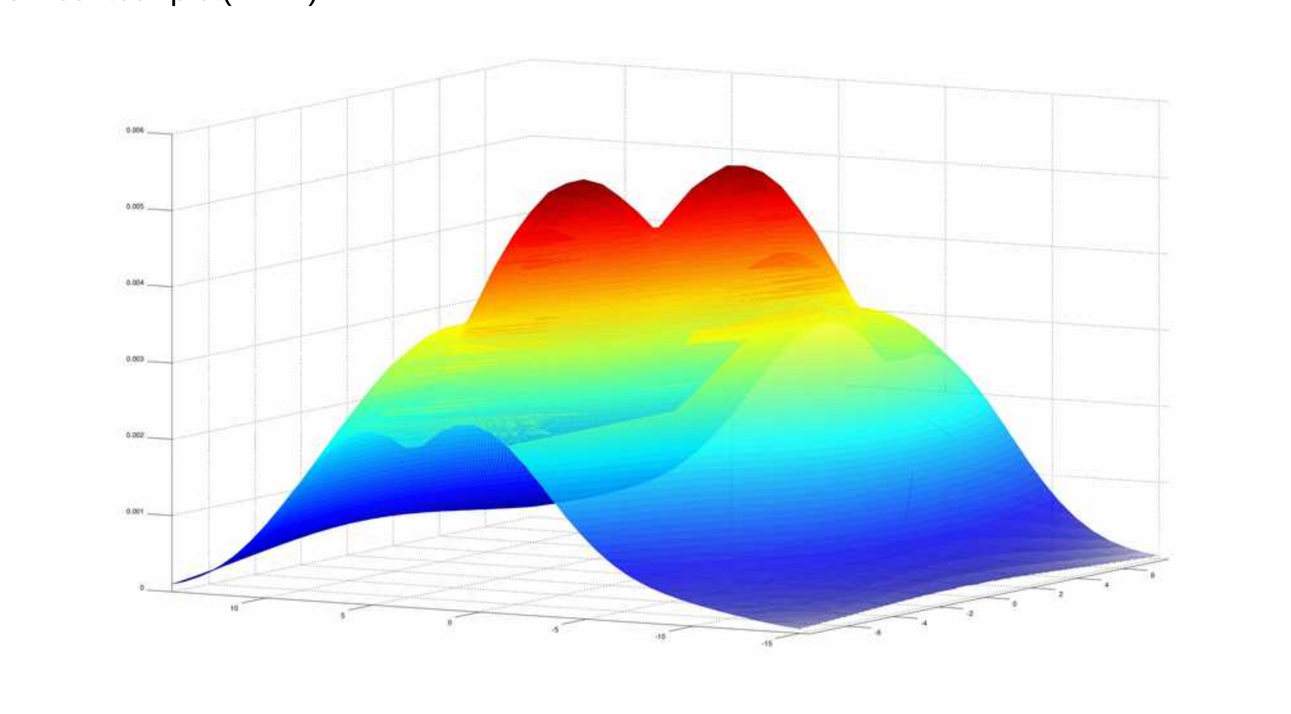
3.2.1 2 clusters



Contour Plot (K = 2):



3D contour plot(K = 2):



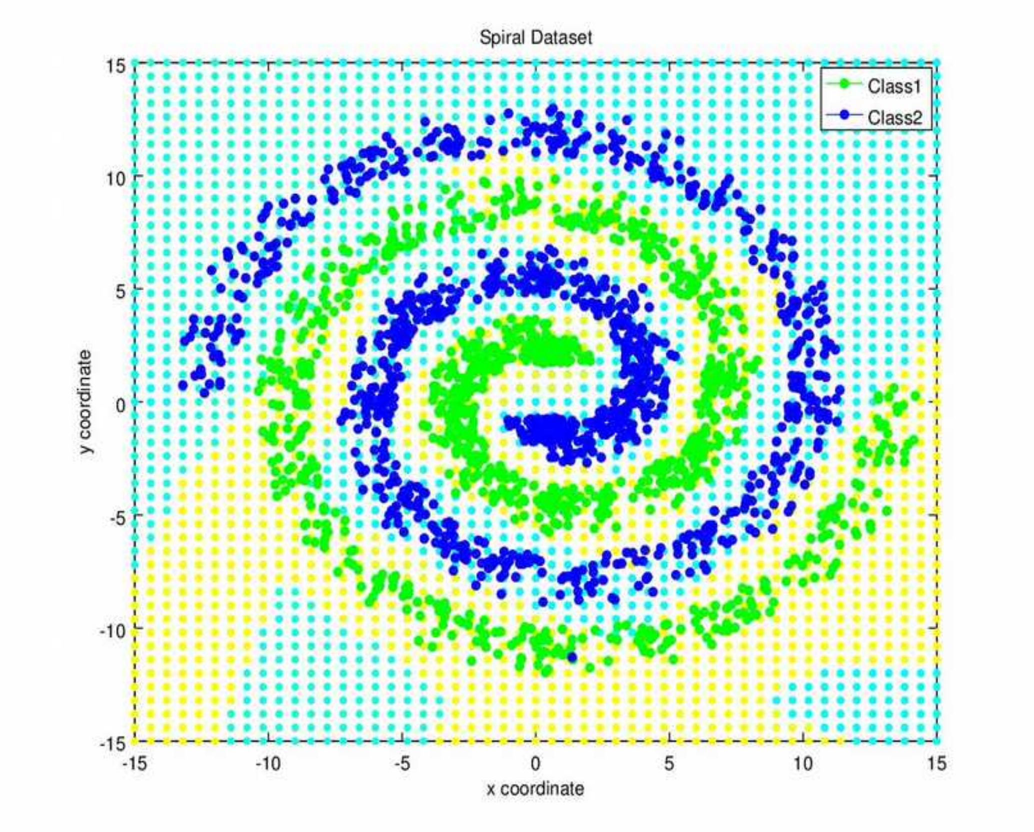
For K = 2

Classification Accuracy (%)	61.19%
Precision for Class1 Precision for Class2	0.6116 0.6123
Mean Precision	0.6119
Recall for Class1 Recall for Class2	0.6135 0.6104
Mean Recall	0.6119
F-measure for Class1 F-measure for Class2	0.6125 0.6113
Mean F-measure	0.6119

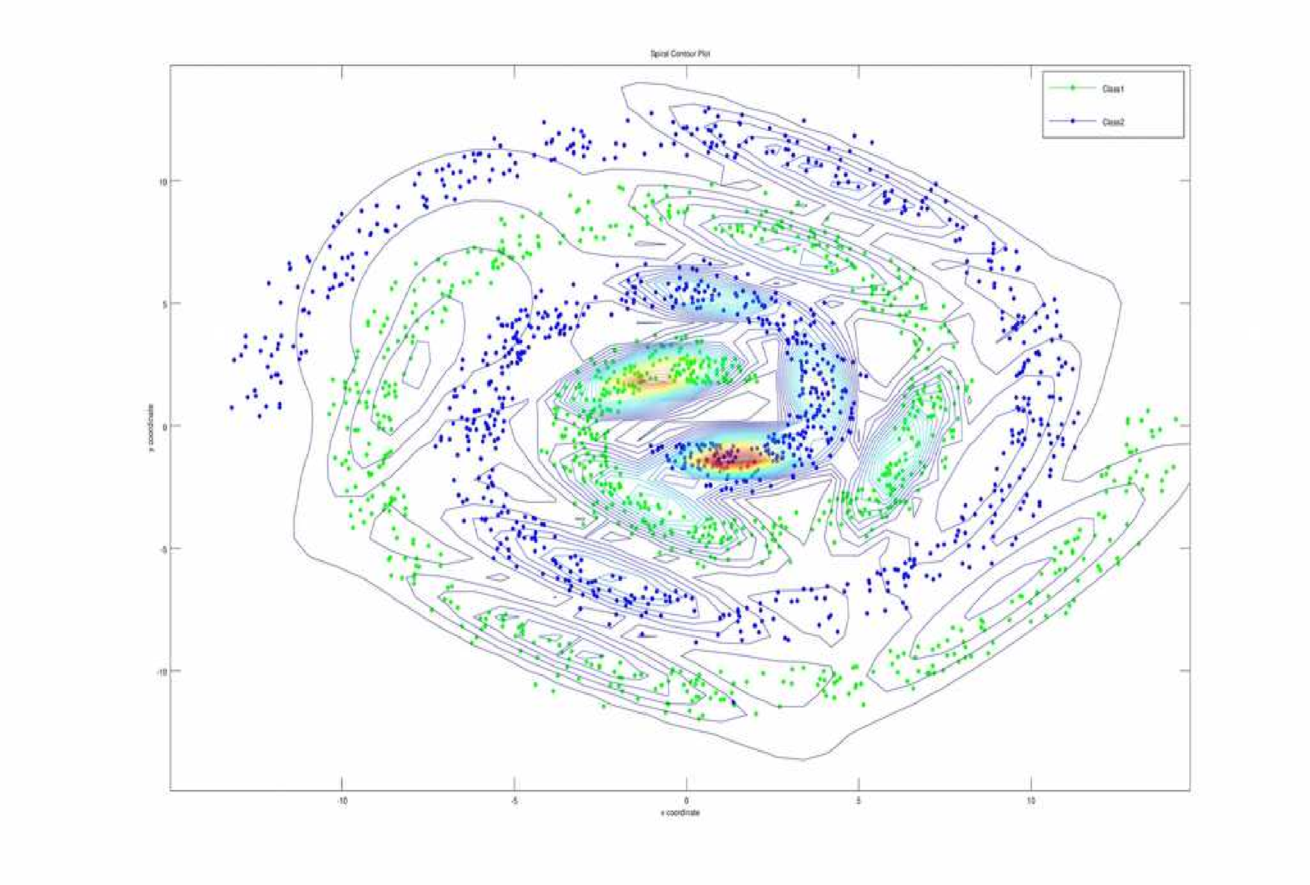
Confusion Matrix :

$$C = \begin{matrix} 200 & 126 \\ 127 & 199 \end{matrix}$$

3.2.2 7 Clusters



Contour Plot (K = 7):



