Pattern Recognition
CS669

Course Instructor : Dr. Dileep A. D.

Final Report

Fisher Discriminant Analysis,
Perceptron based classifier
&
Support Vector Machines

Amrendra Singh : B16010

Contents

List of Tables

1. Problem Description

Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs on the basis of a training set of data containing observations (or instances) whose category membership is known.

Data-sets:

1. 2 dimensional artificial data
   (a) Linearly separable dataset used in Assignment1
   (b) Nonlinearly separable data set used in Assignment1
2. 3 class scene image dataset: Consider the 64-dimensional BoVW representation from Assignment-2

Classifiers to be built:

1. Apply Fisher linear discriminant analysis (FDA) on Dataset-1 and Dataset-2. Use Bayes classifier using both unomodal Gaussian and GMM
2. Perceptron-based classifier on Dataset-1(a).
3. SVM-based classifier using (a) linear kernel, (b) polynomial kernel and (c) Gaussian/RBF kernel on Dataset-1 and Dataset-2

2. Solution Approach

1 Fisher Discriminant Analysis

Procedure

1. For every pair of classes:
   - Calculate the direction for data projection ω
     $\omega \ = \ \lambda \ S^{-1} \ (\mu_+ \ - \ \mu_-)$
   - Project the data of the two classes on the obtained ω
   - Now use a density approximation method with bayes classifier to classify the data using a voting based method.

2 Perceptron

I used a batch perceptron based method in this assignment wherein we take all the training points of all classed at a time and feed it to our system as a batch in a single iteration. This method is continued till no training examples get wrongly classified.

For the classification I have used a one-one approach as that seemed more intuitive. Then I took a voting method to assign the final class to a test point.

3 Support Vector Machines

I have used the sklearn.svm python package for this.
I have used three kernels with the SVM

- Linear Kernel
- Polynomial Kernel with degree as 2 and 3
- RBF or Gaussian Kernel

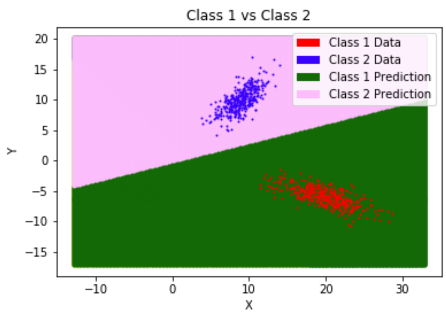
Deafult parameters for sklearn.svm

1. Default kernel is RBF
2. For polynomial kernel, default degree is 3
3. $a = \frac{1}{No.of dimensions}$
4. $b = 0$

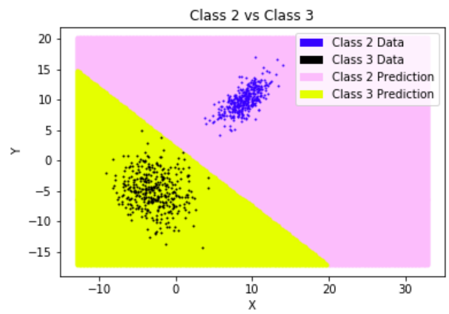
3. Results

1 Dataset 1(a): Artificial Linearly Separable Data

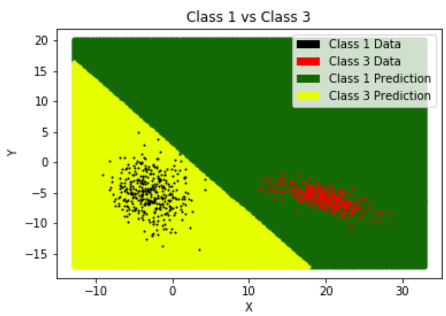
1.1 Perceptron based classification



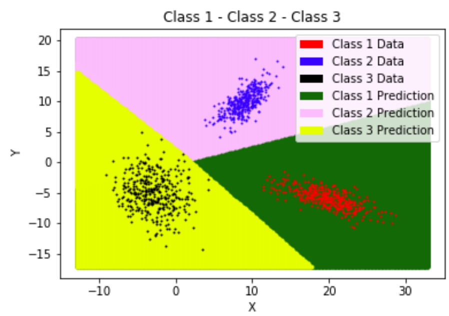
(a) Decision regions and Training data Class 1 and 2



