PATTERN RECOGNITION

ASSIGNMENT 1

MEGHANA VENKATA PALUKURI
CH12B083

RAJAT DESHMUKH
EP13B006

ABHISHIKTH MALLAMPALLI
EP13B002

CONTENTS

1. DEFINITIONS

- Classification accuracy (overall) = $\frac{\text{Number of correctly classified examples}}{Total\ number\ of\ examples}$

- Classification accuracy (class) = $\frac{\text{Number of correctly classified examples for the class}}{Total\ number\ of\ examples\ belonging\ to\ the\ class}$

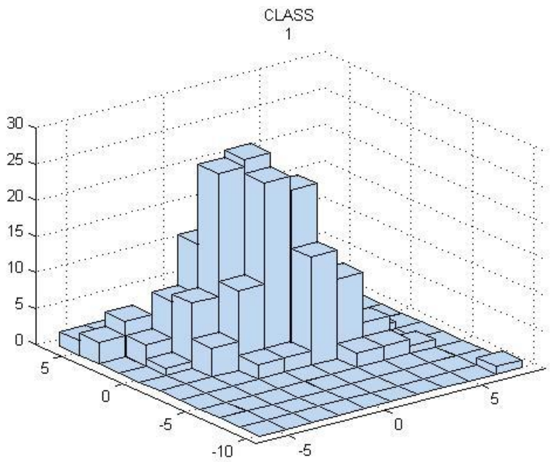
  (indicated by the diagonal elements of the confusion matrix)
- Confusion matrix: A(i,j) = Number of examples classified as j belonging to class i

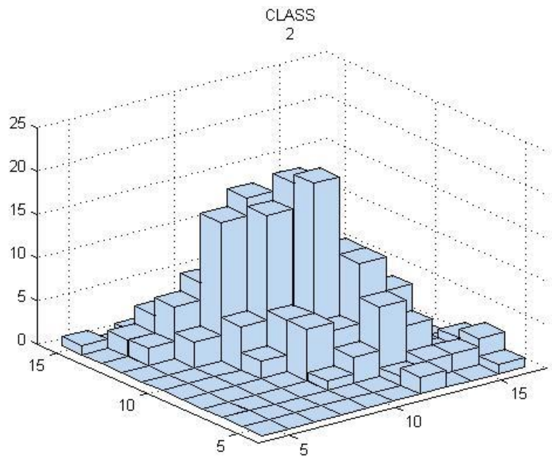
DATASET 1(Static Pattern Classification)

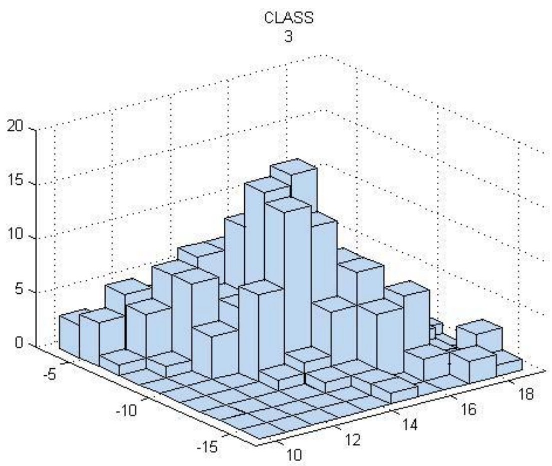
2. LINEARLY SEPARABLE DATA

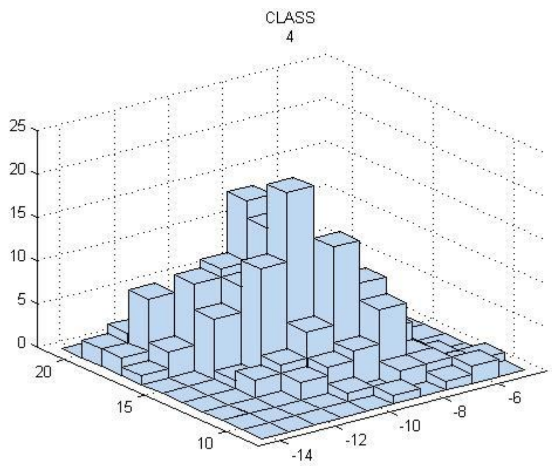
Since the data is 2-dimensional, histograms can be plotted for each class to get an idea of the distribution of the data (Refer training data plots of each class below). We find that distribution of each class is close to a Bivariate Gaussian Distribution, and hence proceed to find the parameters for this distribution for each class in the Naïve Bayes & Bayes Classifiers. MLE estimates are used, i.e sample mean & sample covariance. Priors are assumed to be same for all classes.

There are 4 classes in this dataset, with 2-dimensional feature vectors. We are given 250 training data points, 150 validation data points and 100 test data points. Histograms for the training data of each class are given below:



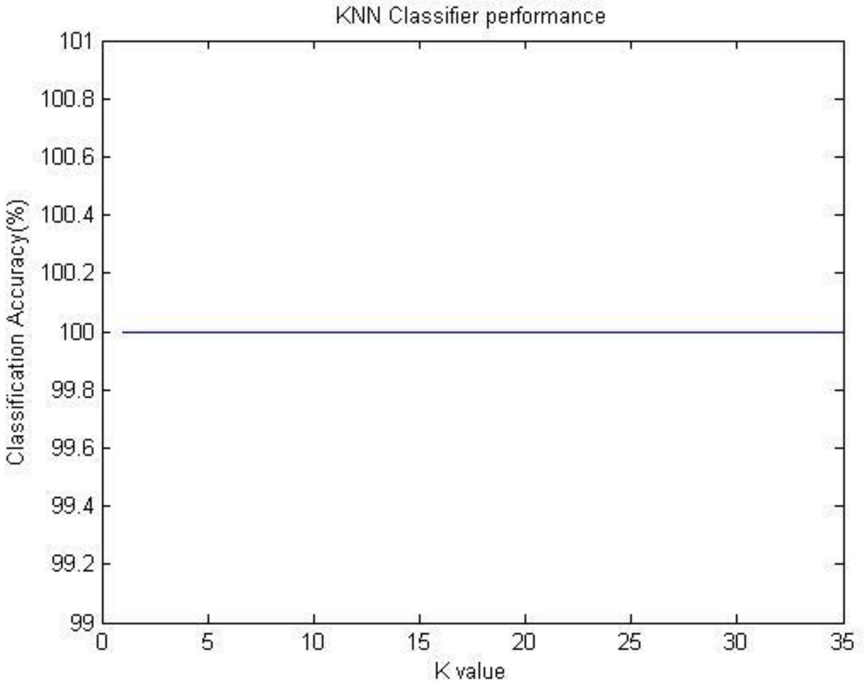




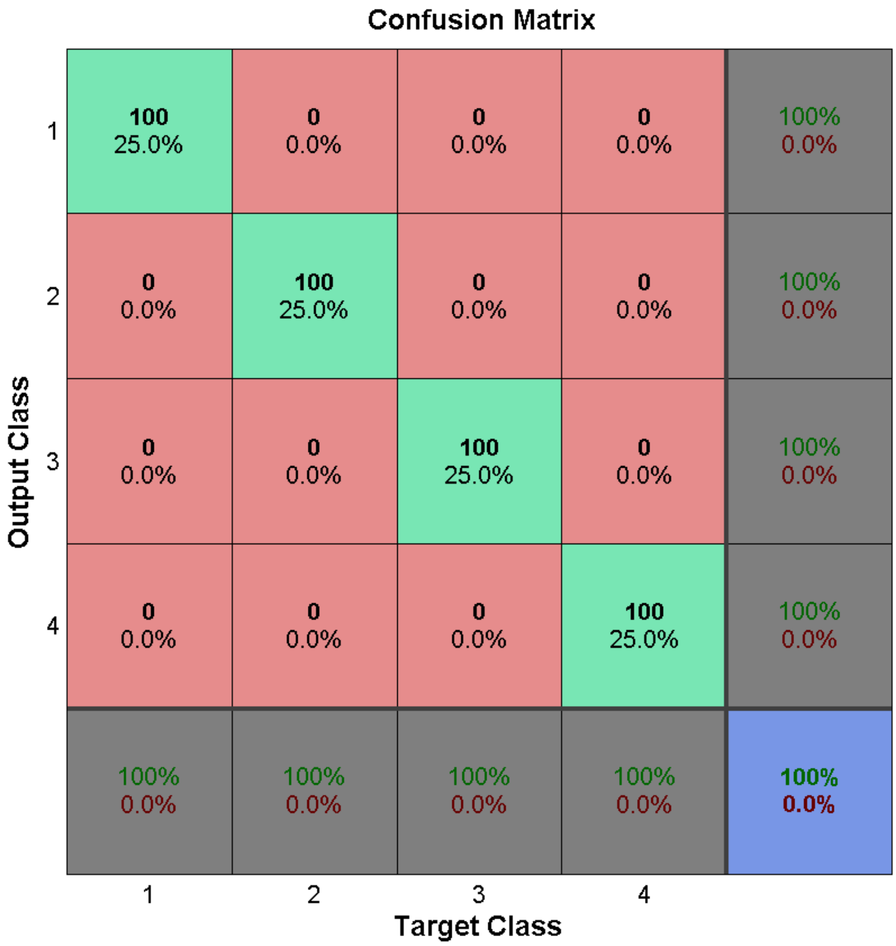


2.1. K-NEAREST NEIGHBOURS CLASSIFIER

- An optimal value of k obtained using validation data is 1, since the classifier gives 100% Accuracy for small value of k. This is justified as the data is linearly separable and from subsequent decision region plots, we can see that the data is very far apart.