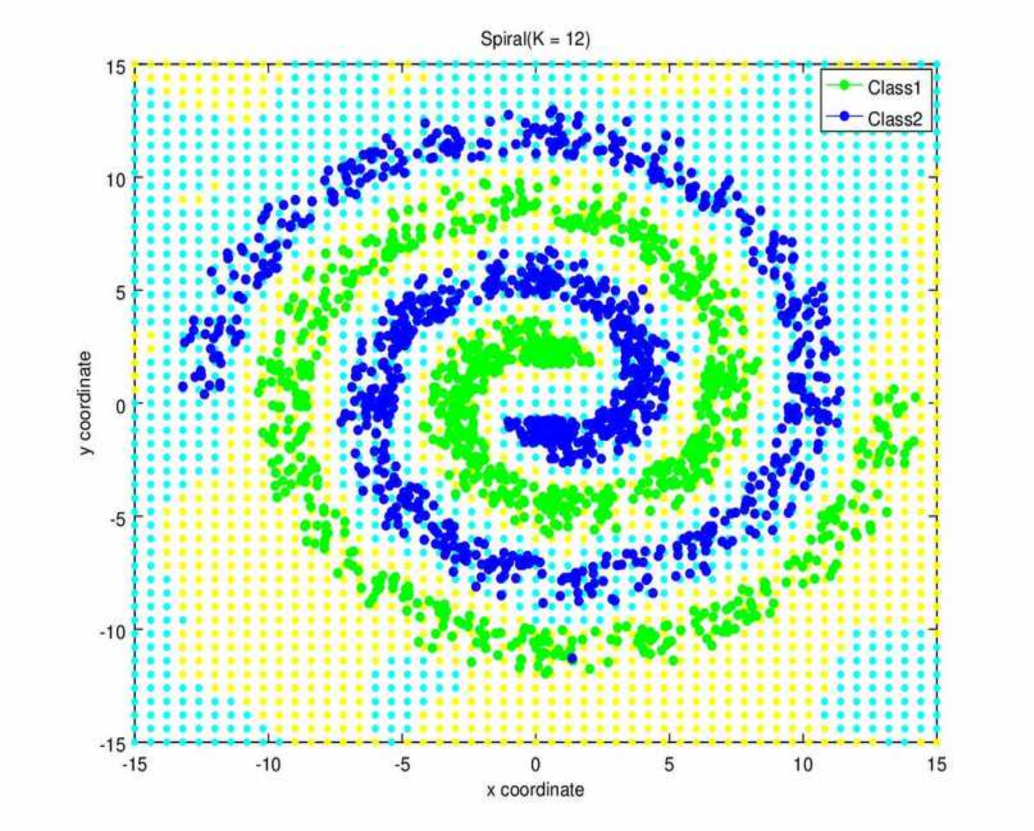
For K = 7 -

Classification Accuracy (%)	94.01%
Precision for Class1	0.9736
Precision for Class2	0.9111
Mean Precision	0.9423
Recall for Class1	0.9409
Recall for Class2	0.9754
Mean Recall	0.9401
F-measure for Class1	0.9380
F-measure for Class2	0.9422
Mean F-measure	0.9401

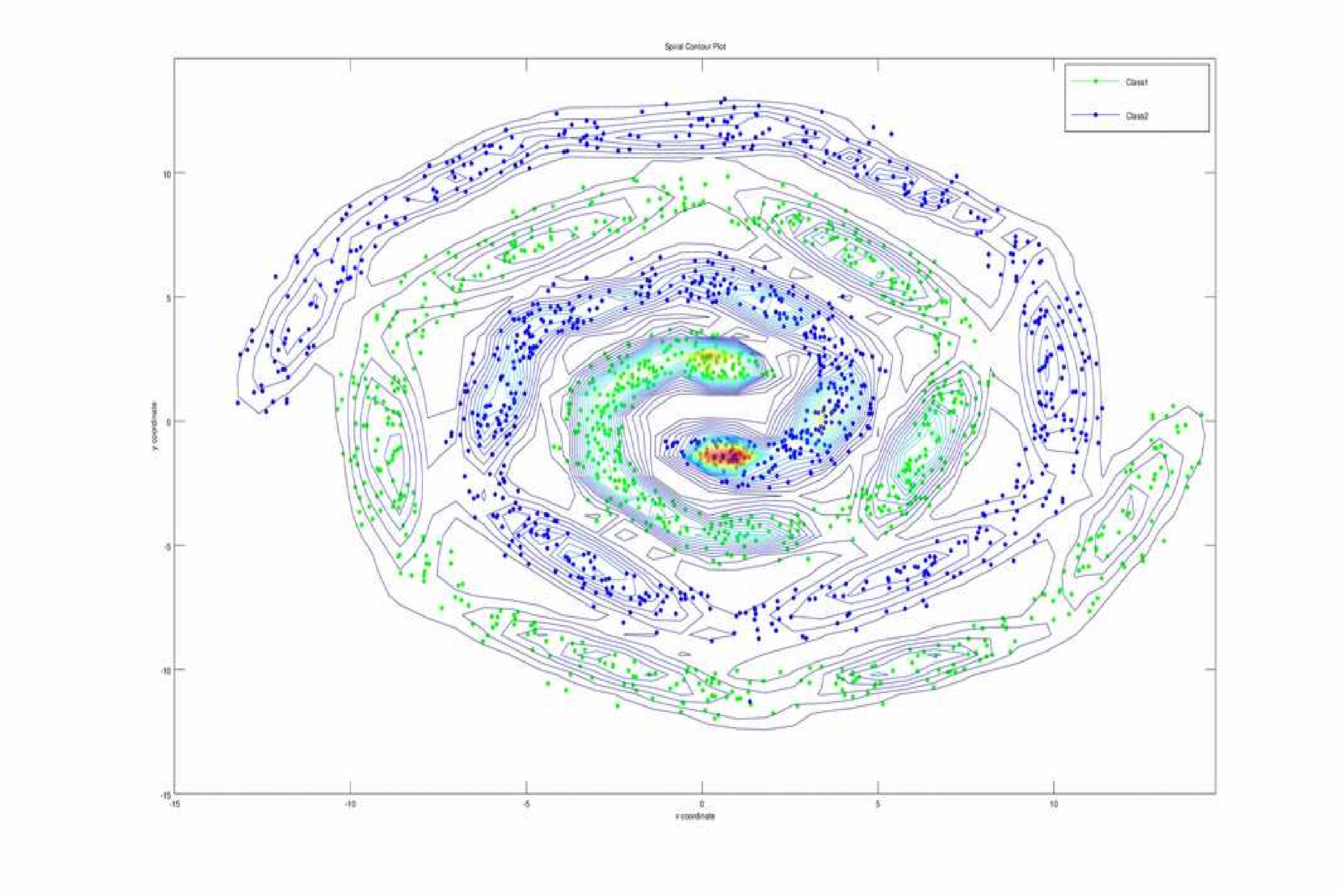
Confusion Matrix :

$$C = \begin{matrix} 295 & 31 \\ 8 & 318 \end{matrix}$$

3.2.3 12 clusters



Contour plot (K = 12) -



For K = 12 -

Classification Accuracy (%)	99.69%
Precision for Class1 Precision for Class2	0.9969 0.9969
Mean Precision	0.9969
Recall for Class1 Recall for Class2	0.9969 0.9969
Mean Recall	0.9969
F-measure for Class1 F-measure for Class2	0.9969 0.9969
Mean F-measure	0.9969

Confusion Matrix :

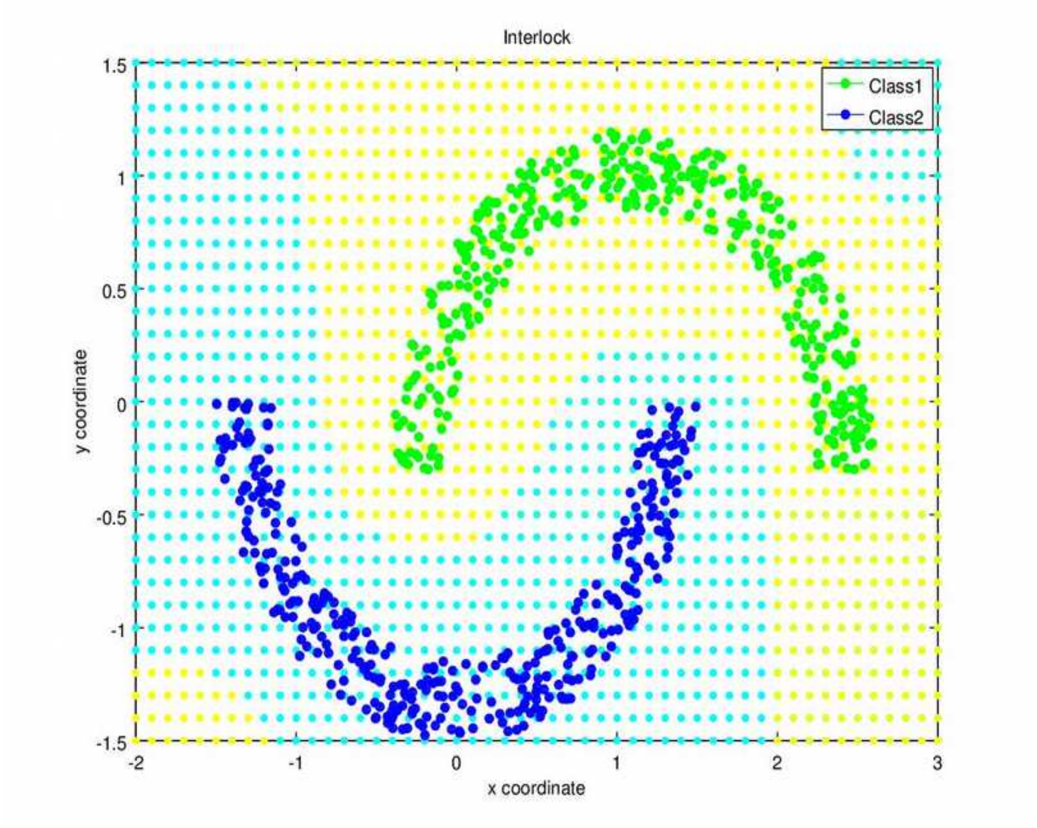
$$C = \begin{matrix} 325 & 1 \\ 1 & 325 \end{matrix}$$