(b) Decision regions and Training data Class 2 and 3



(c) Decision regions and Training data Class 1 and 3



(d) Decision regions and Training data Class 1, 2, 3

Figure 3..1. Perceptron : Linearly Separable Data - Decision Regions

	Class 1	Class 1	Class 3
Class 1	125	0	0
Class 2	0	125	10
Class 3	0	0	125

(a) Confusion Matrix

	Class 1	Class 2	Class 3
Precision	1	1	1
Recall	1	1	1
F-Measure	1	1	1

(b) Analysis

Accuracy	100%
Precision	1
Recall	1
F-Measure	1

(c) Results

Table 3..1. Perceptron : Linearly Separable Data - Confusion Matrix and Analysis

1.2 Support Vector Machines

Linear Kernel

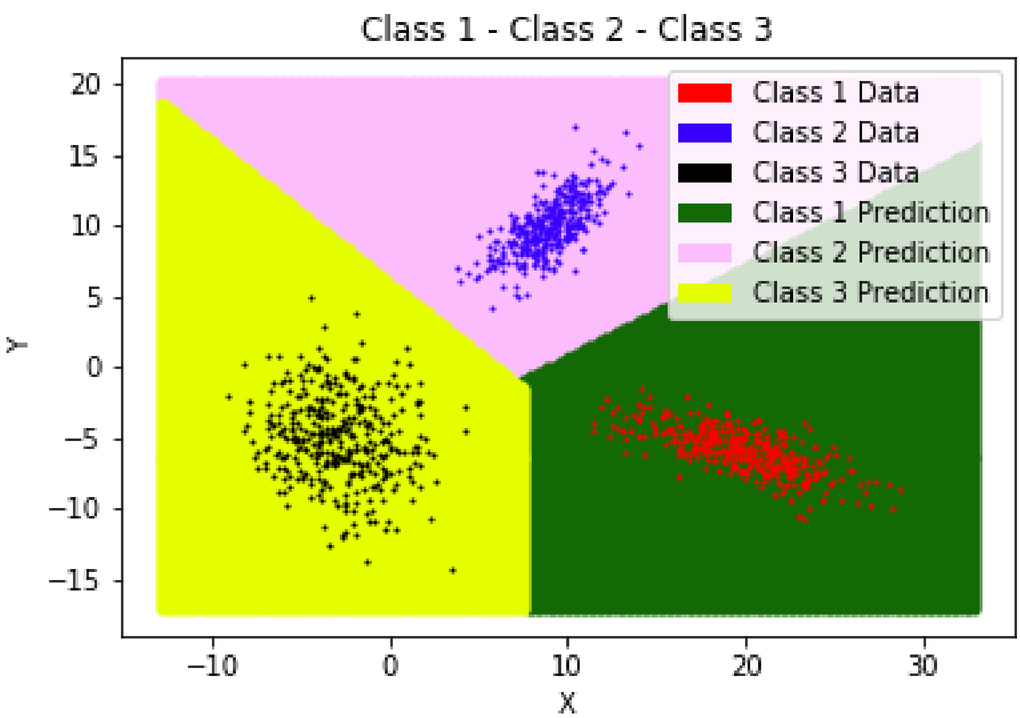


Figure 3..2. SVM : Linearly Separable Data - Decision Regions

	Class 1	Class 1	Class 3
Class 1	125	0	0
Class 2	0	125	10
Class 3	0	0	125

(a) Confusion Matrix

	Class 1	Class 2	Class 3
Precision	1	1	1
Recall	1	1	1
F-Measure	1	1	1

(b) Analysis

Accuracy	100%