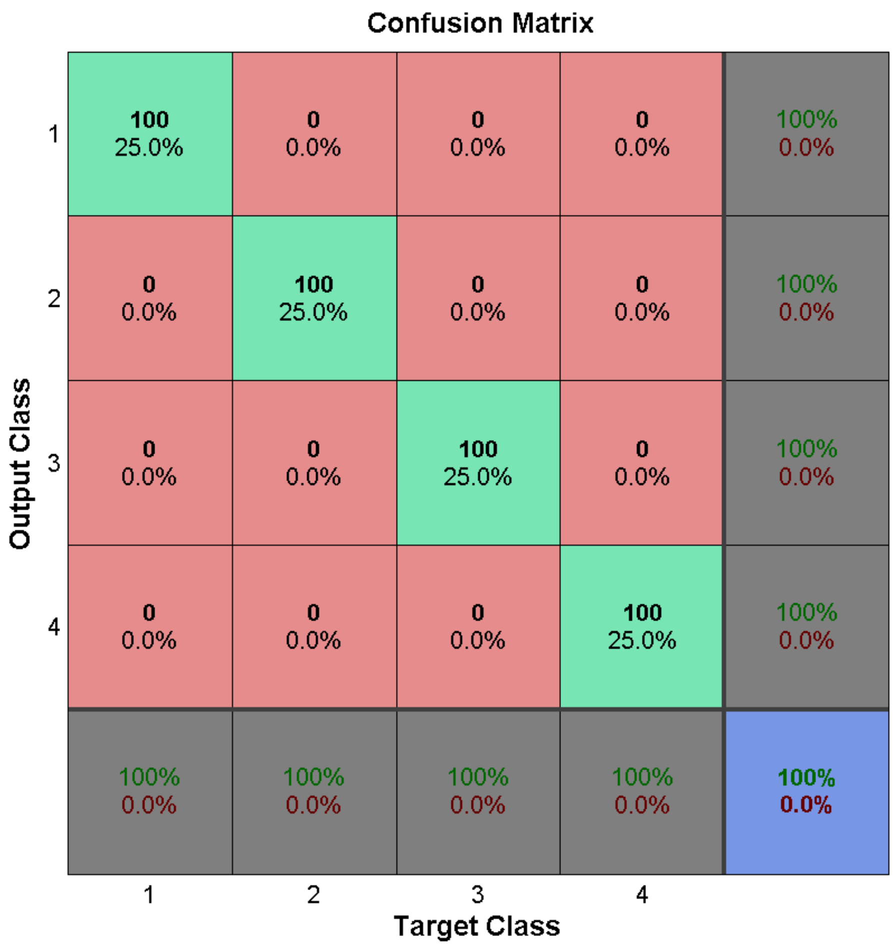
- Classification Accuracy= 1



2.2. NAÏVE BAYES CLASSIFIER

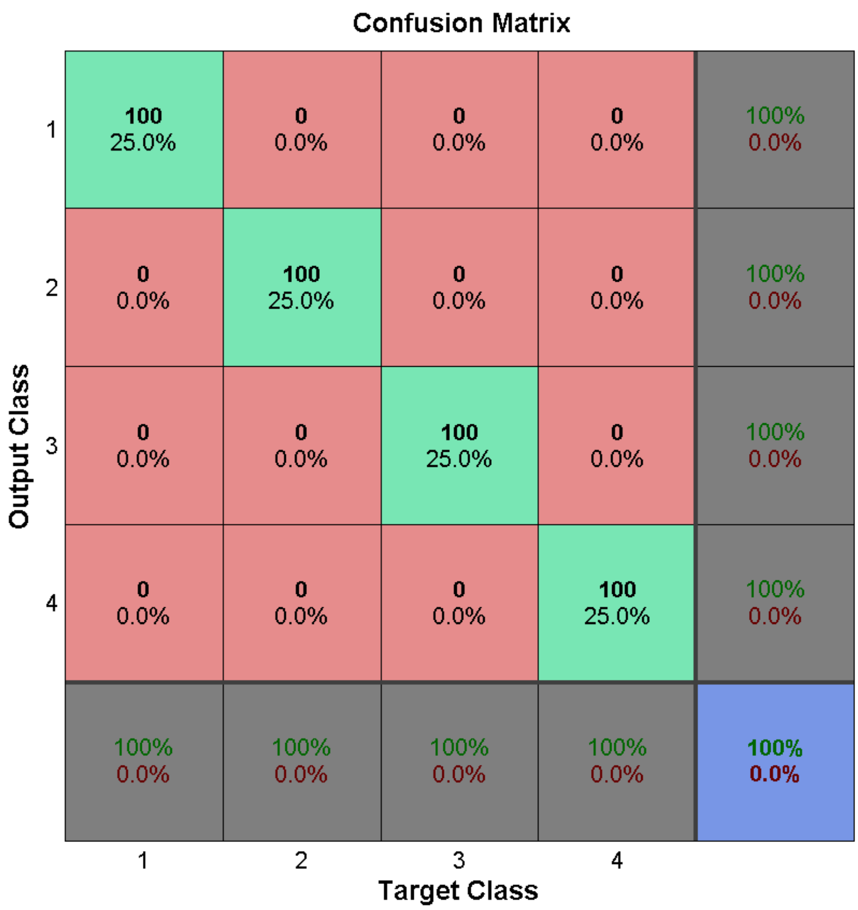
2.2.1. Covariance Matrix is same for all classes and is equal to σ^2 I

- Classification Accuracy= 1



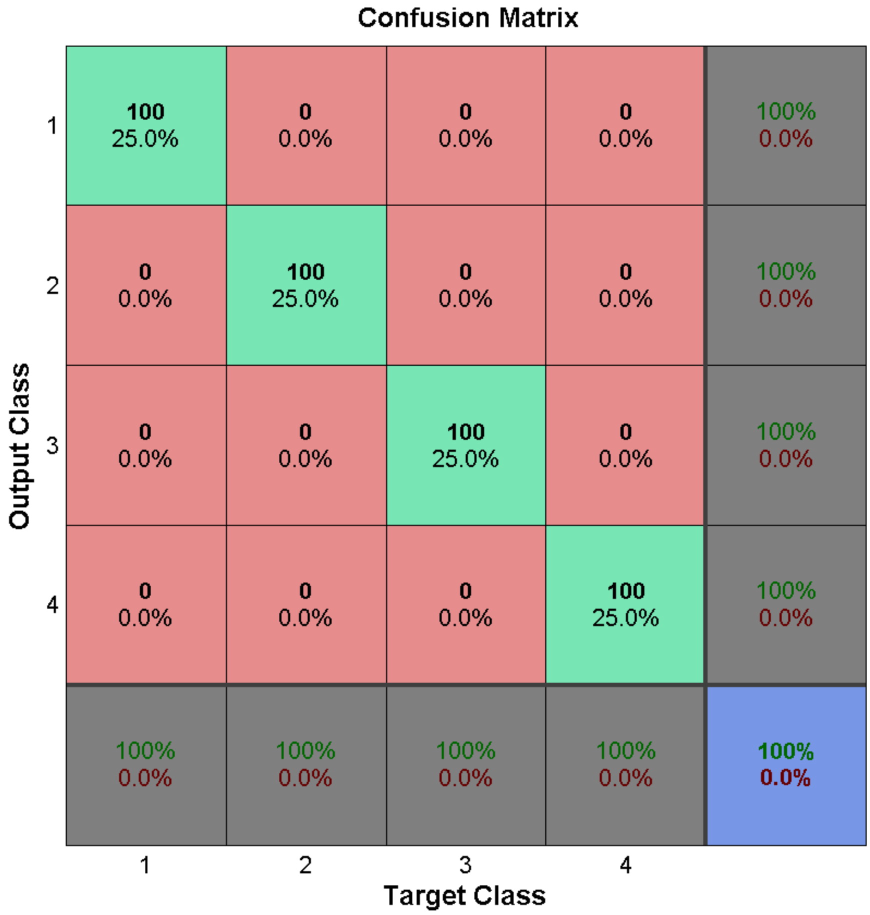
2.2.2. Covariance Matrix is same for all classes and is equal to C

- Classification Accuracy=1



2.2.3. Covariance Matrix is different for all classes

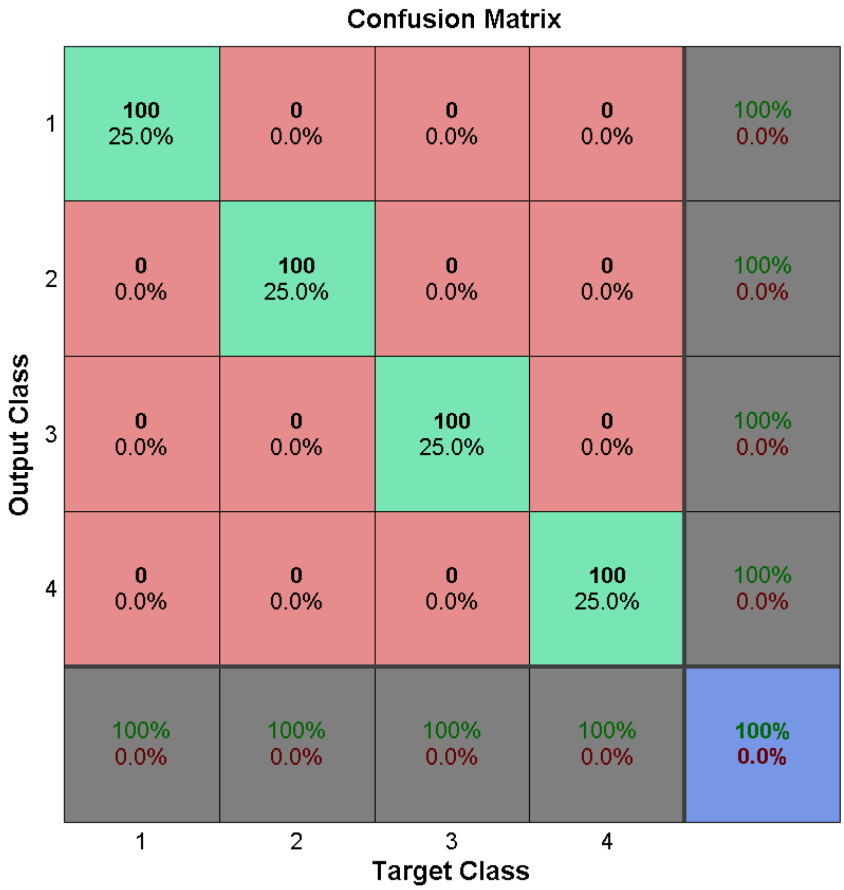
- Classification Accuracy=1



2.3. BAYES CLASSIFIER

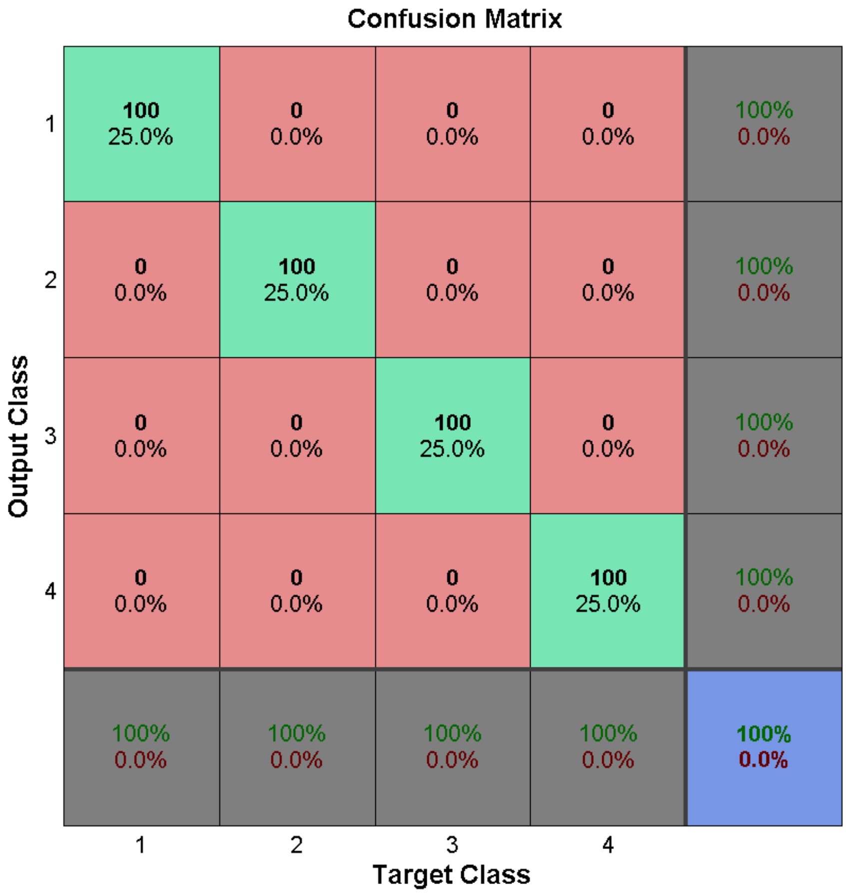
2.3.1. Covariance Matrix is same for all classes and is equal to C