Observations :

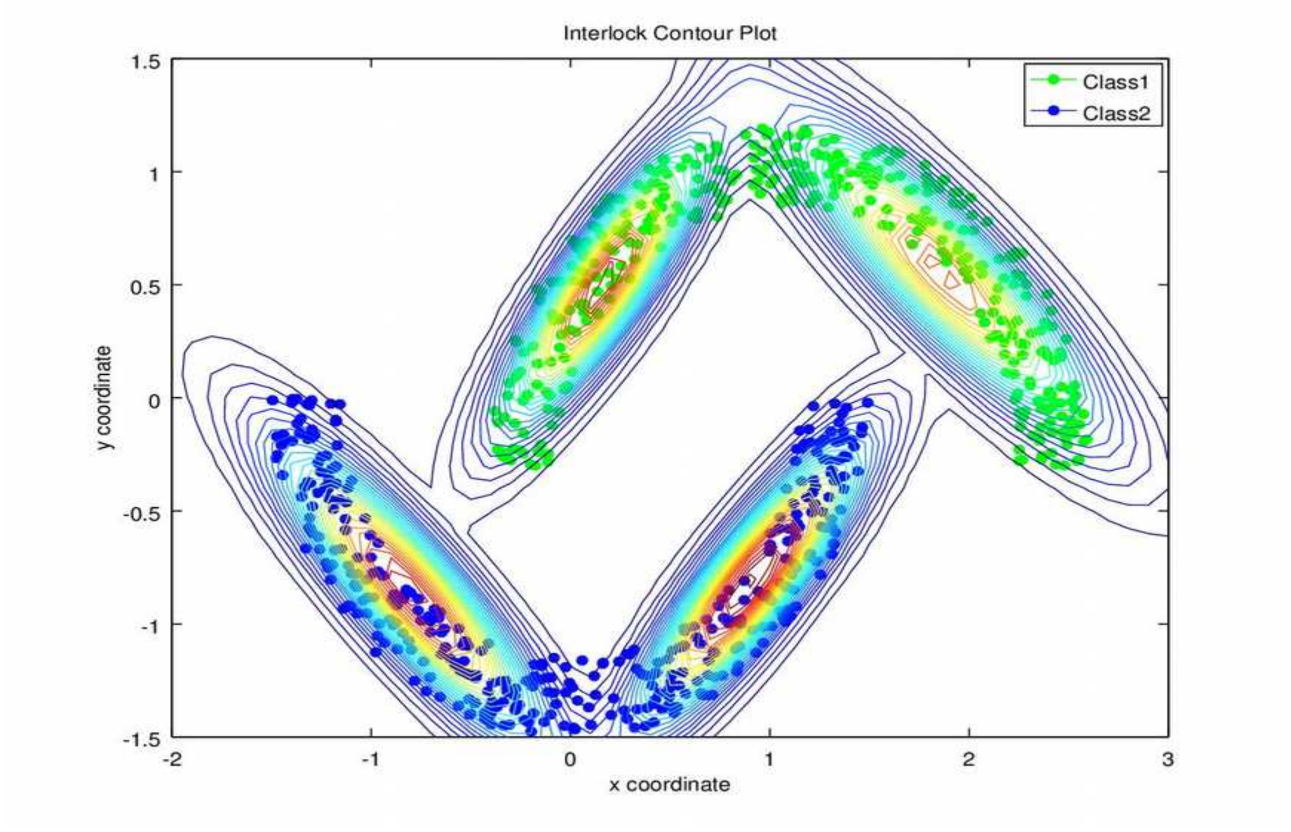
- The accuracy value increases with K and the maximum accuracy obtained is 99.69%.
- The accuracy obtained in this case is far better than that obtained in unimodal gaussian model(53.7%).
- Spiral Dataset requires a very complex decision boundary for classification. In GMM we assume that every class has several gaussian clusters which leads to a non linear envelope of the constituent gaussian components. This explains the vast difference in the accuracies obtained in unimodal case and GMM.
- From the contours plot we see that the packing of data in K = 12 clusters per class is more precise than that in K = 7.

3.3 Interlock

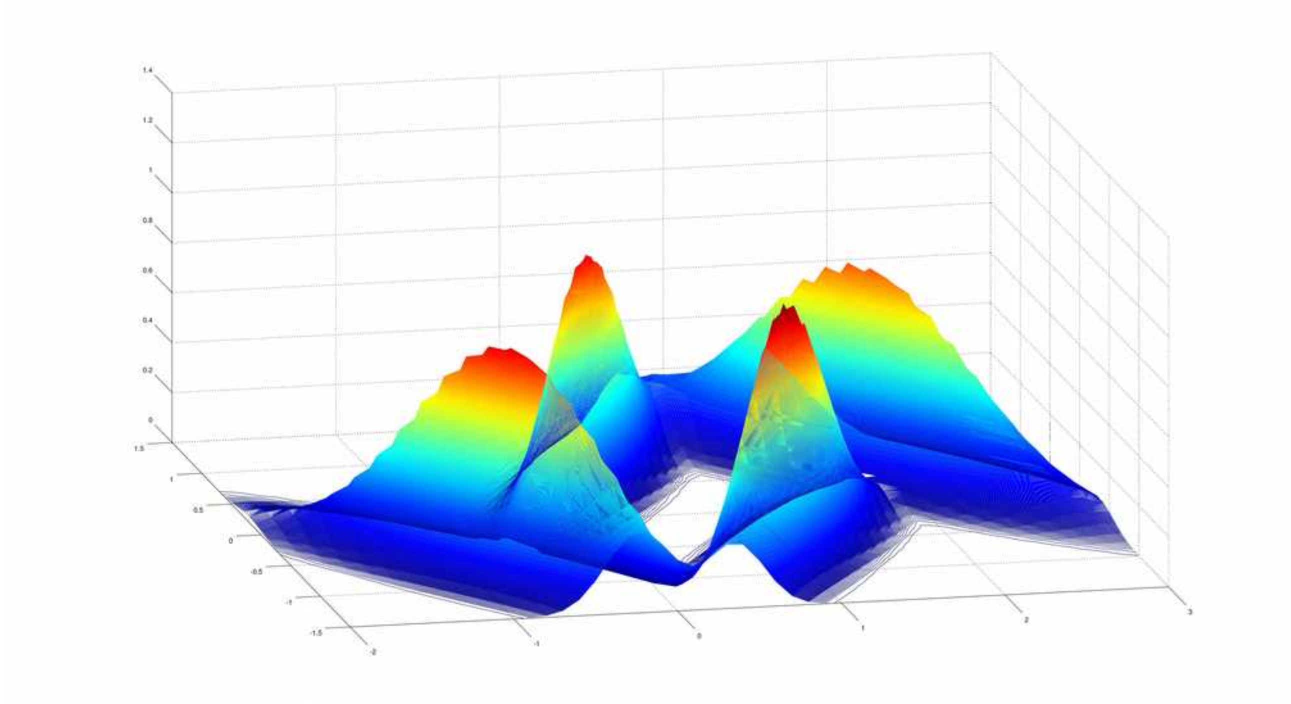
3.3.1 2 Clusters



Contour Plot (K = 2):



3D Contour Plot (K = 2):



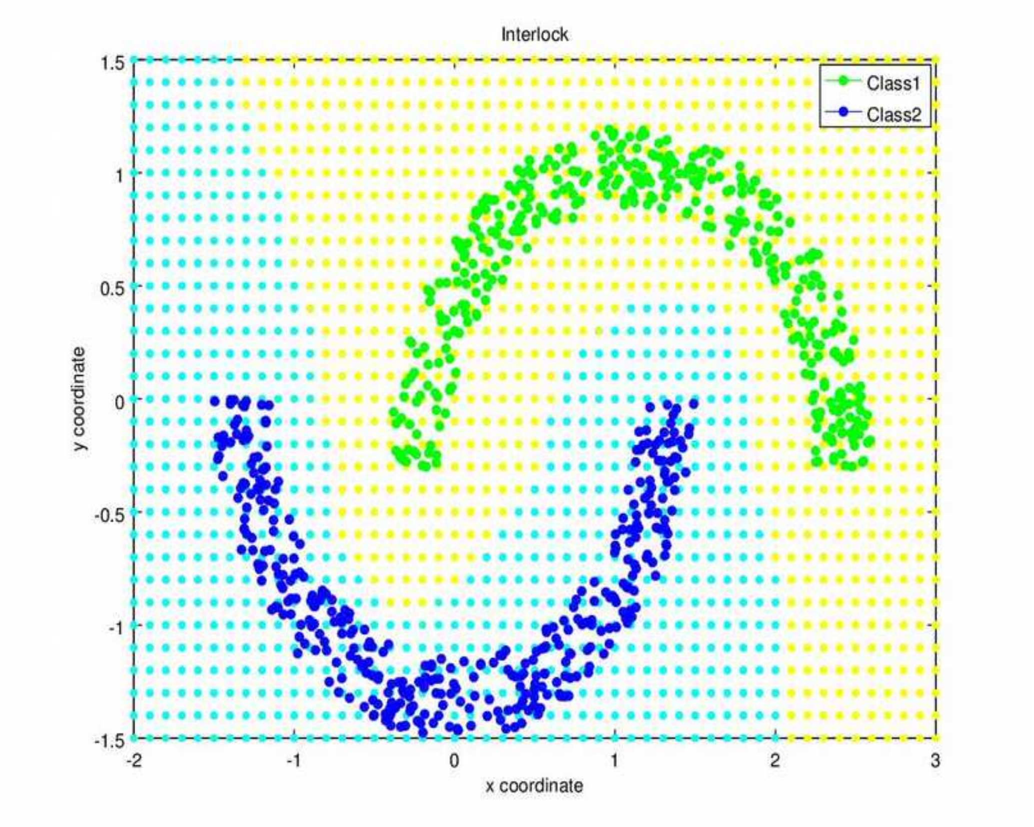
For K = 2 and 4 -

Classification Accuracy (%)	100%
Precision for Class1 Precision for Class2	1 1
Mean Precision	1
Recall for Class1 Recall for Class2	1 1
Mean Recall	1
F-measure for Class1 F-measure for Class2	1 1
Mean F-measure	1