Precision	1
Recall	1
F-Measure	1

(c) Results

Table 3..2. SVM Linear Kernel: Linearly Separable Data - Results

Polynomial Kernel

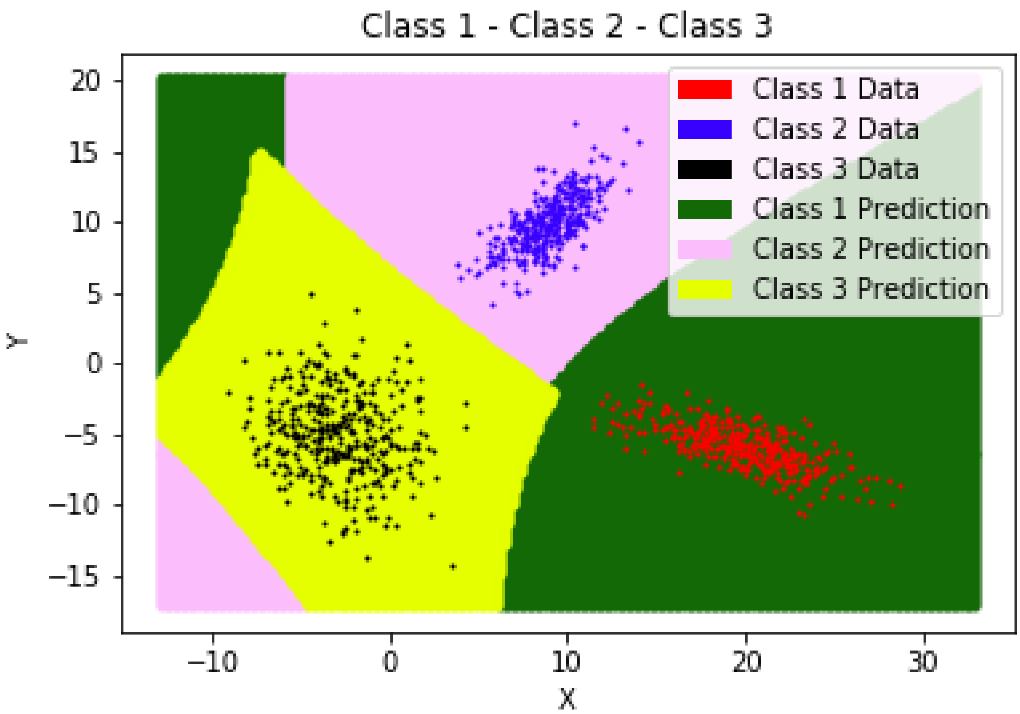


Figure 3..3. SVM Polynomial Kernel: Linearly Separable Data - Decision Regions

	Class 1	Class 1	Class 3
Class 1	125	0	0
Class 2	0	125	10
Class 3	0	0	125

(a) Confusion Matrix

	Class 1	Class 2	Class 3
Precision	1	1	1
Recall	1	1	1
F-Measure	1	1	1

(b) Analysis

Accuracy	100%
Precision	1
Recall	1
F-Measure	1

(c) Results

Table 3..3. SVM Polynomial Kernel: Linearly Separable Data - Results

RBF Kernel

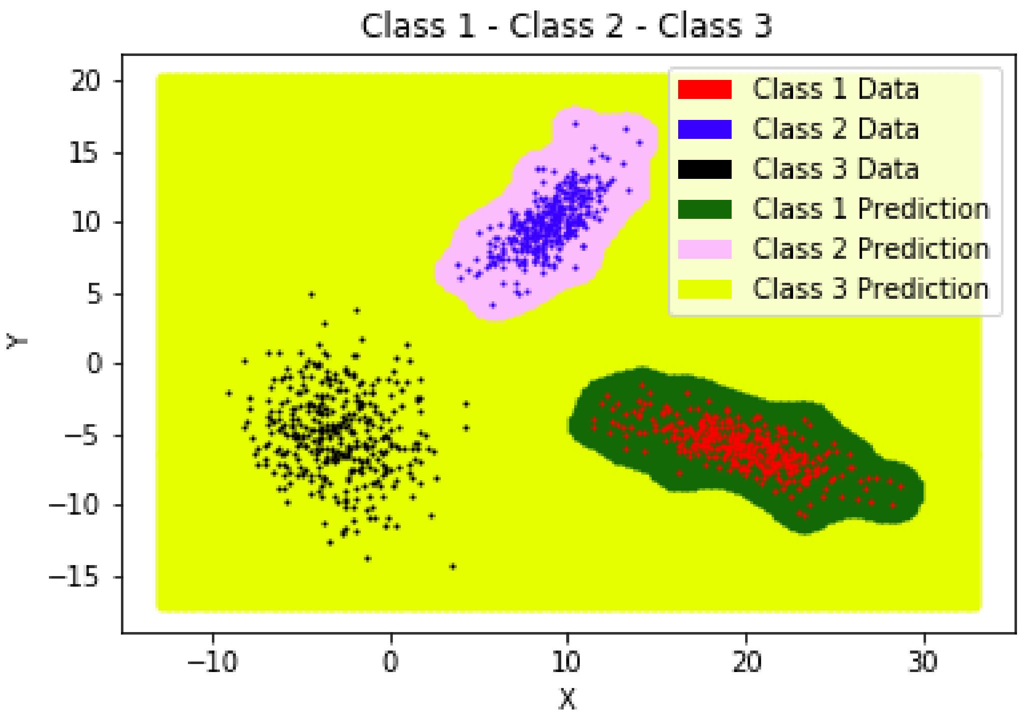


Figure 3..4. SVM RBF Kernel: Linearly Separable Data - Decision Regions

	Class 1	Class 1	Class 3
Class 1	125	0	0
Class 2	0	125	10
Class 3	0	0	125

(a) Confusion Matrix

	Class 1	Class 2	Class 3
Precision	1	1	1
Recall	1	1	1
F-Measure	1	1	1

(b) Analysis

Accuracy	100%
Precision	1
Recall	1
F-Measure	1

(c) Results

Table 3..4. SVM RBF Kernel: Linearly Separable Data - Results

Observations and Inferences

- Perceptron was able to correctly classify all the data
- SVM with a linear kernel gave a linear boundary which was better than that given by the perceptron. The boundaries were centered between the data points of the two corresponding classes.
- Polynomial as well as RBF kernels with SVM gave accurate results with a good approximation.
- 100% accuracy was obtained using both the perceptron as well as SVM with linear, polynomial, RBF kernels

Comparisons

For the artificial linearly separable dataset, Bayes classifier with unimodal gaussian distribution, GMM as well as perceptron and SVM based classification gave us a 100% accuracy. hence we can say that linearly separable classes are easy to classify and even simple classifiers can give good results.