- Classification Accuracy=1

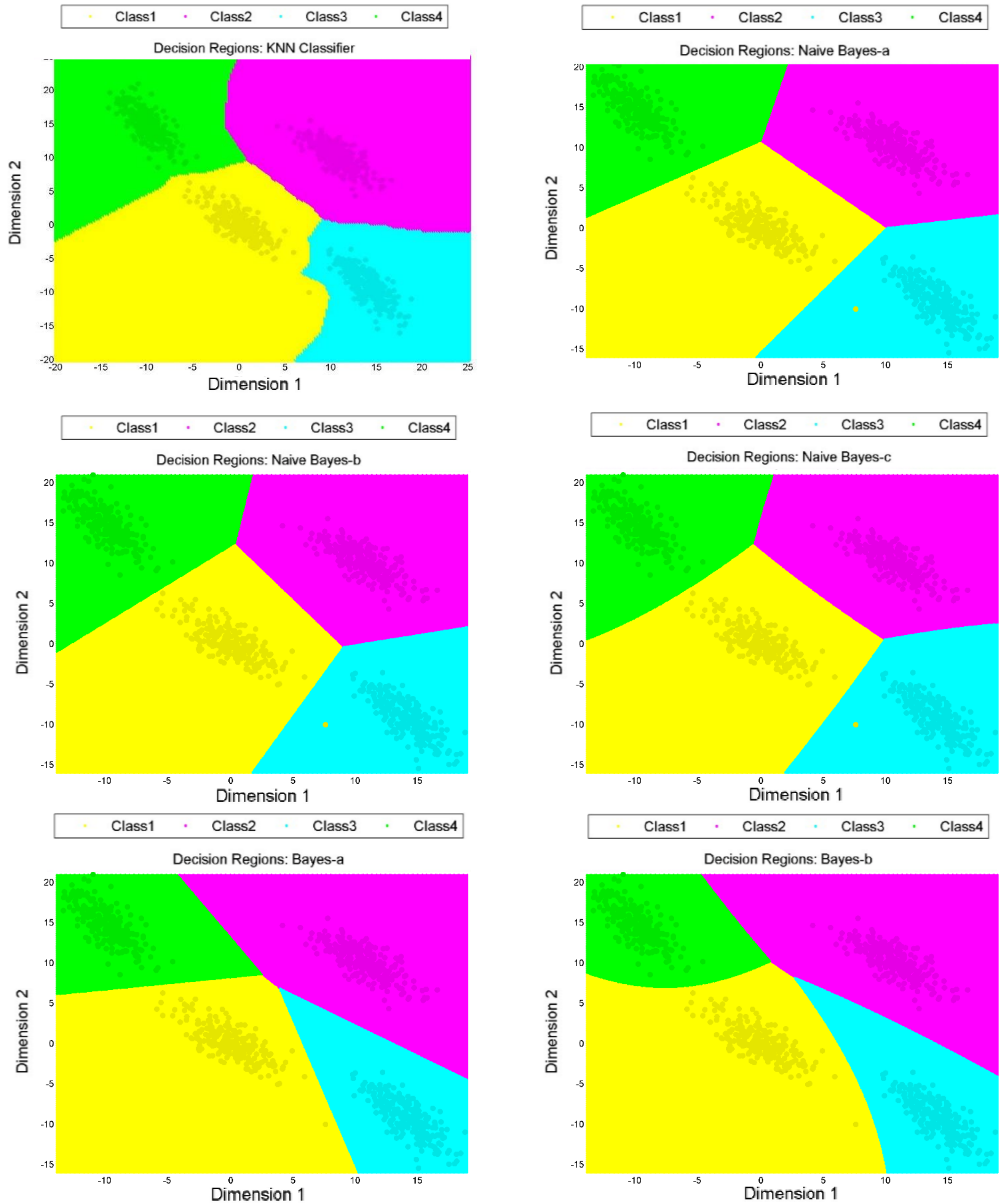


2.3.2. Covariance Matrix is different for all classes

- Classification Accuracy=1



2.4. DECISION REGIONS & OBSERVATIONS

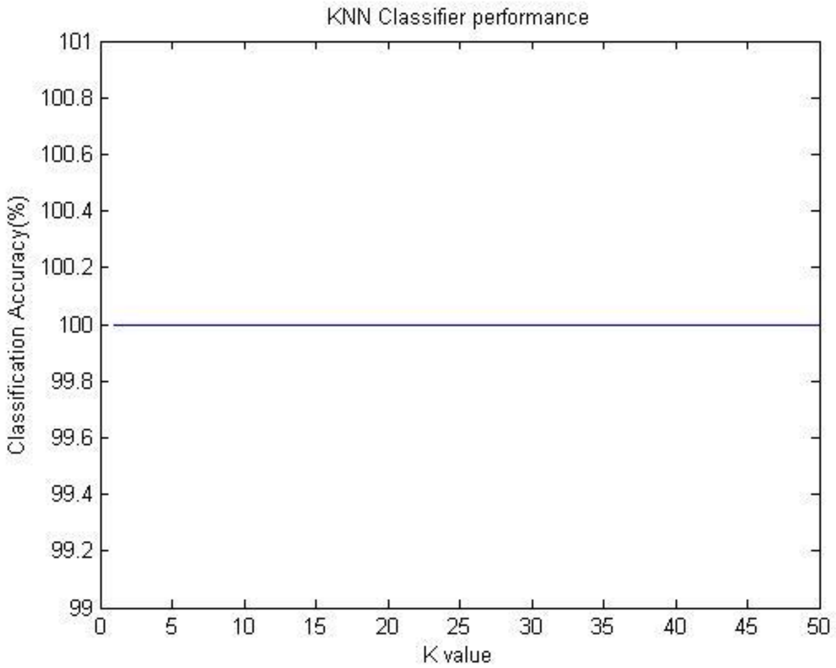


- **Nature of Decision Surface:**
    - **Linear-** Same covariances for all classes: Naïve Bayes-a, Naïve Bayes-b, Bayes-a
    - **Quadratic-** Different covariances for classes: Naïve Bayes-c, Bayes-b
    - **Piece-wise linear-** KNN Classifier
- **Performance of the classifiers** (based on test data classification) **:**
  KNN = Naïve Bayes-a = Naïve Bayes-b = Naïve Bayes-c = Bayes-a = Bayes-b
- The effect of outliers (or points far away from the majority of points in a cluster) is more pronounced in KNN Classifier than other classifiers.

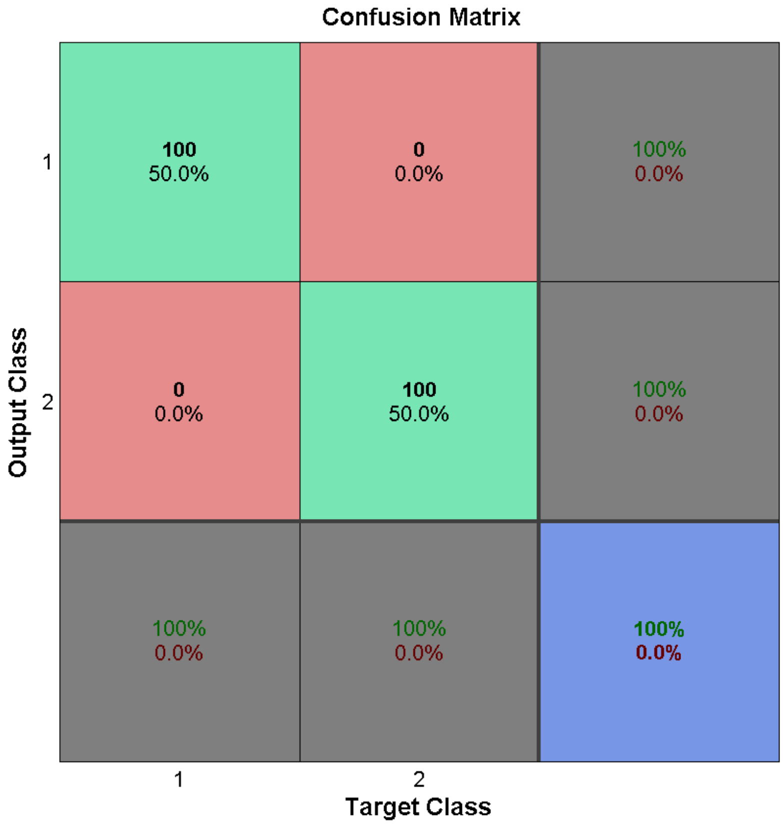
3. NON-LINEARLY SEPARABLE DATA

3.1. K-NEAREST NEIGHBOURS CLASSIFIER

- An optimal value of k obtained using validation data is 1, since the classifier gives 100% accuracy for small value of k. This is justified as the data is well separable as can be seen from subsequent decision region plots.



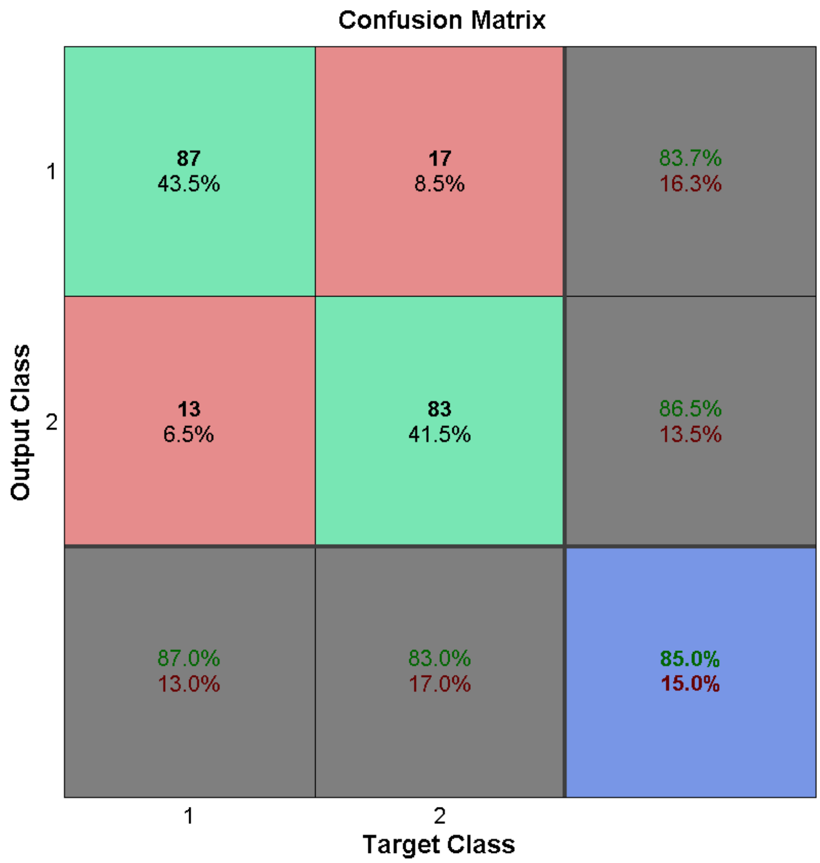
- Classification Accuracy=1



3.2. NAÏVE BAYES CLASSIFIER

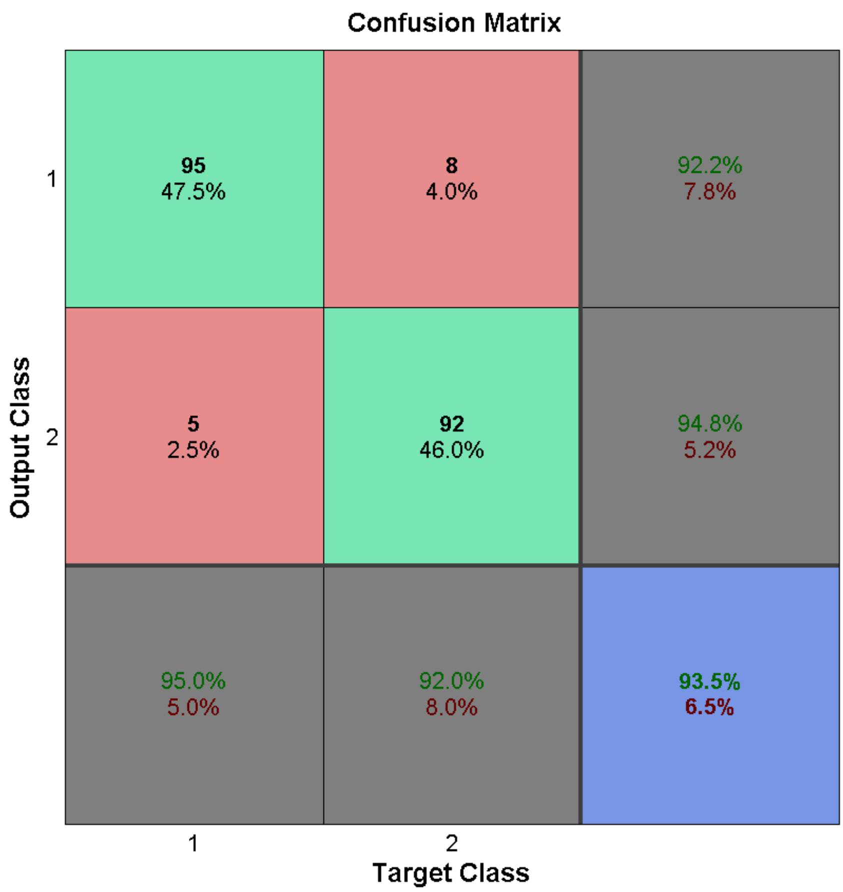
3.2.1. Covariance Matrix is same for all classes and is equal to σ^2 I