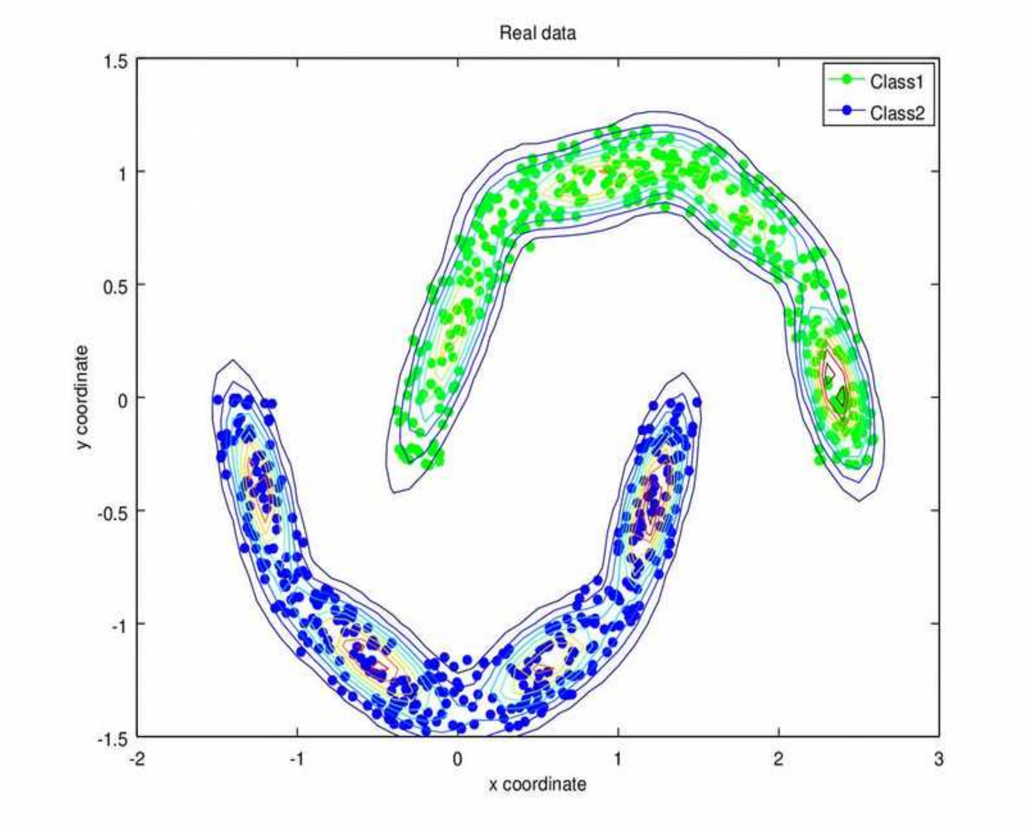
Confusion Matrix :

$$C = \begin{matrix} 125 & 0 \\ 0 & 125 \end{matrix}$$

3.3.2 4 Clusters



Contour Plot (K = 4) :

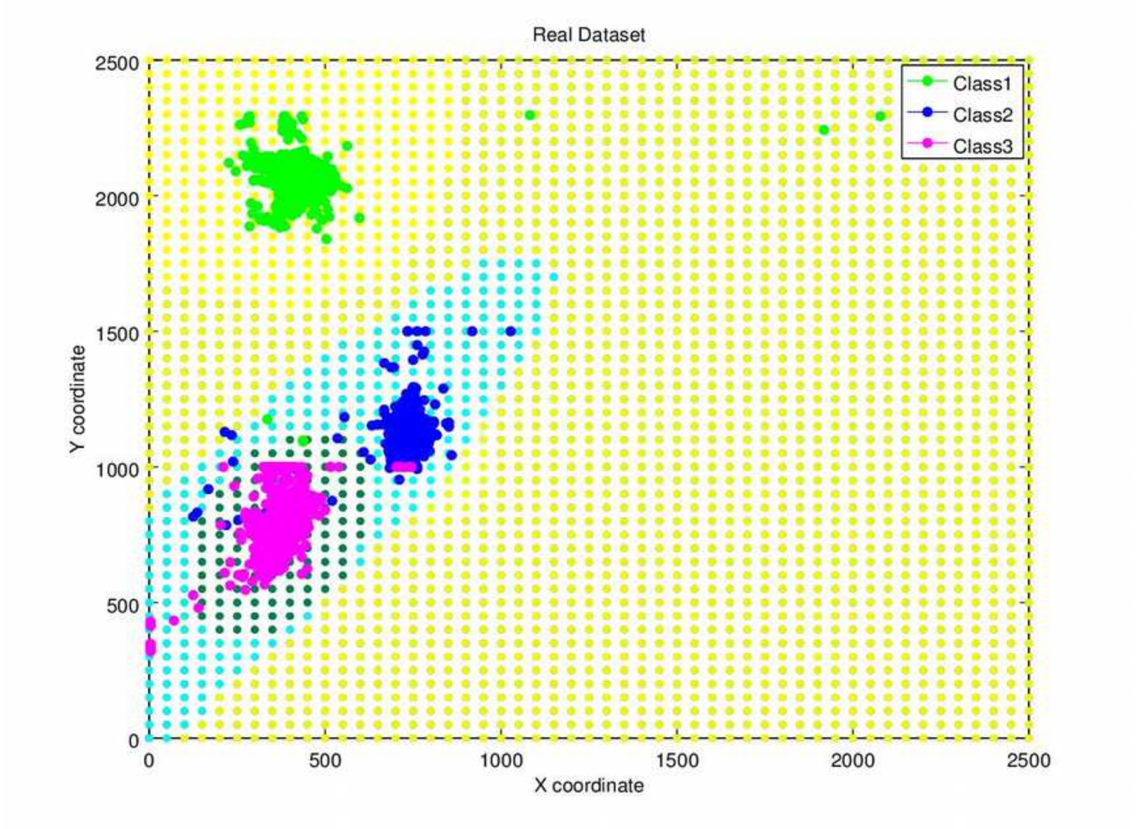


Observations:

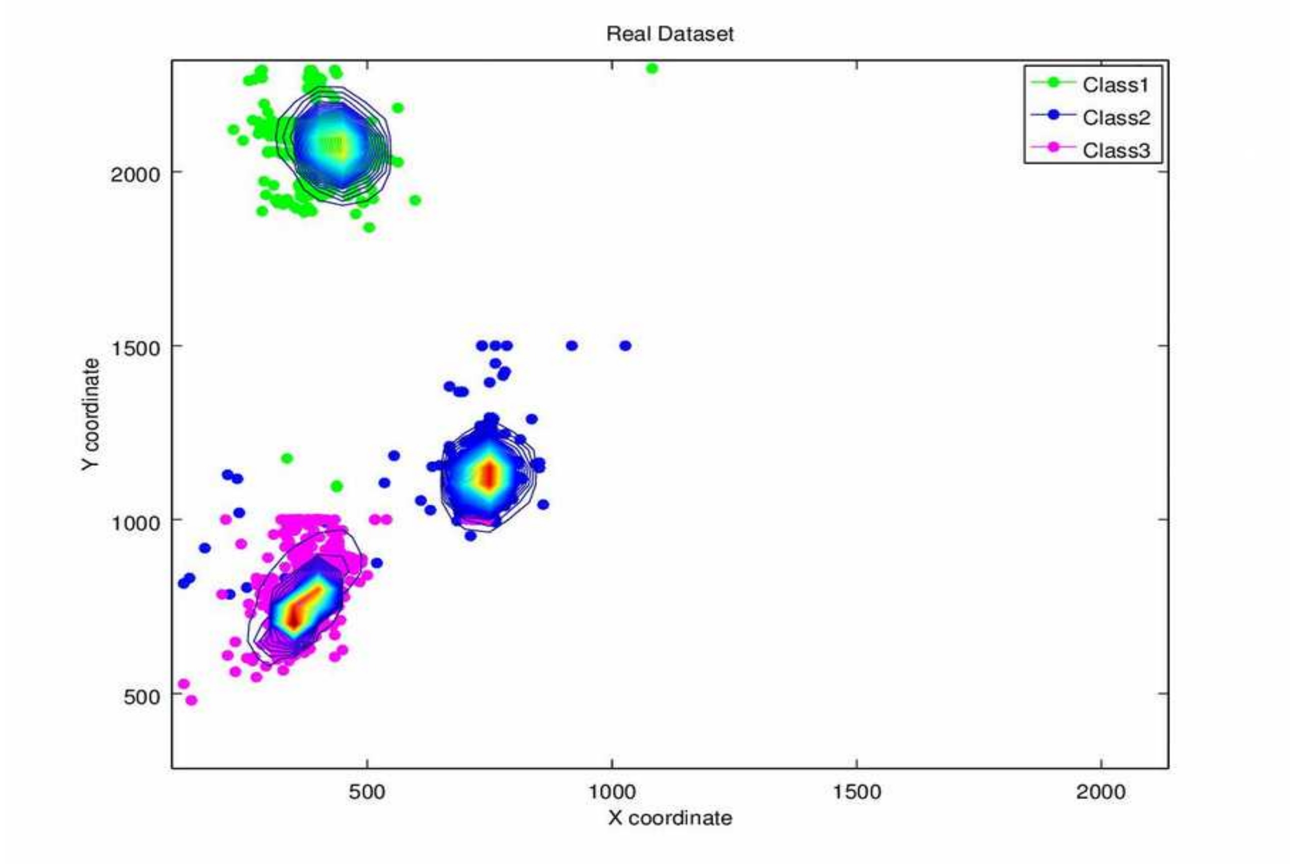
- Classification accuracy for K = 2 and K = 4 is 100% but from the contour plot of K= 4, we see that K = 4 has more precise and tighter packing than K = 2.
- The data of both the classes is similar and hence the contours are also similar in height and relative position.