2 Dataset 1(b): Artificial Non-Linearly separable dataset

2.1 Support Vector Machines

Linear Kernel

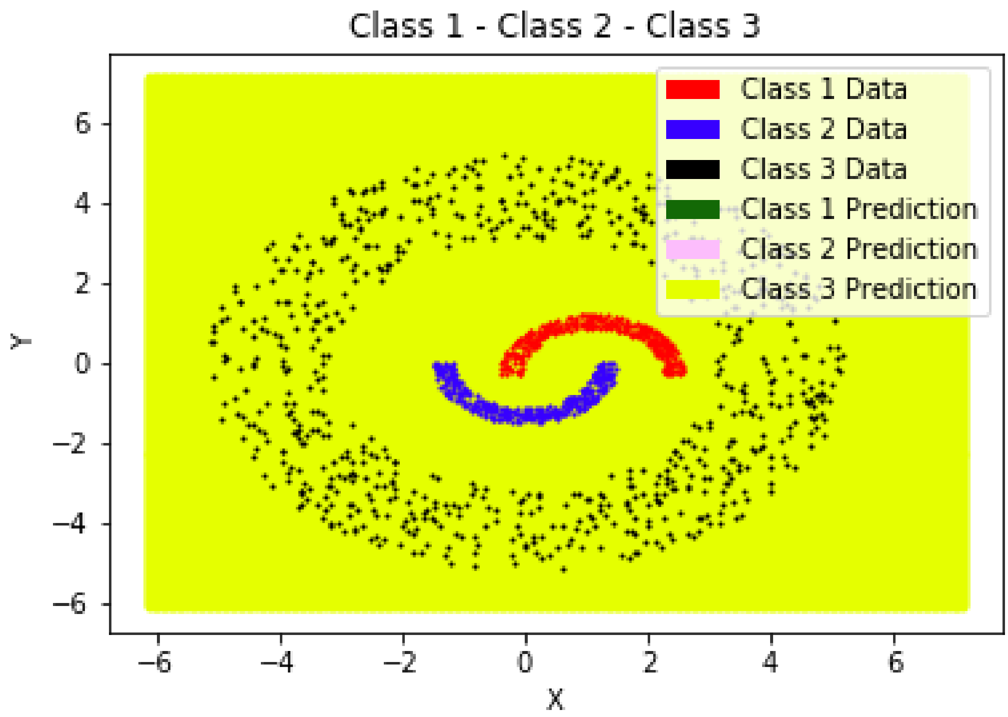


Figure 3..5. SVM Linear Kernel : Non-Linearly Separable Data - Decision Regions

	Class 1	Class 1	Class 3
Class 1	0	0	125
Class 2	0	0	125
Class 3	0	0	250

(a) Confusion Matrix

	Class 1	Class 2	Class 3
Precision	0	0	1
Recall	0	0	0.5
F-Measure	0	0	0.66

(b) Analysis

Accuracy	50%
Precision	0.33
Recall	0.166
F-Measure	0.22

(c) Results

Table 3..5. SVM Linear Kernel : Non-Linearly Separable Data - Results

Polynomial Kernel

Degree = 2

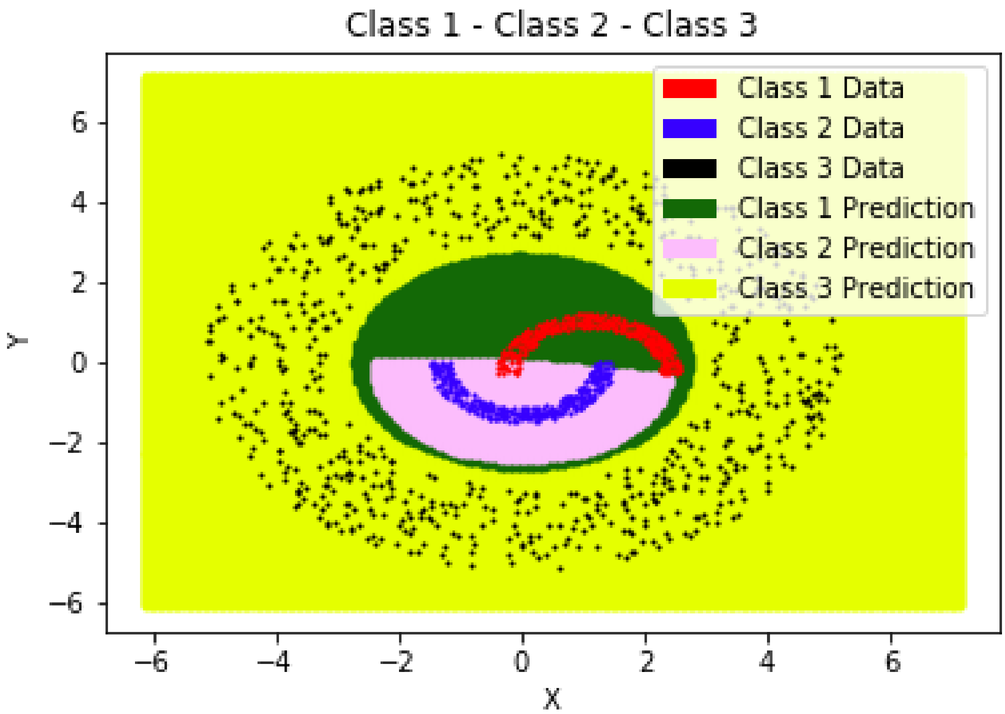


Figure 3..6. SVM Polynomial Kernel, Degree = 2 : Non-Linearly Separable Data - Decision Regions

	Class 1	Class 1	Class 3
Class 1	116	9	0
Class 2	9	116	10
Class 3	0	0	250

(a) Confusion Matrix

	Class 1	Class 2	Class 3
Precision	0.928	0.928	1
Recall	0.928	0.928	1
F-Measure	0.928	0.928	