- Classification Accuracy= 0.85



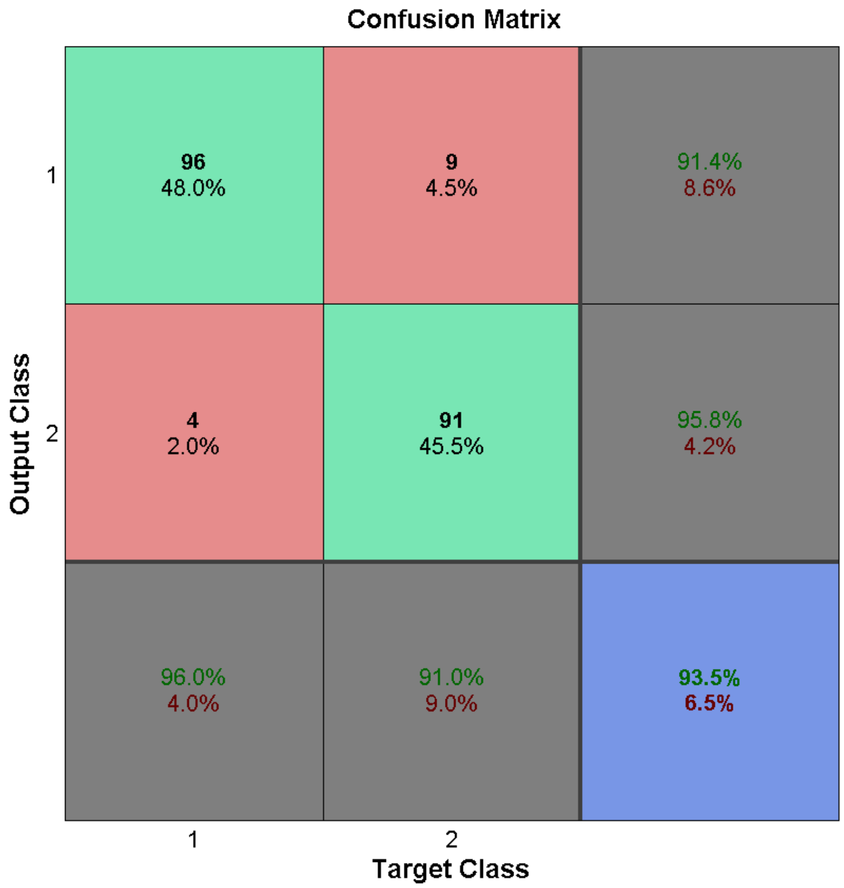
3.2.2. Covariance Matrix is same for all classes and is equal to C

- Classification Accuracy=0.935



3.2.3. Covariance Matrix is different for all classes

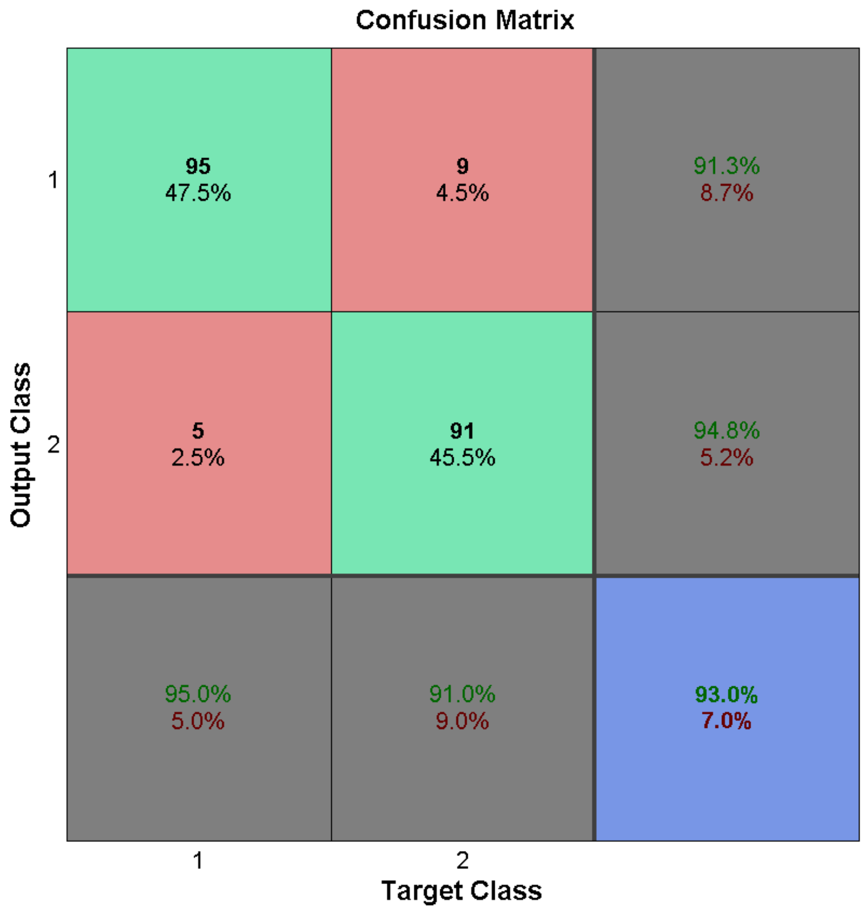
- Classification Accuracy=0.935



3.3. BAYES CLASSIFIER

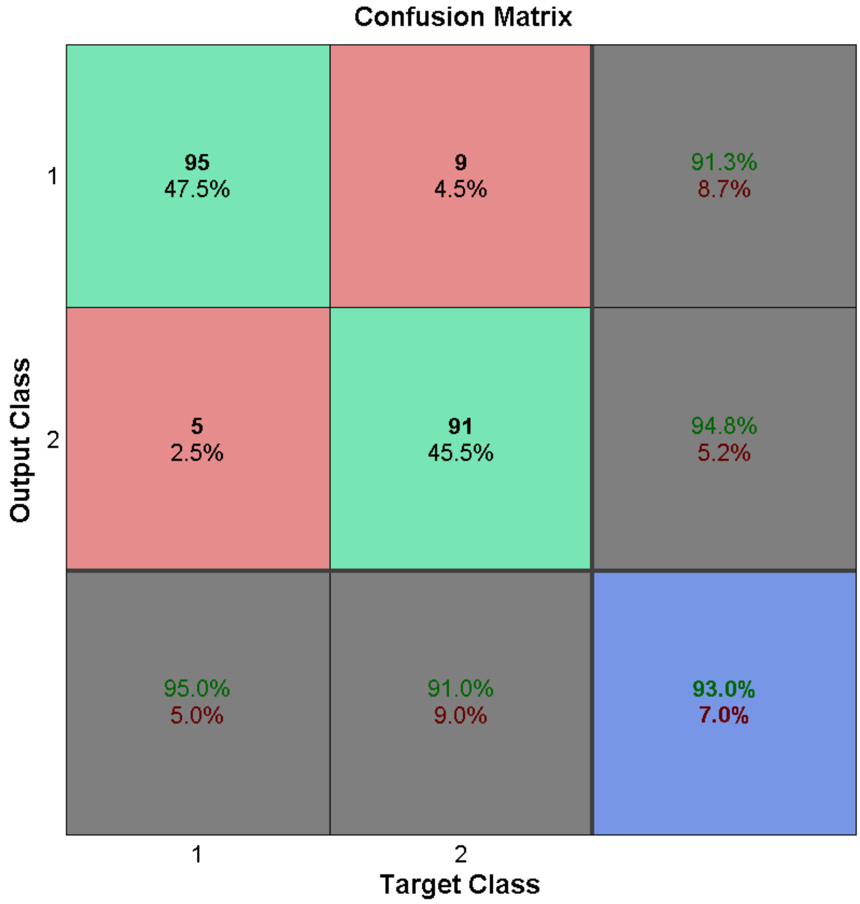
3.3.1. Covariance Matrix is same for all classes and is equal to C