3.4 Real Data set:

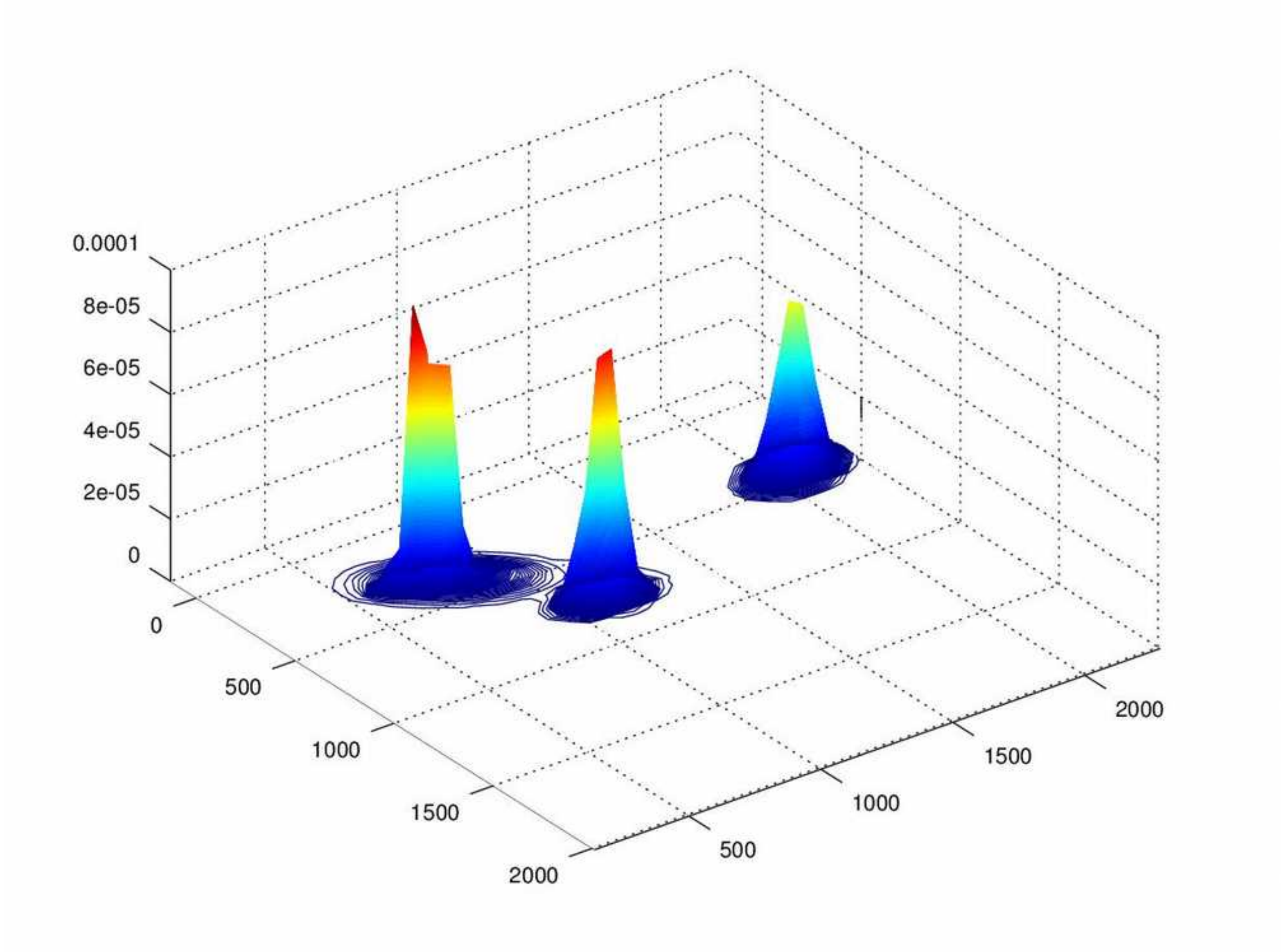
3.4.1 2 Clusters



Contour plot (K = 2):



3D Contour plot (k = 2):



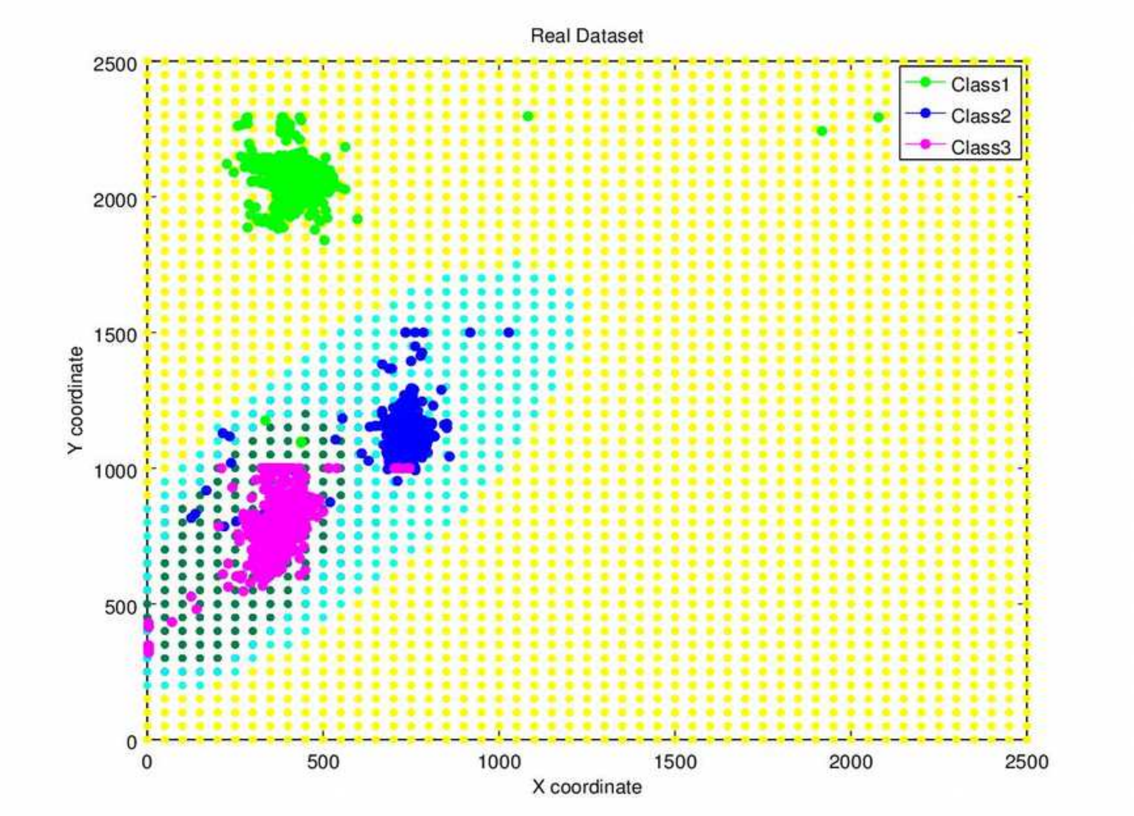
For K = 2,

Classification Accuracy (%)	96.60%
Precision for Class1 Precision for Class2 Precision for Class3	0.9893 0.9558 0.9538
Mean Precision	0.9663
Recall for Class1 Recall for Class2 Recall for Class3	0.9738 0.9611 0.9630
Mean Recall	0.9660
F-measure for Class1 F-measure for Class2 F-measure for Class3	0.9815 0.9585 0.9584
Mean F-measure	0.9661