(b) Analysis

Accuracy	96.4%
Precision	0.952
Recall	0.952
F-Measure	0.952

(c) Results

Table 3..6. SVM Polynomial Kernel, Degree = 2 :
Non-Linearly Separable Data - Results

Degree = 3

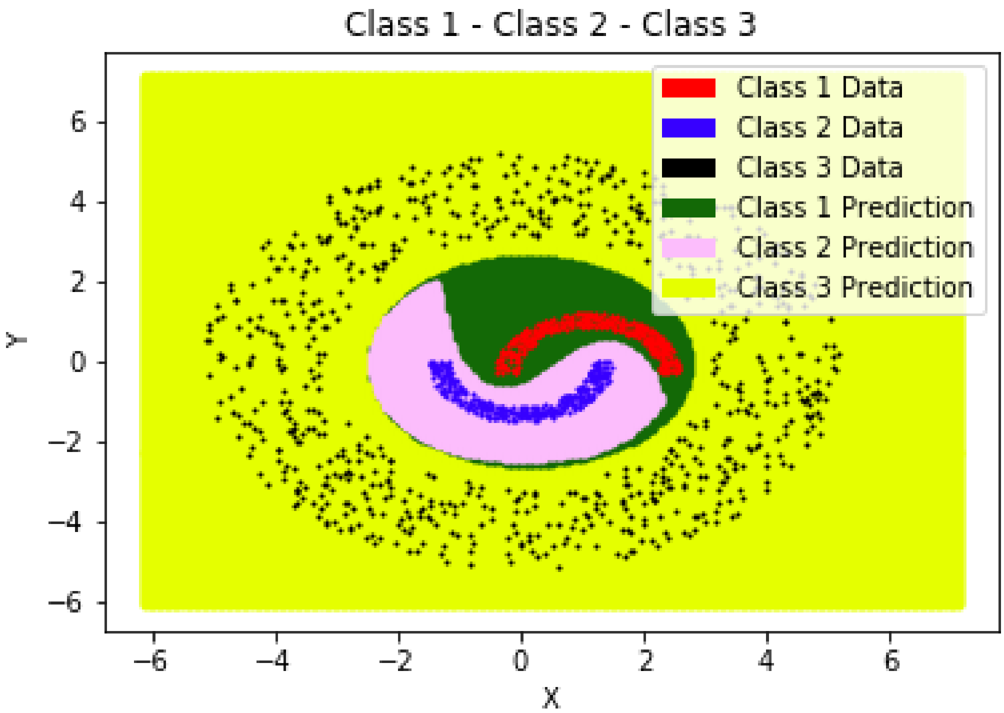


Figure 3..7. SVM Polynomial Kernel, Degree = 3 :
Non-Linearly Separable Data - Decision Regions

	Class 1	Class 1	Class 3
Class 1	125	0	0
Class 2	0	125	10
Class 3	0	0	250

(a) Confusion Matrix

	Class 1	Class 2	Class 3
Precision	1	1	1
Recall	1	1	1
F-Measure	1	1	1

(b) Analysis

Accuracy	100%
Precision	1
Recall	1
F-Measure	1

(c) Results

Table 3..7. SVM Polynomial Kernel, Degree = 3 :
Non-Linearly Separable Data - Results