- Classification Accuracy=0.93

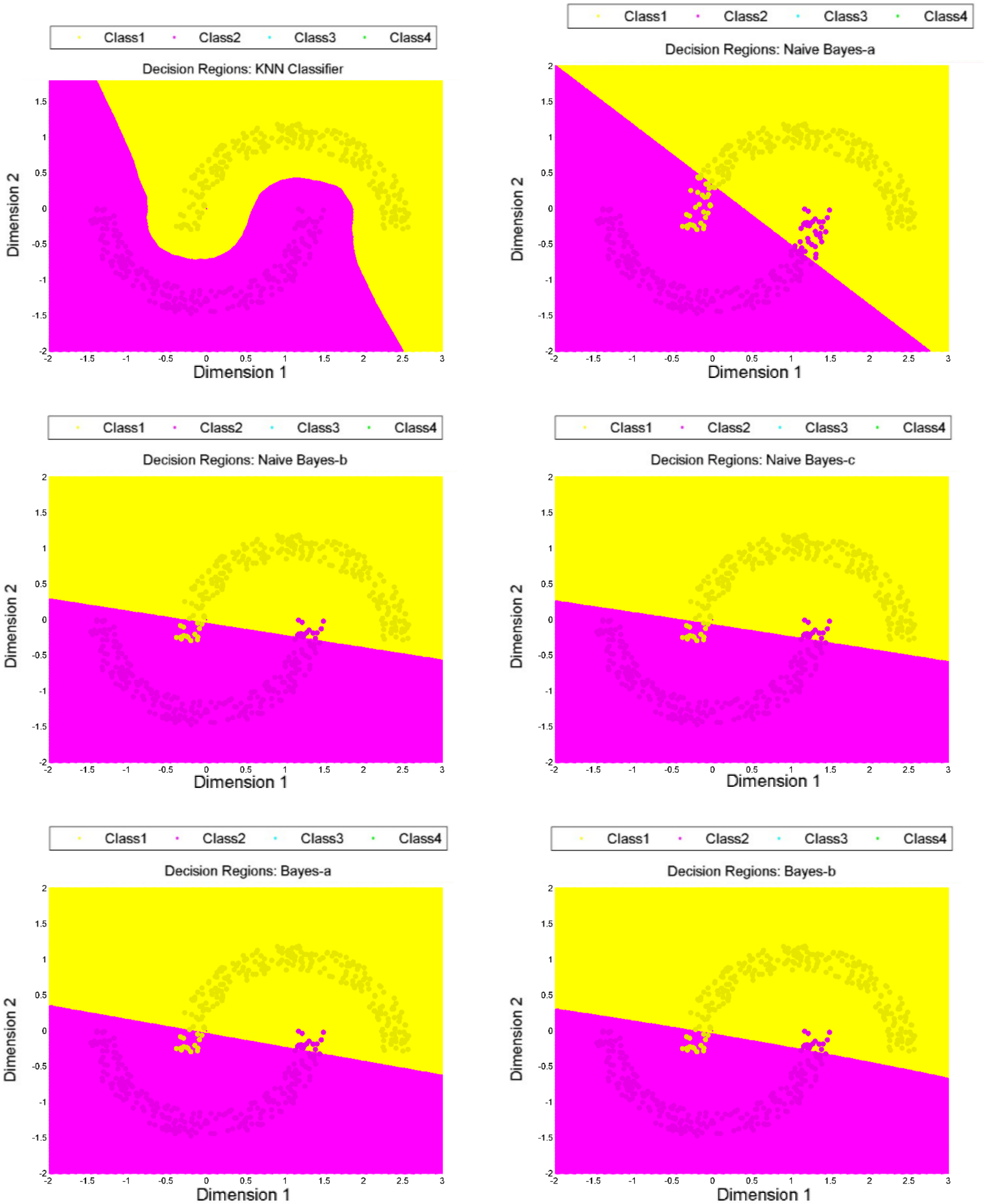


3.3.2. Covariance Matrix is different for all classes

- Classification Accuracy=0.93



3.4. DECISION REGIONS & OBSERVATIONS



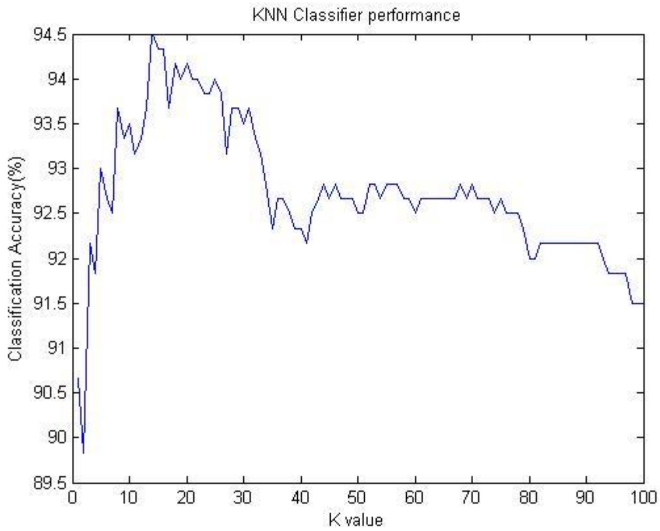
- **Nature of Decision Surface:**
  - ➢ **Linear-** Same covariances for all classes: Naïve Bayes-a, Naïve Bayes-b, Bayes-a
  - ➢ **Almost linear-** Different covariances for classes: Naïve Bayes-c, Bayes-b
  - ➢ **Piece-wise linear-** KNN Classifier
- **Performance of the classifiers** (based on test data classification) **:**
  KNN > Naïve Bayes-c =Naïve Bayes-b >Bayes-b=Bayes-a > Naïve Bayes-a

4. OVERLAPPING DATA

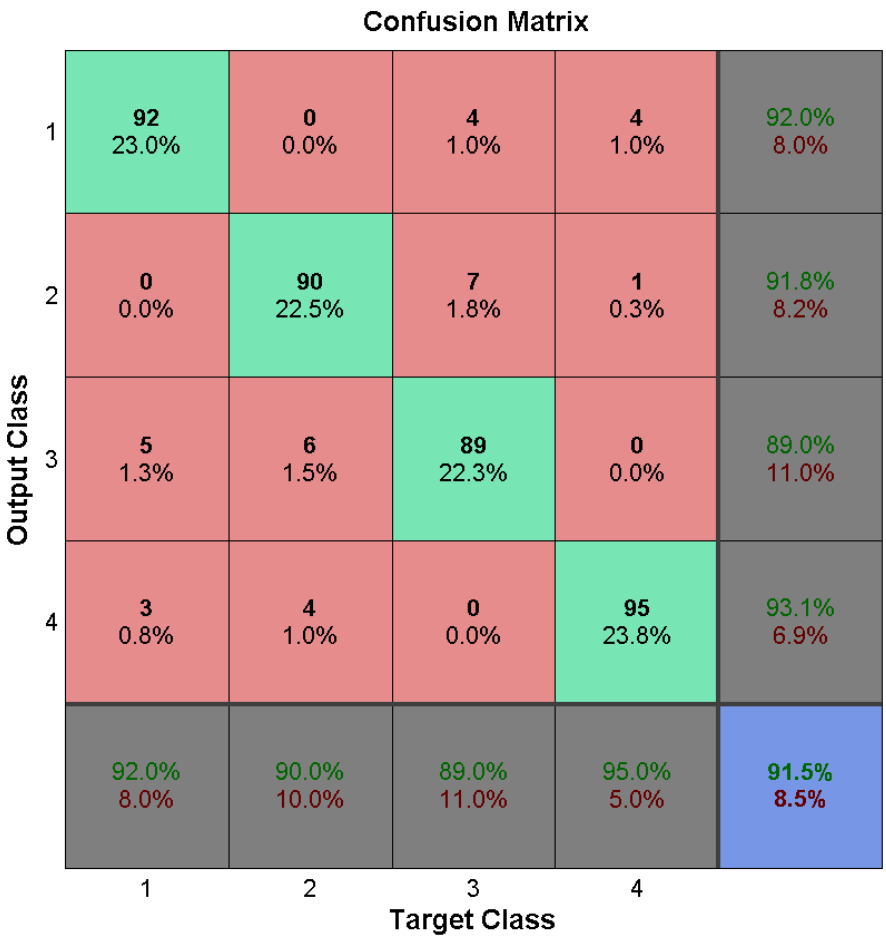
4.1. K-NEAREST NEIGHBOURS CLASSIFIER

NEW RESULTS:

- An optimal k value obtained using validation data classification is k=15. We can see that the performance increases with k up till this optimal value and then decreases. As k value increases to large values beyond 70, the performance decreases due to more points of other classes being included, due to the closeness of points in overlapping data. For very small value of k, the performance is again low because of the few points compared with, a majority could be from other classes due to the overlapping nature.



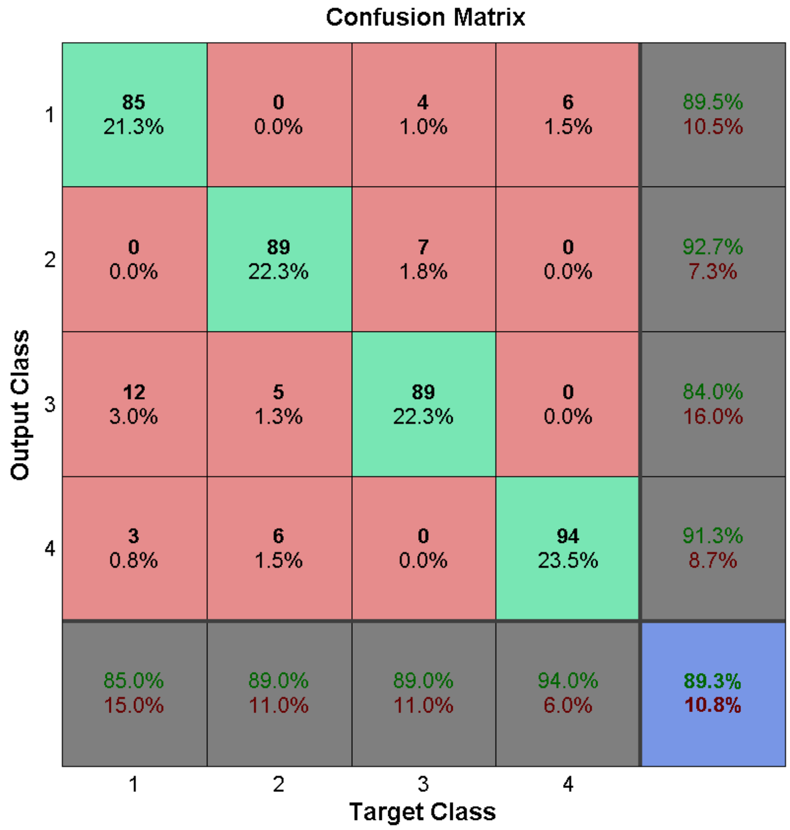
- **Classification Accuracy=0.915**



4.2. NAÏVE BAYES CLASSIFIER

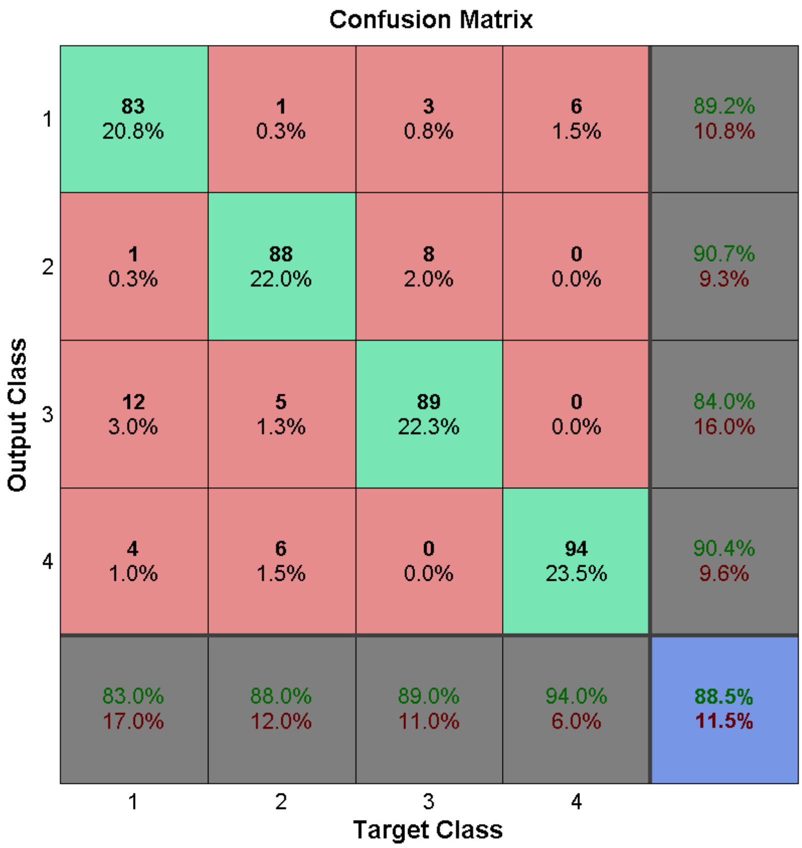
4.2.1. Covariance Matrix is same for all classes and is equal to σ^2 I

- Classification Accuracy=0.8925



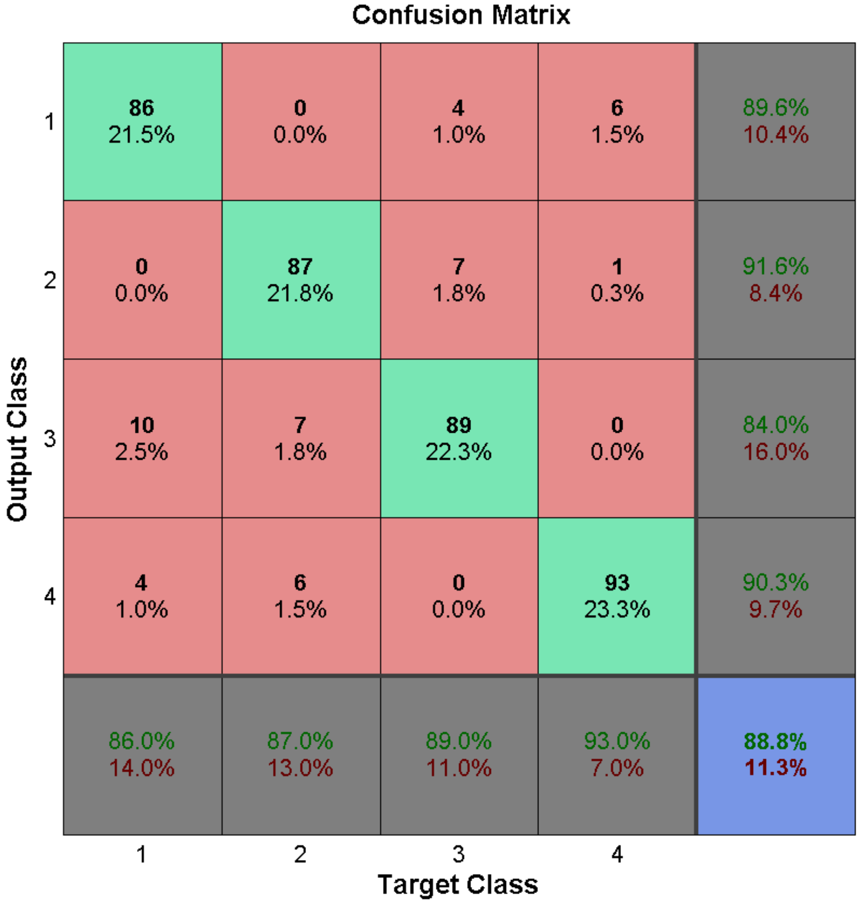
4.2.2. Covariance Matrix is same for all classes and is equal to C