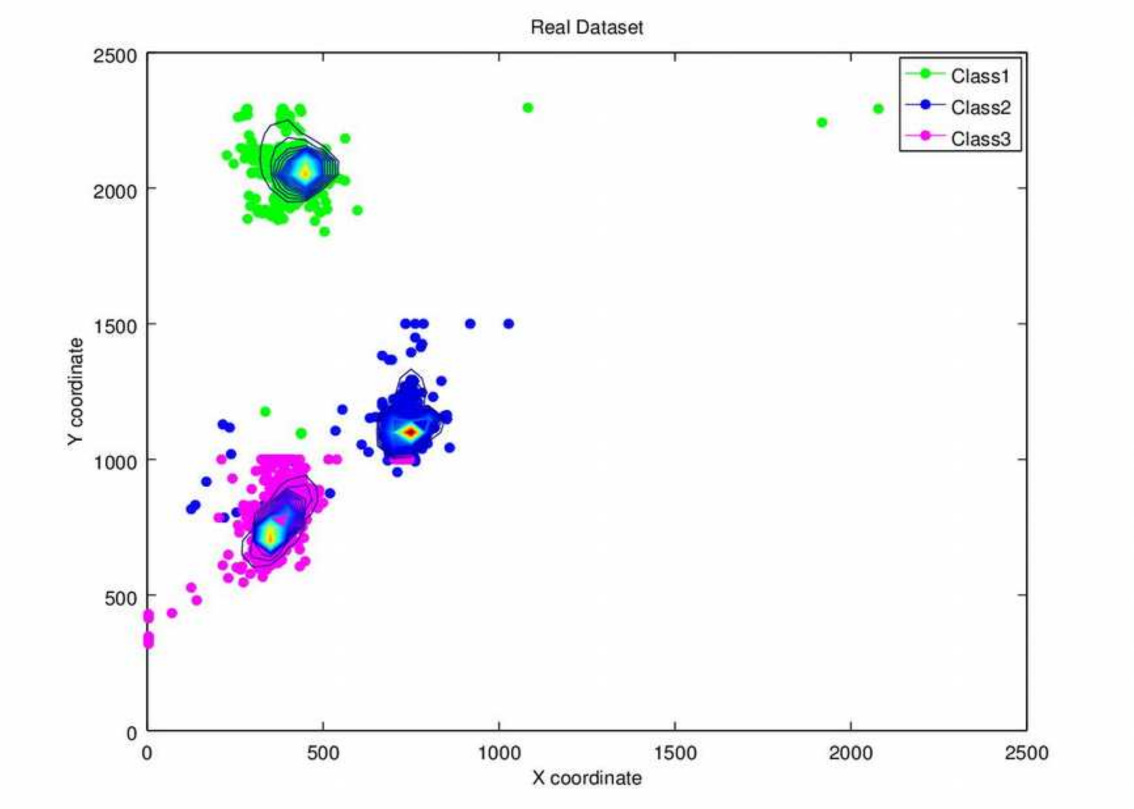
Confusion Matrix :

$$C = \begin{matrix} 558 & 5 & 10 \\ 2 & 520 & 19 \\ 4 & 19 & 599 \end{matrix}$$

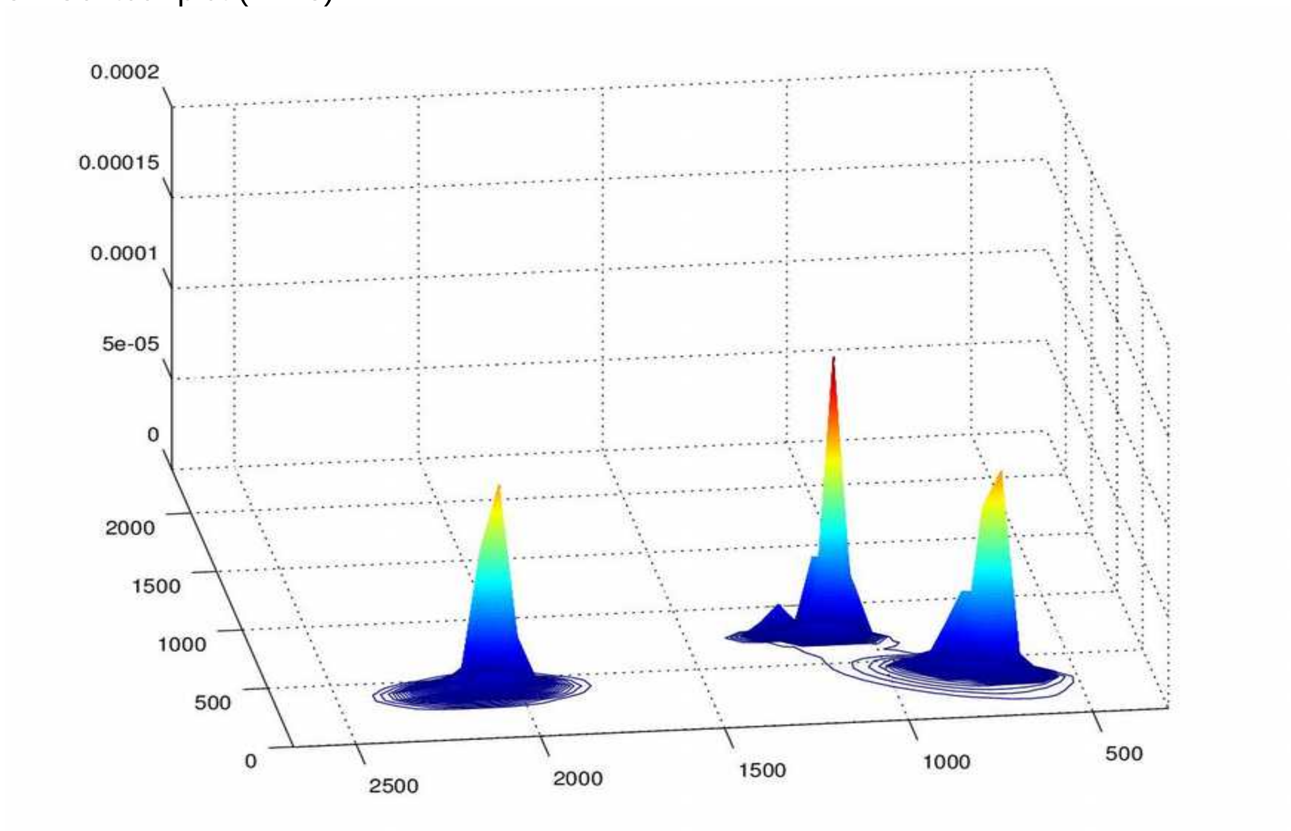
3.4.2 3 Clusters



Contour Plot (K = 3):



3D Contour plot (K = 3):



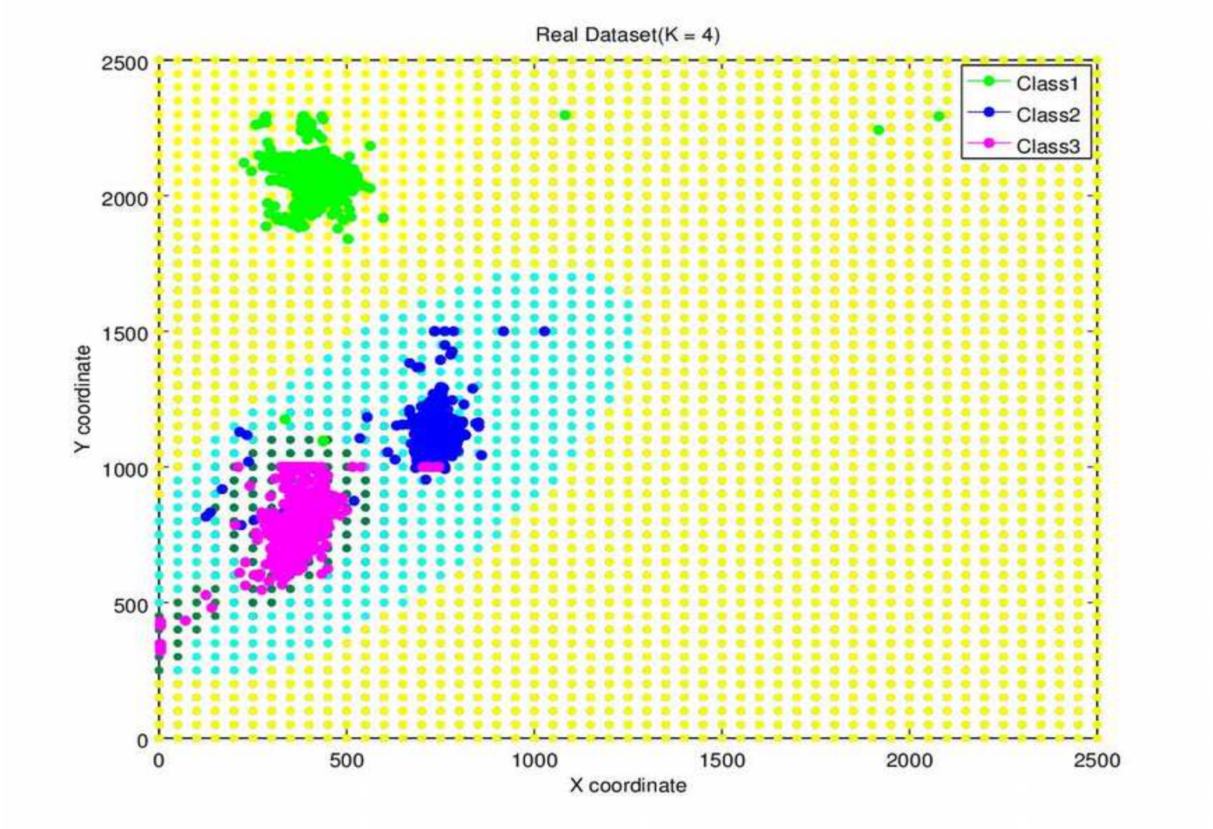
For K = 3,

Classification Accuracy (%)	96.71%
Precision for Class1 Precision for Class2 Precision for Class3	0.9928 0.9715 0.9411
Mean Precision	0.9685
Recall for Class1 Recall for Class2 Recall for Class3	0.9738 0.9482 0.9774
Mean Recall	0.9665
F-measure for Class1 F-measure for Class2 F-measure for Class3	0.9832 0.9597 0.9589
Mean F-measure	0.9673