RBF Kernel

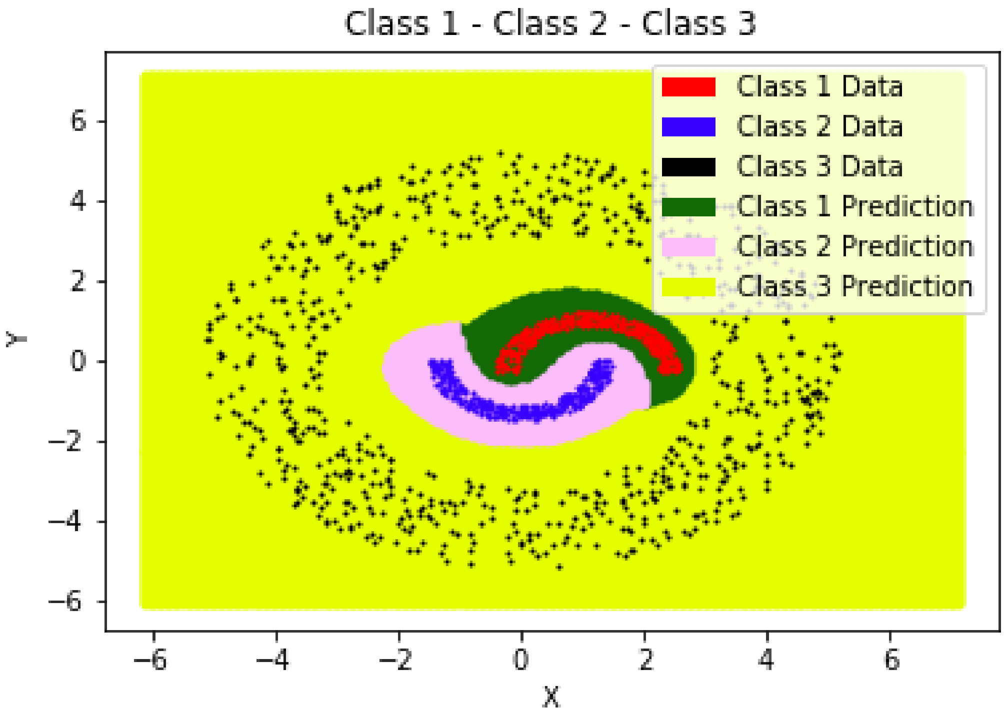


Figure 3..8. SVM RBF Kernel : Non-Linearly Separable
Data - Decision Regions

	Class 1	Class 1	Class 3
Class 1	125	0	0
Class 2	0	125	10
Class 3	0	0	250

(a) Confusion Matrix

	Class 1	Class 2	Class 3
Precision	1	1	1
Recall	1	1	1
F-Measure	1	1	1

(b) Analysis

Accuracy	100%
Precision	1
Recall	1
F-Measure	1

(c) Results

Table 3..8. SVM RBF Kernel : Non-Linearly Separable Data - Results

Comparisons

With unimodal gaussian distribution we achieved a maximum accuracy of 96%. Using Gaussiam Mixture Model based density estimation we were able to achieve 100%
Accuracy of 100% was also achieved using SVM with a polynomial kernel of degree of 2 as well as with the RBF kernel.

3 Dataset 2 : Scene Image Dataset

3.1 Classification using GMM on Color Histograms

	C = 1	C = 2	C = 4	C = 8	C = 16	C = 32
Accuracy	46.0%	24.66%	40.0%	45.33%	40.66%	47.99%
mean precision	0.5146	0.2368	0.3852	0.4769	0.4103	0.4908
mean recall	0.46	0.2466	0.4155	0.4533	0.4066	0.4799
mean F-Measure	0.4419	0.2237	0.3878	0.4397	0.4061	0.4779

Table 3..9. GMM Results for Color Histograms

3.2 Classification using GMM on BOVW representaion

	C = 1	C = 2	C = 4	C = 8	C = 16	C = 32
Accuracy	34.67	46.0	45.33	54.67	46.67	38.0
mean precision	0.3522	0.4706	0.4578	0.5498	0.4899	0.3731
mean recall	0.3467	0.46	0.4533	0.5467	0.4667	0.38
mean F-Measure	0.3014	0.4531	0.4326	0.5474	0.4616	0.3582

Table 3..10. GMM Results for BOVW

3.3 Principal Component Analysis

Variation of Accuracy with L and number of GMM components

	C = 1	C = 2	C = 4	C = 8
L = 2	34.67%	34.67%	34.0%	37.33%
L = 5	29.33%	43.33%	36.0%	40.67%
L = 10	30.67%	42.0%	34.67%	39.33%
L = 15	34.0%	44.67%	42.0%	43.33%
L = 21	30.67%	32.67%	45.33%	45.33%

Table 3..11. Variation of Accuracy with L and C

We observe that in general that accuracy increases with increase in L and the number of GMM components although there are a few spikes in occasionally.