- Classification Accuracy=0.885



4.2.3. Covariance Matrix is different for all classes

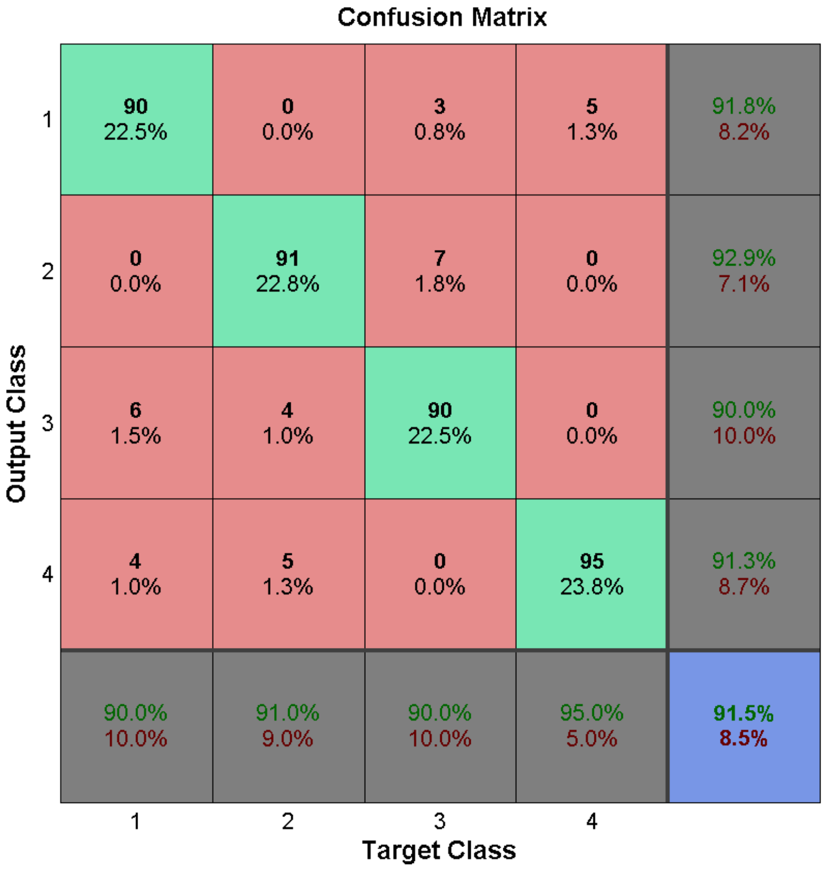
- Classification Accuracy=0.8875



4.3. BAYES CLASSIFIER

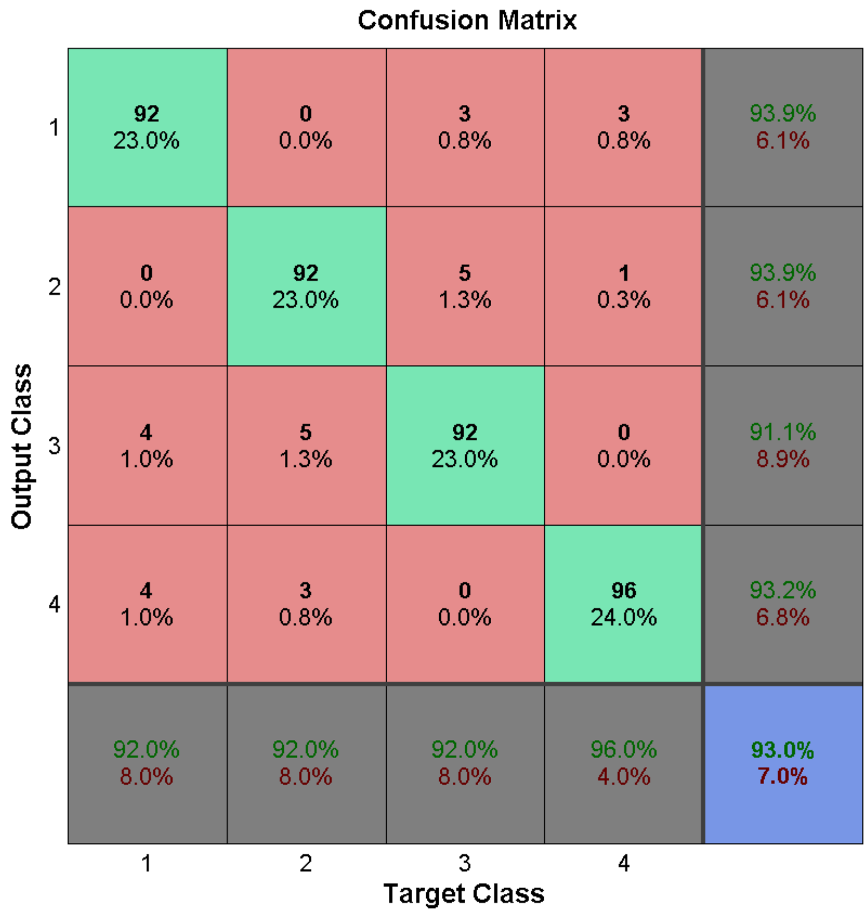
4.3.1. Covariance Matrix is same for all classes and is equal to C

- Classification Accuracy=0.915



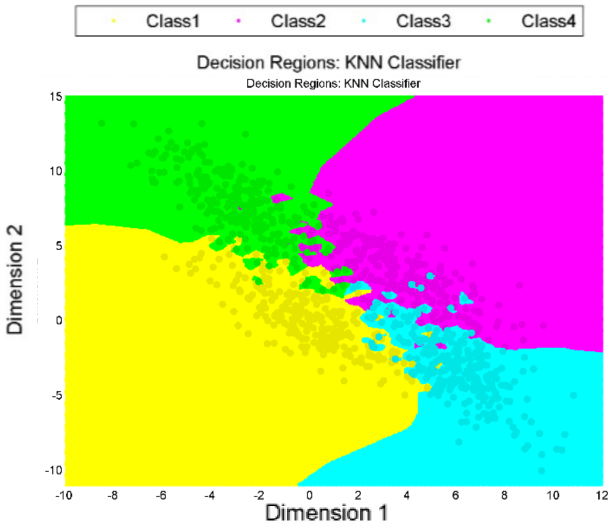
4.3.2. Covariance Matrix is different for all classes

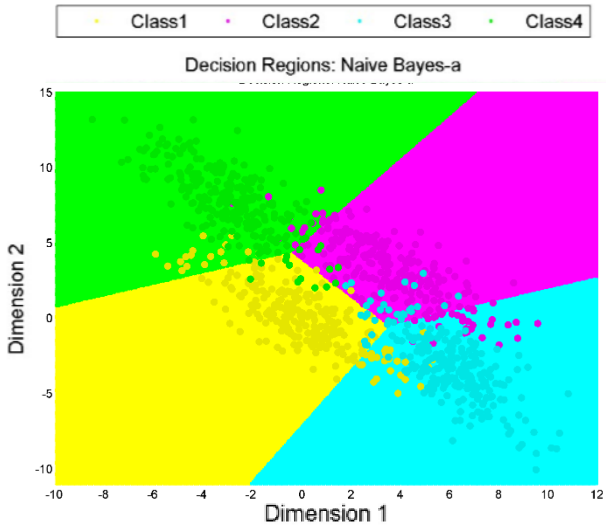
- Classification Accuracy=0.93

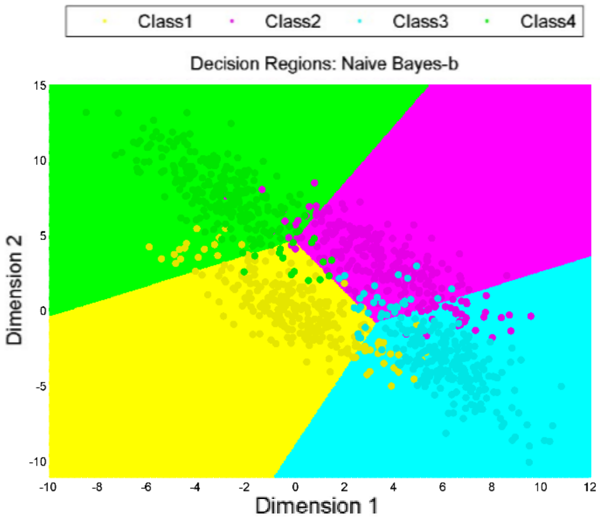


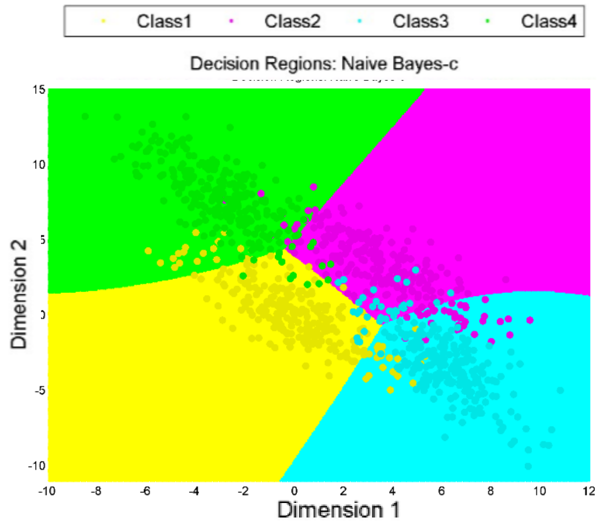
4.4. DECISION REGIONS & OBSERVATIONS

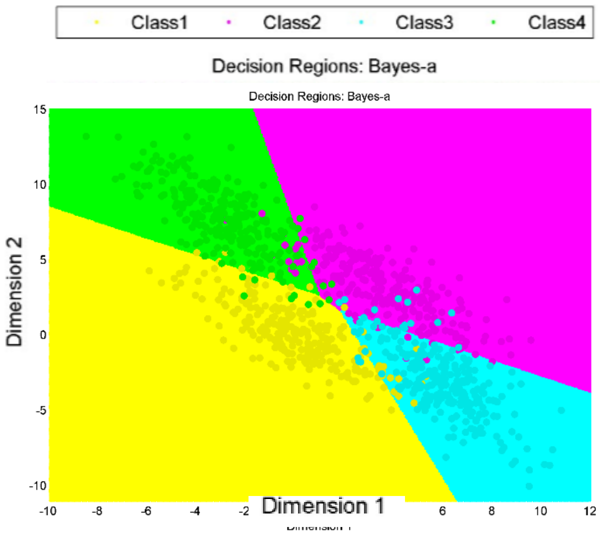
NEW RESULTS

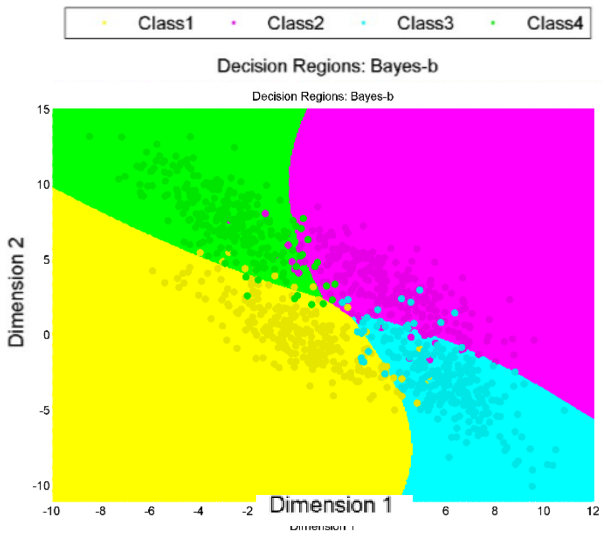












- **Nature of Decision Surface:**
  - **Linear-** Same covariances for all classes: Naïve Bayes-a, Naïve Bayes-b, Bayes-a
  - **Quadratic-** Different covariances for classes: Naïve Bayes-c, Bayes-b
  - **Arbitrary shape-** KNN Classifier
- **Performance of the classifiers** (based on test data classification) **:**
  Bayes-b >KNN > Bayes-a > Naïve Bayes-a > Naïve Bayes-c > Naïve Bayes-b