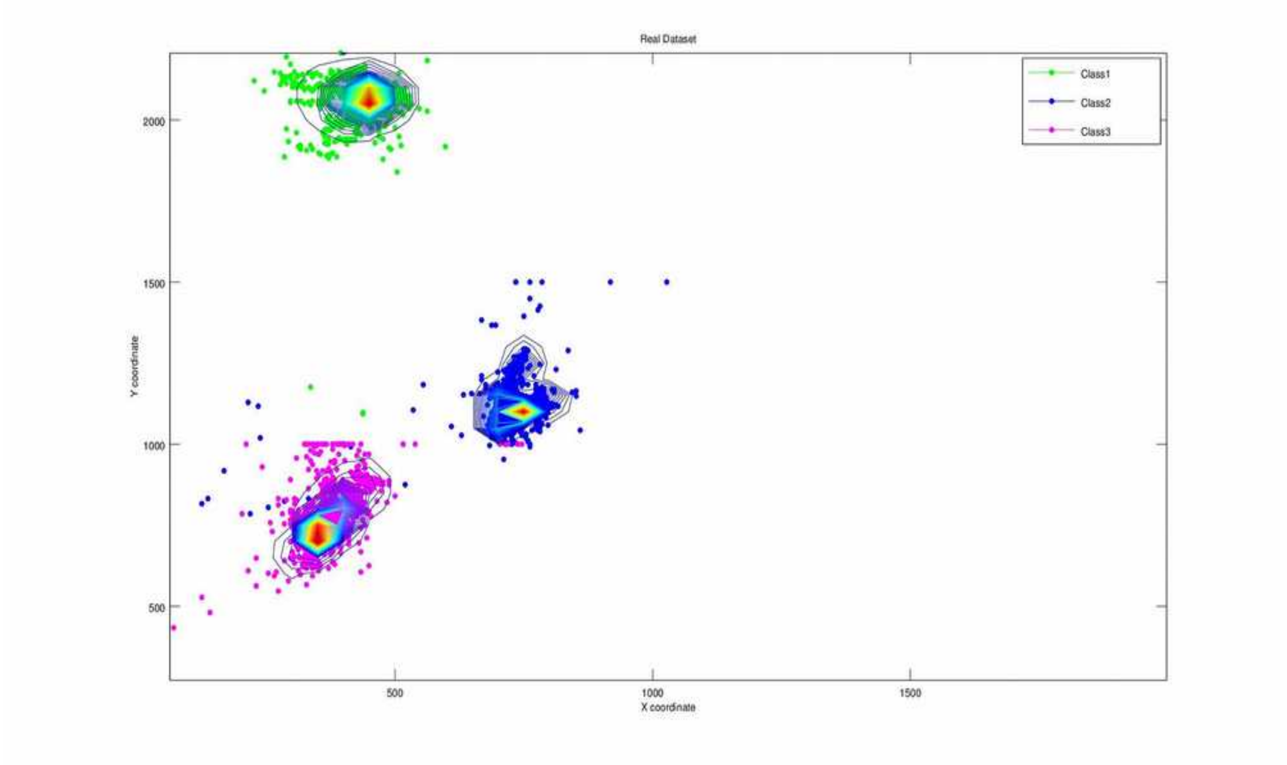
Confusion Matrix :

$$C = \begin{matrix} 558 & 3 & 12 \\ 2 & 513 & 26 \\ 2 & 12 & 608 \end{matrix}$$

3.4.3 4 Clusters



Contour Plot(K = 4):



For K = 4 -

Classifiaction Accuracy (%)	96.42%
Precision for Class1 Precision for Class2 Precision for Class3	0.9928 0.9772 0.9287
Mean Precision	0.9663
Recall for Class1 Recall for Class2 Recall for Class3	0.9738 0.9537 0.9646
Mean Recall	0.9640
F-measure for Class1 F-measure for Class2 F-measure for Class3	0.9832 0.9653 0.9463
Mean F-measure	0.9650

Confusion Matrix :

$$C = \begin{matrix} 558 & 0 & 15 \\ 2 & 516 & 23 \\ 2 & 20 & 600 \end{matrix}$$

Observations :

- From K = 3 to K = 4 over clustering leads to less accuracy as more data points get misclassified.
- Maximum acuracy in unimodal gaussian case is more or less equal to the accuracy obtained in GMM with 3 components.

3.5 Image dataset :

Here,
Class1 = coast dataset
Class2 = mountain dataset
Class3 = insidecity dataset

3.5.1 2 clusters:

Confusion Matrix :

$$C = \begin{matrix} 81 & 7 & 2 \\ 12 & 75 & 7 \\ 6 & 5 & 66 \end{matrix}$$

Images correctly classified = 222
Images incorrectly classified = 39
Classfication accuracy = 85.05 %

3.5.2 3 clusters:

Confusion Matrix :

$$C = \begin{matrix} 79 & 8 & 3 \\ 11 & 76 & 7 \\ 5 & 9 & 63 \end{matrix}$$

Images correctly classified = 218
Images incorrectly classified = 43
Classfication accuracy = 83.52 %

3.5.3 4 clusters:

Confusion Matrix :

$$C = \begin{matrix} 80 & 6 & 4 \\ 6 & 85 & 3 \\ 7 & 10 & 60 \end{matrix}$$

Images correctly classified = 225
Images incorrectly classified = 36
Classfication accuracy = 86.20 %

3.5.4 8 clusters:

Confusion Matrix :

$$C = \begin{matrix} 84 & 4 & 2 \\ 12 & 79 & 3 \\ 6 & 16 & 55 \end{matrix}$$

Images correctly classified = 218
Images incorrectly classified = 43
Classfication accuracy = 83.52 %

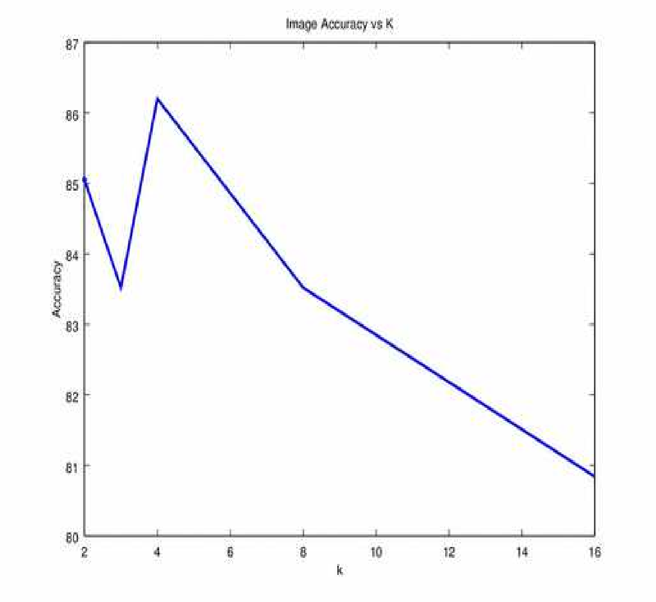
3.5.5 16 clusters:

Confusion Matrix :

$$C = \begin{matrix} 76 & 13 & 1 \\ 13 & 80 & 1 \\ 8 & 14 & 55 \end{matrix}$$

Images correctly classified = 211
Images incorrectly classified = 50
Classfication accuracy = 80.84 %

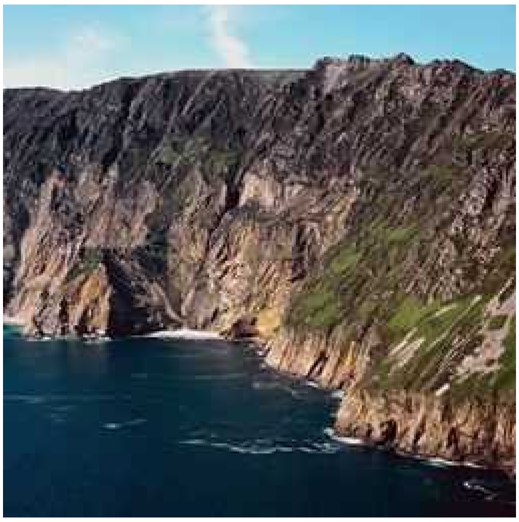
Graph of K vs Classification accuracy -



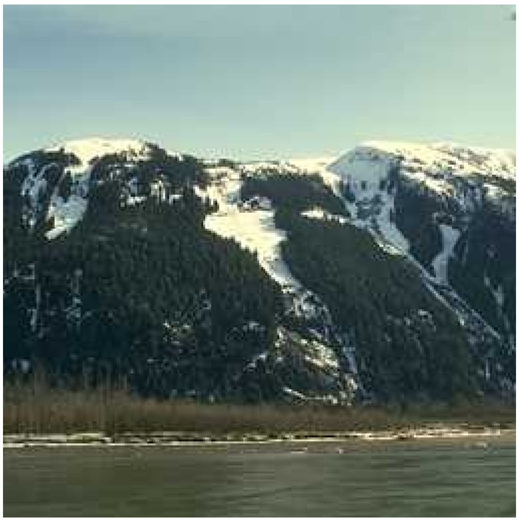
Observations:

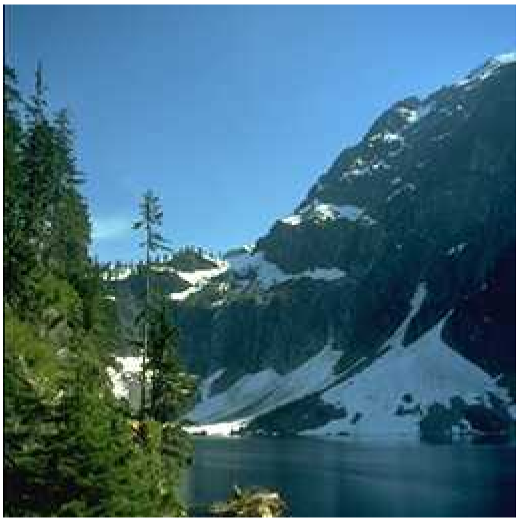
- The maximum accuracy is obtained at K = 4 clusters per class and decreases thereafter.
- The classification accuracy of the image dataset is highly contingent on the feature extraction process. Every image is represented by 36 fixed size blocks and each block was a feature vector of 23 components.