3.4 Support Vector Machines

Linear Kernel

	Bayou	Chalet	Creek
Bayou	21	13	16
Chalet	19	21	10
Creek	20	8	22

(a) Confusion Matrix

	Bayou	Chalet	Creek
Precision	0.42	0.42	0.44
Recall	10.35	0.5	0.458
F-Measure	0.381	0.456	0.448

(b) Analysis

Accuracy	42.66%
Precision	42.6 6
Recall	43.61
F-Measure	42.91

(c) Results

Table 3..12. SVM Linear Kernel : Image BoVW Data - Results

Polynomial Kernel

Degree = 2

	Bayou	Chalet	Creek
Bayou	30	10	10
Chalet	20	21	9
Creek	16	18	16

(a) Confusion Matrix

	Bayou	Chalet	Creek
Precision	0.6	0.42	0.32
Recall	0.45	0.42	0.45
F-Measure	0.51	0.42	0.37

(b) Analysis

Accuracy	44.6%
Precision	0.446
Recall	0.446
F-Measure	0.439

(c) Results

Table 3..13. SVM Polynomial Kernel, Degree = 2 : Image BoVW Data - Results

Degree = 3

	Bayou	Chalet	Creek
Bayou	31	11	8
Chalet	29	24	7
Creek	14	18	18

(a) Confusion Matrix

	Bayou	Chalet	Creek
Precision	0.62	0.48	0.36
Recall	0.48	0.45	0.54
F-Measure	0.54	0.46	0.43

(b) Analysis

Accuracy	48.6%
Precision	0.486
Recall	0.494
F-Measure	0.481

(c) Results

Table 3..14. SVM Polynomial Kernel, Degree = 3 : Image BoVW Data - Results

RBF Kernel

	Bayou	Chalet	Creek
Bayou	46	4	0
Chalet	42	7	1
Creek	46	4	0

(a) Confusion Matrix

	Bayou	Chalet	Creek
Precision	0.92	0.14	0
Recall	0.34	0.46	0
F-Measure	0.5	0.21	nan

(b) Analysis

Accuracy	0.35%
Precision	0.35
Recall	0.26
F-Measure	nan

(c) Results

Table 3..15. SVM RBF Kernel : Image BoVW Data - Results

Comparisons

The best accuarcy was obtained using a GMM based desnity estimation. Accuracy using a PCA based dimensionality reduction also gave us comparable results but not better.

SVM with a degree 3 polynomial kernel gave an accuracy of 48% which is comparable to the one obtained with GMM based density estimation.

4. Conclusion & Inferences

- Even a single perceptron was able to efficiently classify multiple linearly separable classes.
- SVM gave a better decision boundary in in comparison to a single perceptron.
- SVM with a polynomial or RBF kernel is in general more effective than that using a linear kernel for non linearly separable data.
- Increasing the degree of the polynomial kernel in SVM increases accuracy to some extent.
- For the scene image dataset, SVM was not much effective than the previously used PCA and GMM combinations.
- For scene image dataset, accuracy was observed to increase with increase in the degree of the polynomial kernel.

Bibliography

[1] Bayes Classifier
https://en.wikipedia.org/wiki/Bayes_classifier
https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysi
https://onlinecourses.science.psu.edu/stat505/node/49/
https://medium.com/@aptrishu/understanding-principle-component-analysis-e32be0

[2] Stack Overflow
https://stackoverflow.com

[3] Naive Bayes Classifier
https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[4] K-Means clustering
https://en.wikipedia.org/wiki/K-means_clustering
https://www.geeksforgeeks.org/k-means-clustering-introduction/
https://towardsdatascience.com/understanding-k-means-clustering-\
in-machine-learning-6a6e67336aa1

[5] Matplotlib Contours
https://matplotlib.org/api/_as_gen/matplotlib.pyplot.contour.
html

[6] K-Nearest Neighbor Method
https://stats.stackexchange.com/questions/252852/
k-value-vs-accuracy-in-knn

[7] Baum Welch Algorithm
https://en.wikipedia.org/wiki/Baum\OT1\textendashWelch_
algorithm

[8] Fisher Discriminant Analysis
https://en.wikipedia.org/wiki/Linear_discriminant_analysis