5. TRAJECTORY DATA (Sequential Pattern Classification)

5.1. Discrete HMM based Classifier

- Classification Accuracy:

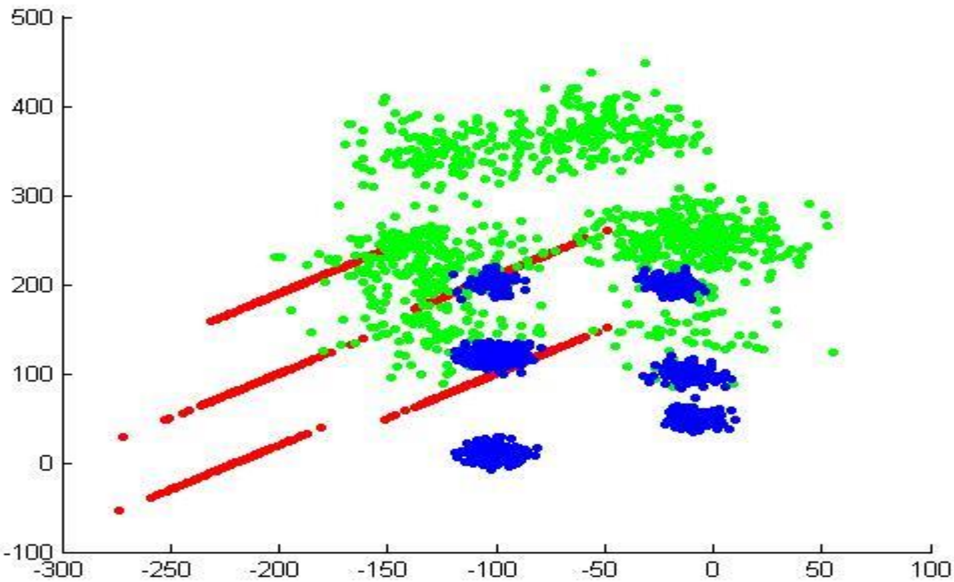
Data	Classification accuracy
Overall	16.44%
Class 5	34.67%
Class 6	14.67%
Class 7	0

- **Confusion Matrix:**

Actual\Predicted	Class 5	Class 6	Class 7
Class 5	26	26	23
Class 6	64	11	0
Class 7	75	0	0

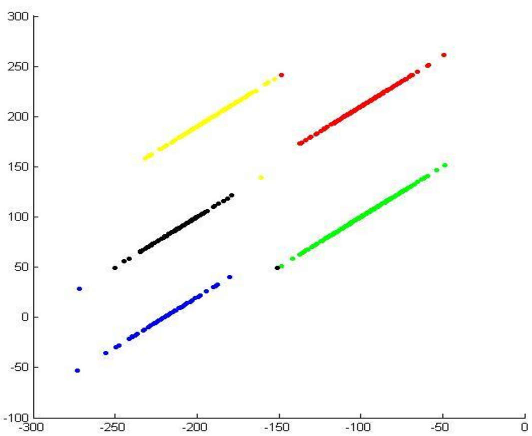
Few plots:

1. Data of all 3 classes

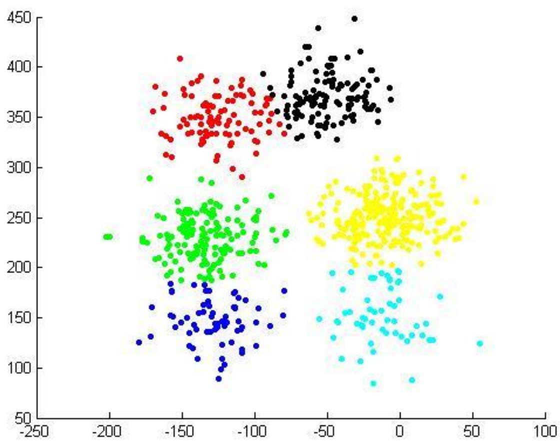


2. After clustering

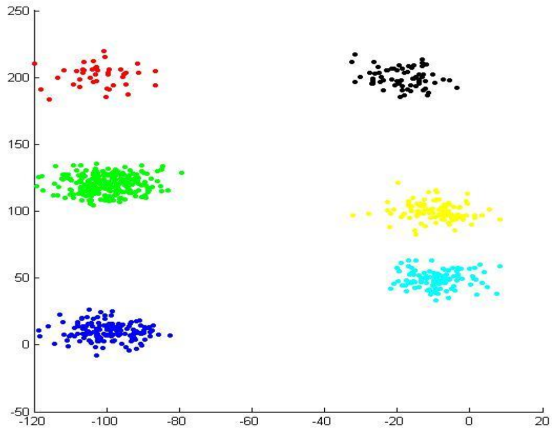
Class5:



Class 6:



Class 7:



Inference:
The classification is not very good and this might be mainly due to the fact that the code given is exclusively for left-to-right model HMM. The given data might not satisfy this assumption and hence the poor classification.

Notes:

- Each test example was converted into 3 sequences using different quantization for each class.
- Then log likelihoods were computed for each example assuming it belongs to a given class and using the specific example sequence.
- The class which gives the maximum log likelihood was chosen as the class of the test data.
- The number of observations was taken to be the same for all the states of a given class and this was equal to the number of clusters in the data
- The number of states was decided based on performance on validation data.

The plots show that the clustering code can be expected to perform good clustering even for the real world data set.

5.2. Continuous density HMM based Classifier

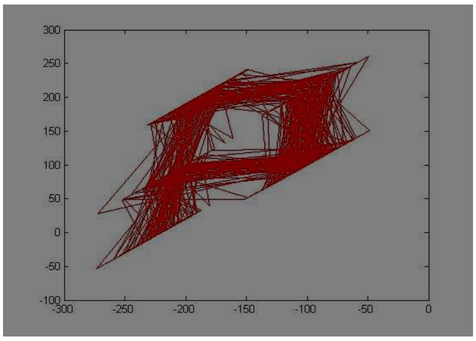
- Classification Accuracy:

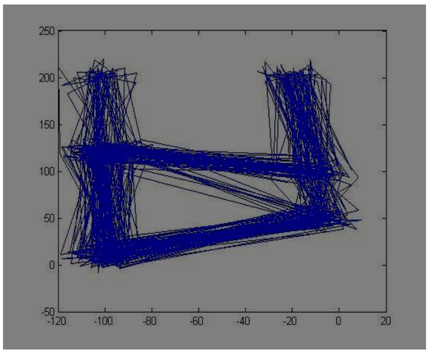
Data	**Classification accuracy**
Overall	100%
Class 5	100%
Class 6	100%
Class 7	100%

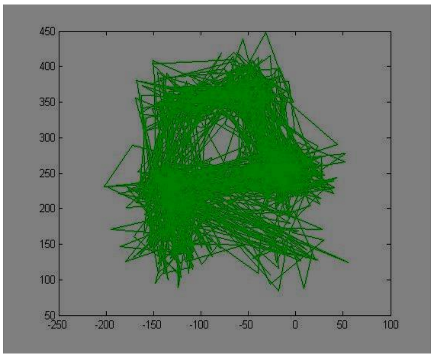
- **Confusion Matrix:**

Actual\Predicted	Class 5	Class 6	Class 7
Class 5	75	0	0
Class 6	0	75	0
Class 7	0	0	75

Plots:







Class 5 **Class 6** **Class 7**

Notes:

- The expected number of states in the 3 classes can be guessed from a look at the above plots to be equal to 5, 6 and 6 respectively.
- The number of states=3 for the initial iteration just to see the performance. But as it gave 100% accuracy no further iterations were performed.
- The input to the htk code should be 5 states because it treats the first and last state as dummy.
- The number of mixtures in each state was chosen to be 5. This was chosen because the points may lie on either the end points of these letter-like figures or on the line segments joiing them.

DATASET 2

6. IMAGE CLASSIFICATION DATA (Static Pattern Classification)

Class 1- Faces
Class 2- Grapes
Class 3- Light House
Class 4- Leopards
Class 5- Mattress
The data is randomly divided into 70% training, 15% validation and 15% test data. Dimensionality of the data is 48.

6.1. Naïve Bayes Classifier

NEW RESULTS:

- The optimal number of clusters (for the GMM) obtained from the validation data set was 2 for all the classes. Prior probabilities for all the classes are assumed to be the same. The method used is Expectation-Maximization with MLE estimates for parameters. The iterations are performed until convergence is achieved.
- Classification Accuracy:

Data	Classification accuracy (Percentage)
Overall	40
Class 1	89.2
Class 2	30.7
Class 3	13.85
Class 4	41.54
Class 5	24.62

- **Confusion Matrix:**

Actual\Predicted	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	**58**	0	3	0	4
Class 2	5	**20**	2	2	1
Class 3	5	8	**9**	0	6

Class 4	0	1	0	**27**	0
Class 5	3	9	1	0	**16**

The classifier performs well, as for all the classes, no. of correctly classified examples is higher than the number of examples classified into any other particular class.

6.2. Bayes Classifier

- The number of clusters = 2 for all classes.
- Classification Accuracy:

Data	Classification accuracy (Percentage)
Overall	17.31
Class 1	3.07
Class 2	26.66
Class 3	0
Class 4	0
Class 5	78.57

- **Confusion Matrix:**

Actual\Predicted	Class 1	Class 2	Class 3	Class 4	Class 5
Class 1	2	20	0	0	43
Class 2	1	8	0	0	21
Class 3	0	6	0	0	22
Class 4	0	6	0	0	22
Class 5	0	6	0	0	22

7. SPEAKER IDENTIFICATION & VERIFICATION DATA (Varying Length Pattern Classification- Set of local features representation)

There are 10 speakers, of which first 4 are female and remaining 6 are male. The

NEW RESULTS:

7.1. Naïve Bayes Classifier

- The optimal number of clusters (for the GMM) obtained from the validation data set was 9 for all the classes. Prior probabilities for all the classes are assumed to be the same. The method used is Expectation-Maximization with MLE estimates for parameters. The iterations are performed until convergence is achieved.
- Classification Accuracy:

Data	Classification accuracy (Percentage)
Overall	**50**
Class 1	60
Class 2	40
Class 3	60
Class 4	80
Class 5	50
Class 6	10
Class 7	60
Class 8	80
Class 9	40
Class 10	20

- **Confusion Matrix:**

Actual\Predicted	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10
Class 1	**6**	0	0	0	1	0	2	0	0	1
Class 2	4	**4**	0	1	0	0	1	0	0	0
Class 3	2	1	**6**	1	0	0	0	0	0	0
Class 4	2	0	0	**8**	0	0	0	0	0	0
Class 5	0	0	2	0	**5**	2	0	0	1	0
Class 6	3	0	0	0	1	**1**	2	1	0	2
Class 7	0	0	0	0	1	0	**6**	0	0	3
Class 8	0	0	0	0	1	0	0	**8**	1	0
Class 9	0	0	0	0	2	0	2	1	**4**	1
Class 10	0	0	2	0	2	0	2	1	1	**2**

Total no. of examples in each class is 10. We can see that the speaker identification does a good job in classifying whether the person is male or female. Here, the first 4 classes are female and the last 6 are male.