- The following image from coast dataset got misclassified into mounatin class due to similarity in features.

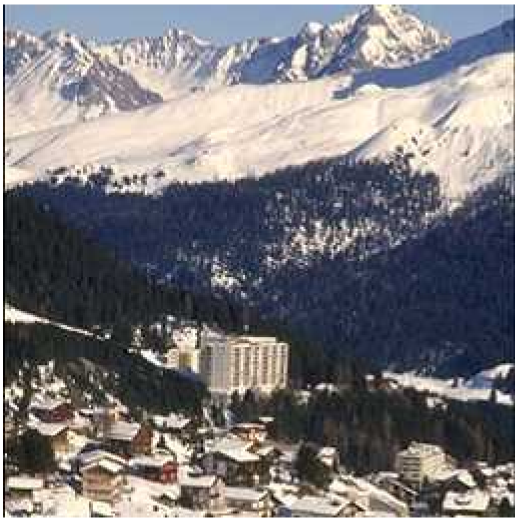


- The following images from mountain dataset got misclassified into coast class.



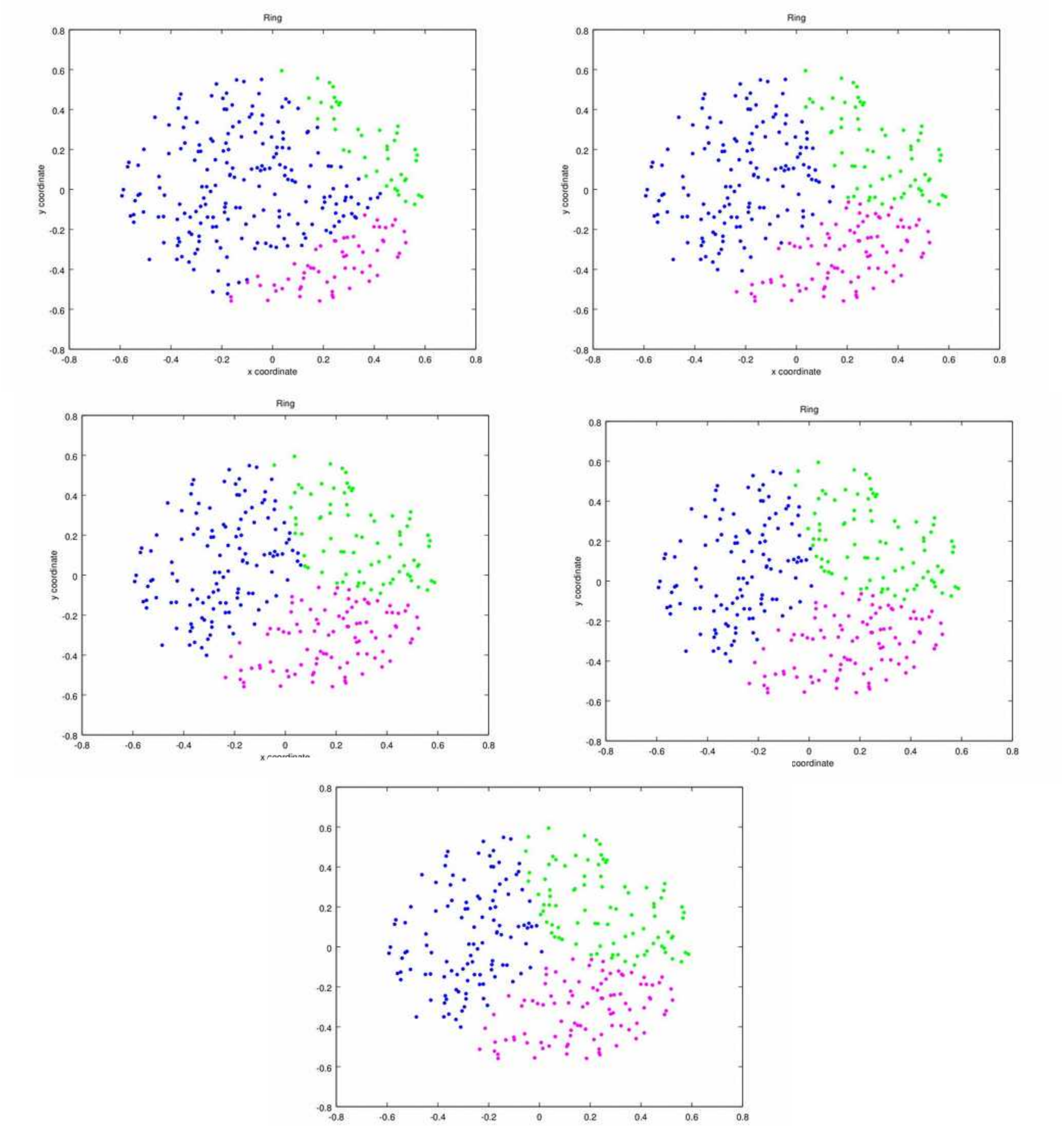


- The following image from mountain dataset got misclassified into insidecity class.



4.0 Additional Graphs

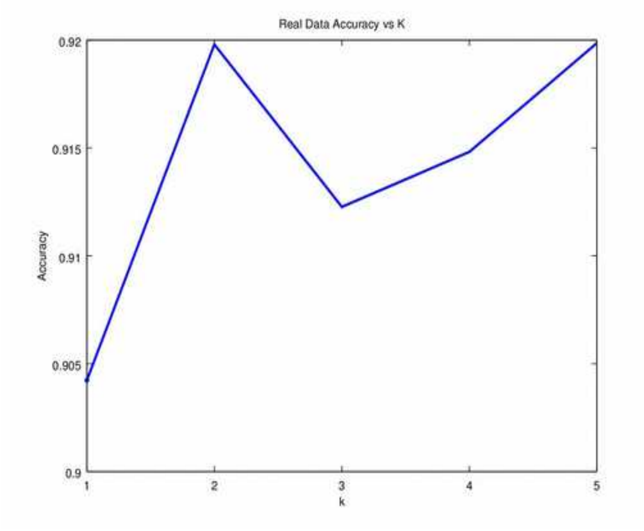
4.1 Simulation of K means clustering for Ring Class1 data with K = 3

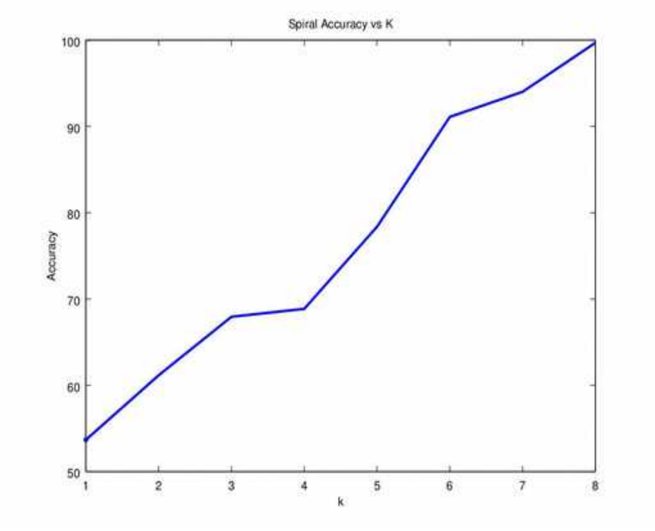


Observations :

- As the number of iterations increase, cluster assignment of the data points changes and approaches convergence.
- The boundary between 2 clusters is linear in the above graphs because distance measure used in Kmeans classification is eucledian distance.

4.2 Graph of accuracy vs K

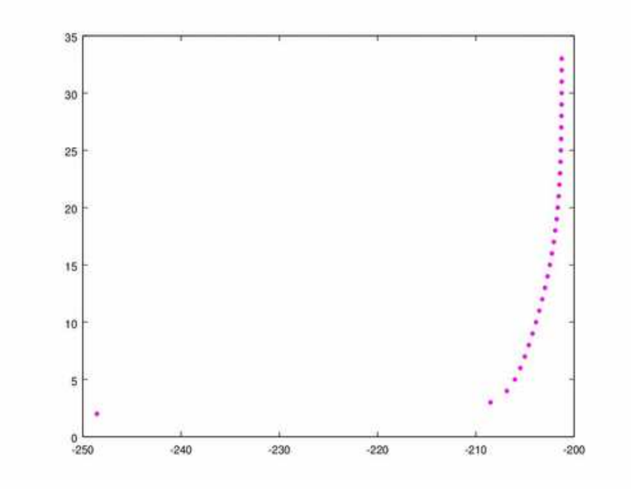




Observations :

- In general the accuracy increases with K till it reaches maxima and decreases thereafter.

4.3 Graph of log likelihood vs number of iterations in EM method (ML parameter estimation)



X axis – log likelihood
Y axis – number of iterations
Dataset – Interlock

Observations :

- The value of log likelihood increases with the number of iterations till it reaches the local / global optima.

5.0 Conclusion:

- The decision boundaries are more precise when the data is modelled using mixture of multiple gaussians as compared to unimodal gaussian.
- Different classification accuracies are obtained when number of clusters K are varied.
- Although the accuracy tends to increase with the number of clusters assumed for a class, but due to overlapping data, over clustering may cause the class to cover non-belonging points as well, this is evident in the real world data with 4 and 2 clusters.
- The decision boundaries appear to be piece wise combinations of quadratic boundaries as in case of unimodal gaussian.
- The boundary between the clusters in K means clustering depends on the distance measure used and this is evident from the graphs in section 4.1.
- The accuracies obtained for the nonlinear datasets are far better than that obtained in unimodal case. Acccuracy for Real dataset is more or less same for both the cases.

6.0 References

- Pattern Classification, Duda, Hart, Stork
- http://www.ee.iisc.ac.in/people/faculty/prasantg/downloads/GMM_Tutorial_Reynolds.pdf