7.2. Bayes Classifier

- The optimal number of clusters obtained from the validation data set was 5 for all the classes. Prior probabilities for all the classes are assumed to be the same. The method used is Expectation-Maximization with MLE estimates for parameters. The iterations are performed until convergence is achieved.
- Classification Accuracy:

Data	**Classification accuracy (Percentage)**
Overall	**68**
Class 1	80
Class 2	70
Class 3	60
Class 4	100
Class 5	40
Class 6	0
Class 7	70
Class 8	100
Class 9	80
Class 10	80

- **Confusion Matrix:**

Actual\Predicted	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10
Class 1	**8**	0	0	0	0	0	1	0	0	1
Class 2	2	**7**	0	0	0	0	0	1	0	0
Class 3	0	0	**6**	2	0	0	0	0	0	2
Class 4	0	0	0	**10**	0	0	0	0	0	0
Class 5	0	0	1	0	**4**	1	2	1	0	1
Class 6	0	0	0	0	1	**0**	3	0	0	6
Class 7	0	0	1	0	0	0	**7**	1	0	1
Class 8	0	0	0	0	0	0	0	**10**	0	0
Class 9	0	0	0	0	0	0	1	0	**8**	1
Class 10	0	0	0	0	0	0	1	1	0	**8**

- Performance in terms of classification accuracy is higher for the Bayes' Classifier when compared to the Naïve Bayes Classifier.
- We can see that the speaker identification does a good job in classifying whether the person is male or female as in the Naïve Bayes case.

- As the features are not independent, the Bayes classifier performs better than the Naïve Bayes Classifier.

8. ISOLATED WORD RECOGNITION DATA (Sequential Pattern Classification)

8.1. Discrete HMM based Classifier

- Old Classification Accuracy:

Data	Classification accuracy
Overall	75%
Class 1	91.67%
Class 2	45.83%
Class 3	62.5%
Class 4	100%

- **Old Confusion Matrix:**

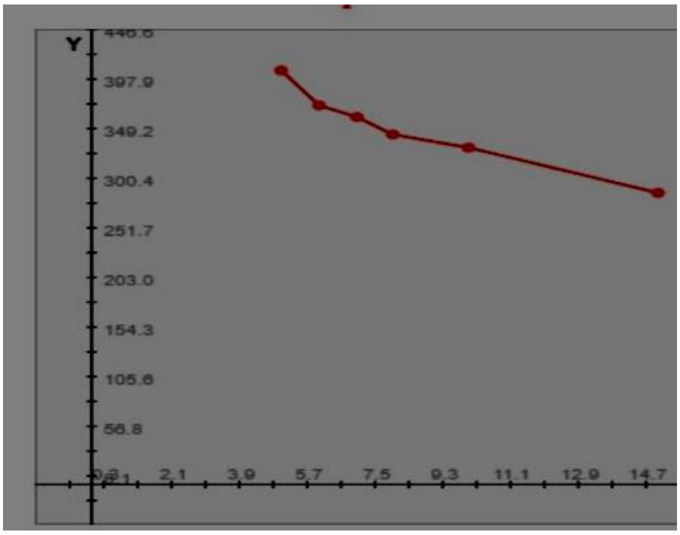
Actual\Predicted	Class 1	Class 2	Class 3	Class 7
Class 1	22	0	0	2
Class 2	0	11	0	13
Class 3	4	0	15	5
Class 7	0	0	0	24

Old Method:

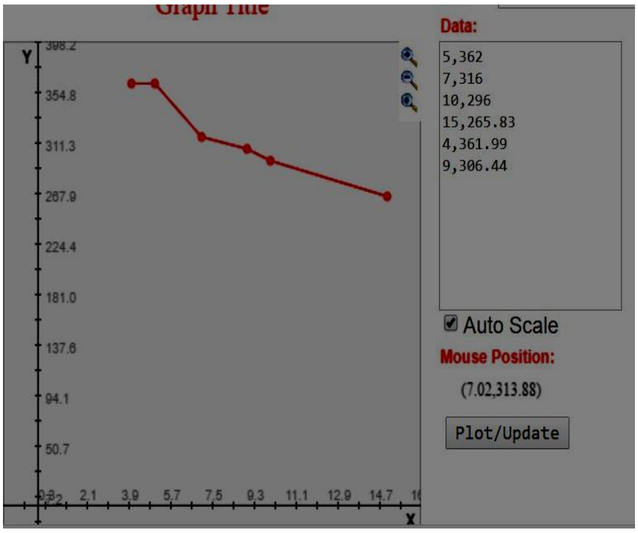
The number of observables for each class was found by plotting number of clusters vs average variance(see below for images) of the clusters. The knee of the curve was chosen as the optimum number of clusters.

The number of states was decided based on performance of validation data.

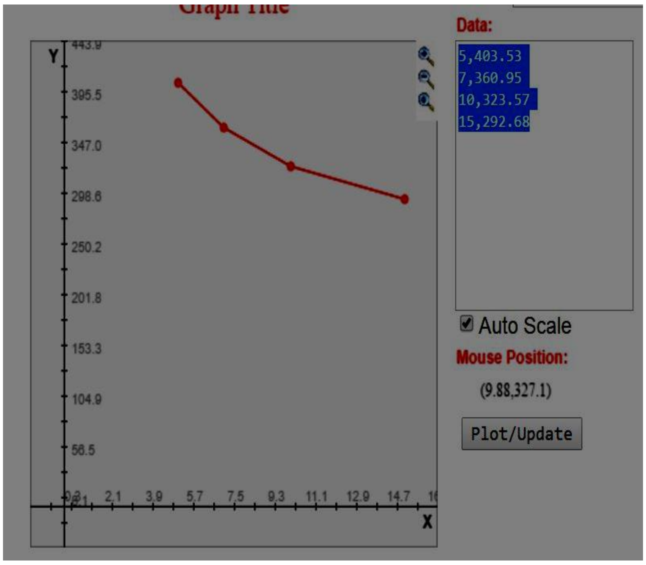
Class1



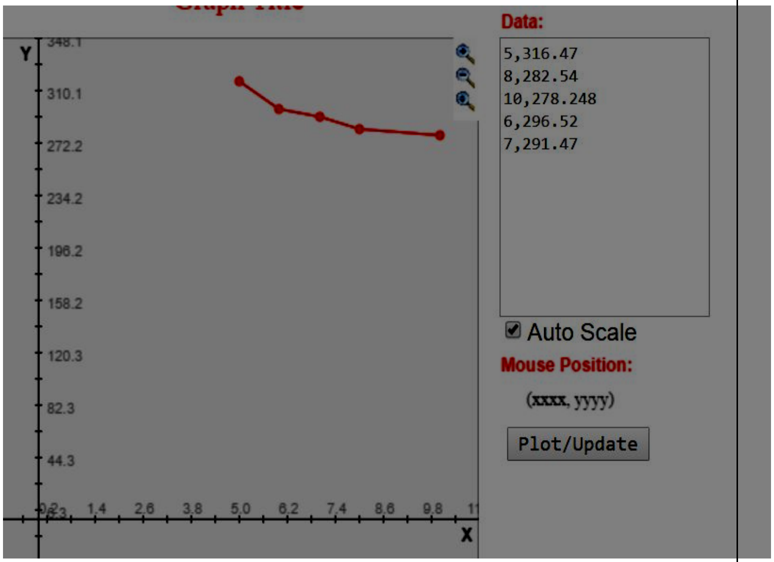
Class2



Class 3



Class 7:



Notes:

- The results are those produced when we use the above-mentioned criterion of cluster selection.
- It was observed that there are cases where this criterion of cluster formation fails and therefore we did the analysis using different number of states and number of clusters for each state by varying one in the range 3 to 5 holding the other fixed. The best results were obtained when each of them was 3.
- The seed decides the initial parameters of the hmm.
- The **revised classification accuracy** in this case is:

Data	Classification accuracy
Overall	82.29%
Class 1	95.83%
Class 2	54.17%
Class 3	79.17%
Class 4	100%

- **Revised Confusion Matrix:**

Actual\Predicted	Class 1	Class 2	Class 3	Class 7
Class 1	23	0	1	0
Class 2	0	13	0	11
Class 3	2	0	19	3
Class 7	0	0	0	24

8.2. Continuous density HMM based Classifier

- Classification Accuracy:

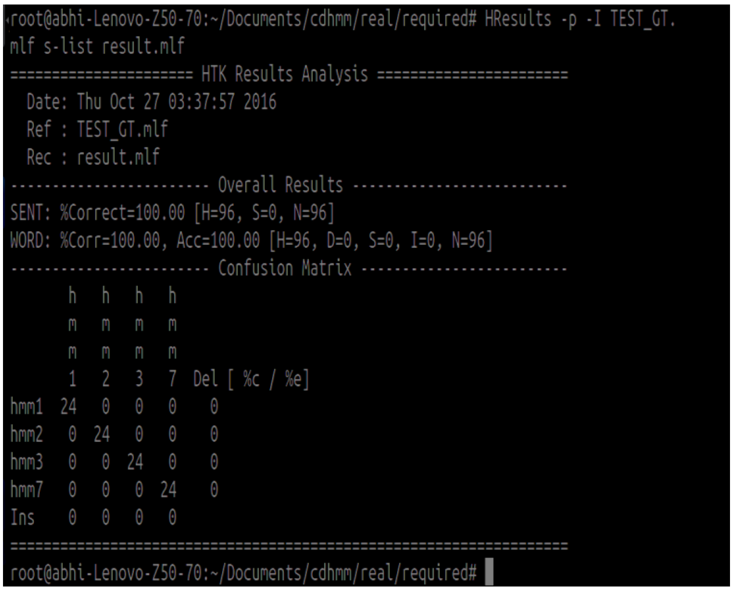
Data	Classification accuracy
Overall	100%
Class 1	100%
Class 2	100%
Class 3	100%
Class 7	100%

- **Confusion Matrix:**

Actual\Predicted	Class 1	Class 2	Class 3	Class 7
Class 1	24	0	0	0
Class 2	0	24	0	0
Class 3	0	0	24	0
Class 7	0	0	0	24

Notes:

- Each class represents a word.
- For the first iteration we used 3 states for each classes(because one=va+a+n,two=tu+oo,three=th+r+ee,seven=se+ve+n) and 3 gaussian mixtures(since each break-up of word(eg. Va,th,se etc.) has a beginning, a middle part and an end)for each state. HRest was used only once.
- As the results obtained were satisfactory, no further iterations were carried out.



```
root@abhi-Lenovo-Z50-70:~/Documents/cdhmm/real/required# HResults -p -I TEST_GT.
mlf s-list result.mlf
==================== HTK Results Analysis =====================
  Date: Thu Oct 27 03:37:57 2016
  Ref : TEST_GT.mlf
  Rec : result.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=100.00 [H=96, S=0, N=96]
WORD: %Corr=100.00, Acc=100.00 [H=96, D=0, S=0, I=0, N=96]
------------------------ Confusion Matrix -------------------------
       h   h   h   h
       m   m   m   m
       m   m   m   m
       1   2   3   7  Del [ %c / %e]
hmm1  24   0   0   0    0
hmm2   0  24   0   0    0
hmm3   0   0  24   0    0
hmm7   0   0   0  24    0
Ins    0   0   0   0
===================================================================
root@abhi-Lenovo-Z50-70:~/Documents/cdhmm/real/required#
```