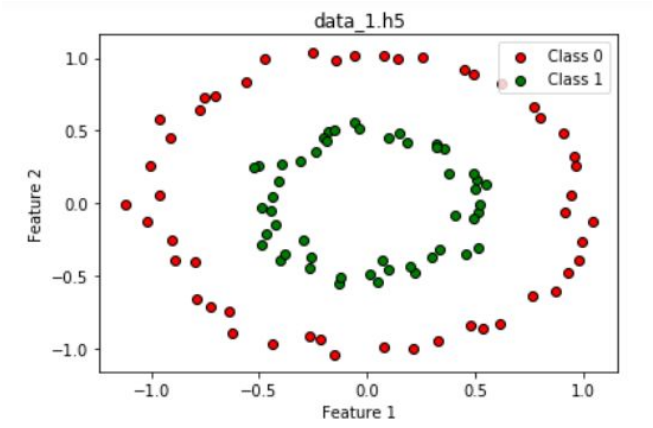
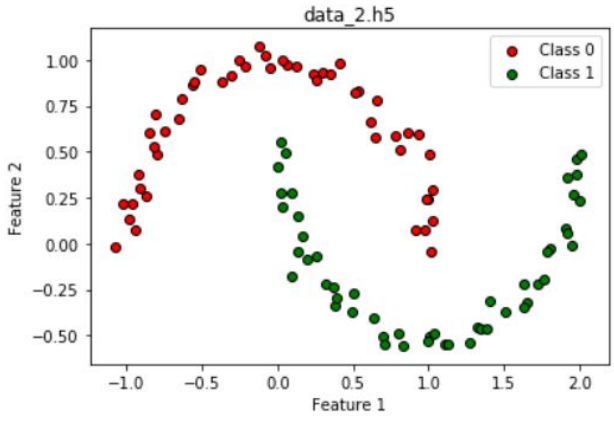
ML Assignment 2 : Report

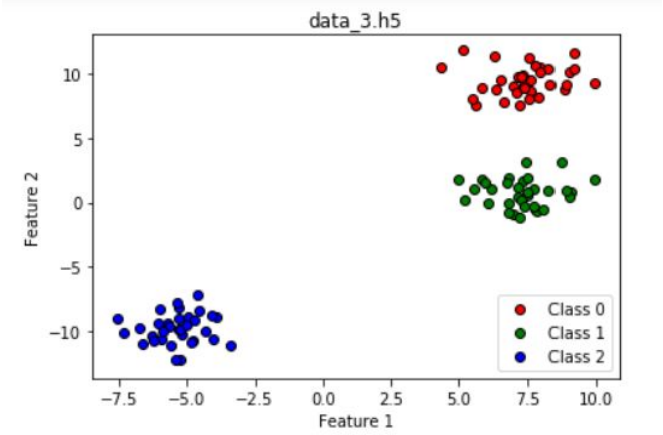
Anand, 2017218

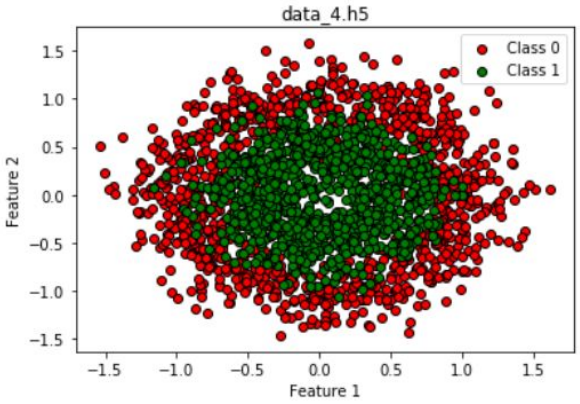
Q1 Part 1,2,3 : **Q1_Part_1_2_3.py**
Q1 Part 4 : **Q1_Part_4.py**
Q1 2 : **Q2_CIFAR.py**

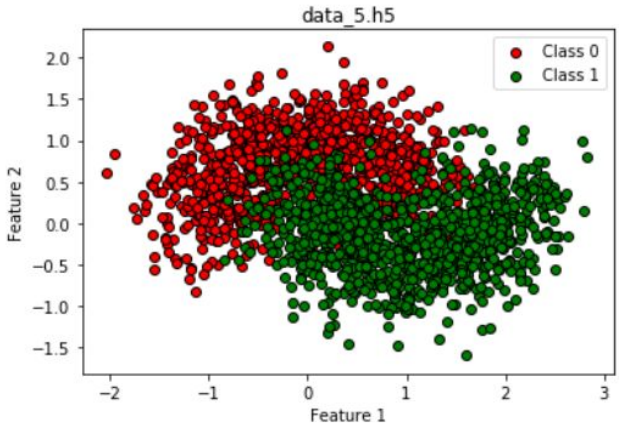
Q1- Part1: Plotting the data:







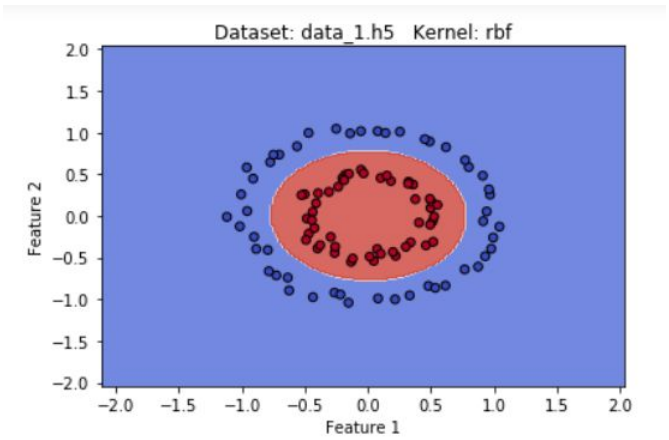


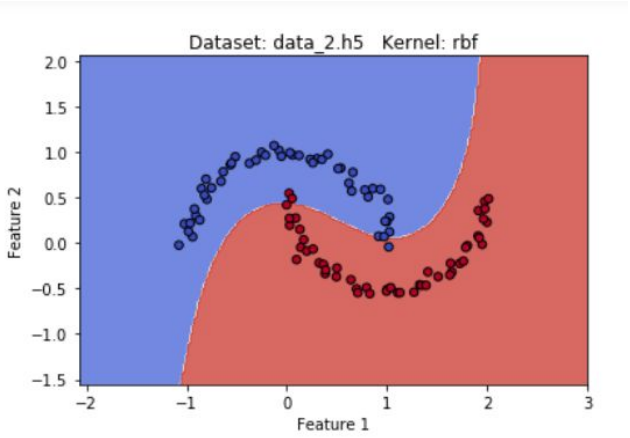


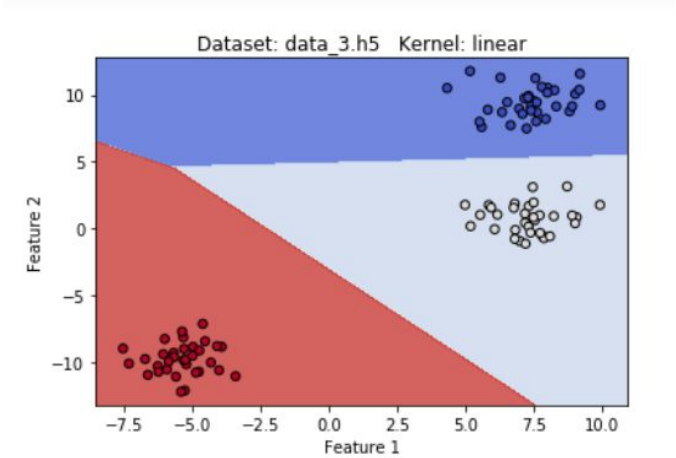
Observations and findings:

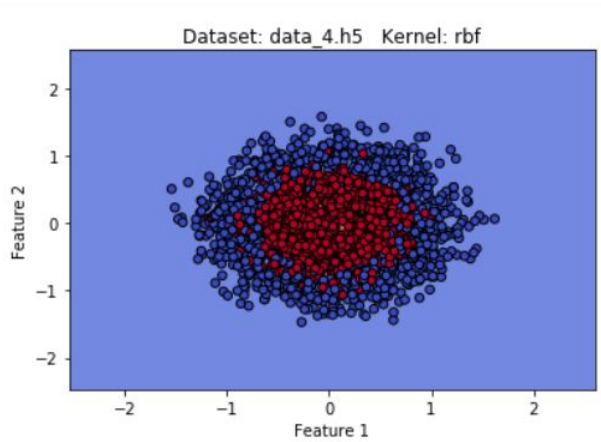
i) Dataset 1 has very less noise and seems that hyperplane would be some ellipse or circle.
ii) Dataset 2 has very less noise and seems that hyperplane would be a cubic polynomial
iii) Dataset 3 also has less noise and it is clear that it is linearly separable.
iv) Dataset 4 has some noise and is linearly inseparable.
v) Dataset 5 has some noise and is linearly inseparable.

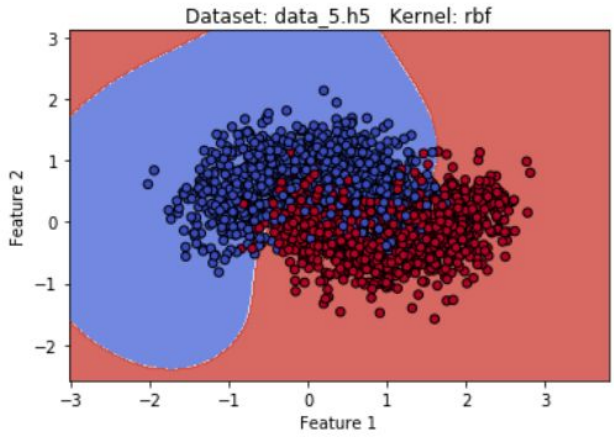
Q1- Part2 : Decision boundaries









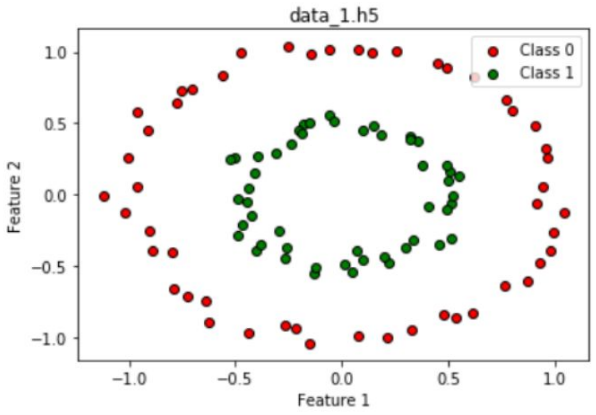


Explanation for choice of kernels:

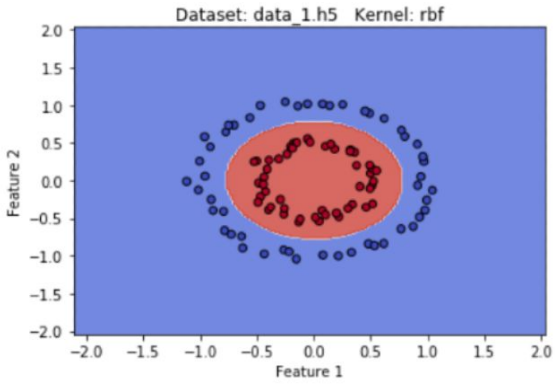
i) Dataset 1,2,4,5: RBF : as it perfectly separates the classes and gives the highest accuracy.
ii) Dataset 3: Linear kernel: the data was linearly separable and hence linear kernel.

Q1 - Part 3 - Outlier removed dataset

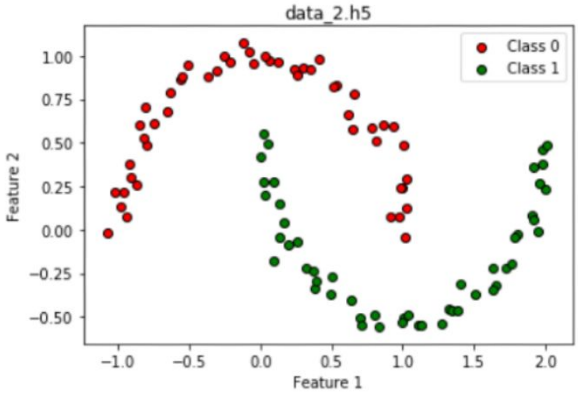
```
Original No of Examples:  100
No of Examples after removing outliers:  100
```



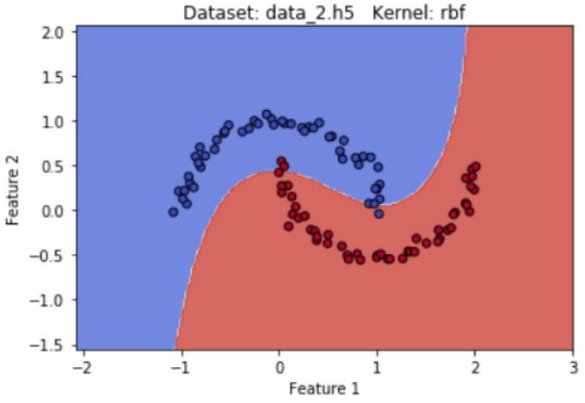

```
Original No of Examples:  100
No of Examples after removing outliers:  100
```



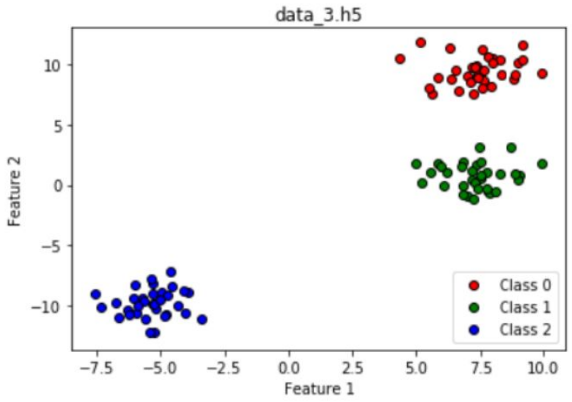

```
Original No of Examples:  100
No of Examples after removing outliers:  100
```



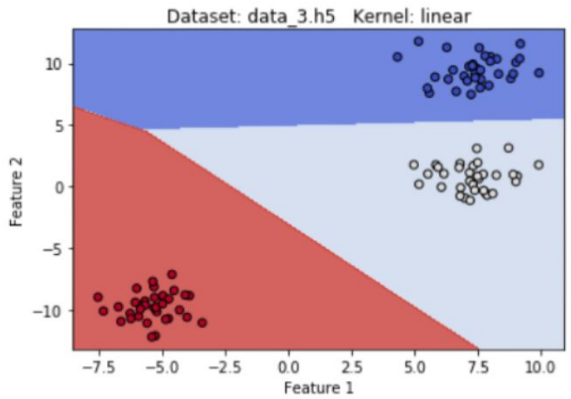

```
Original No of Examples:  100
No of Examples after removing outliers:  100
```



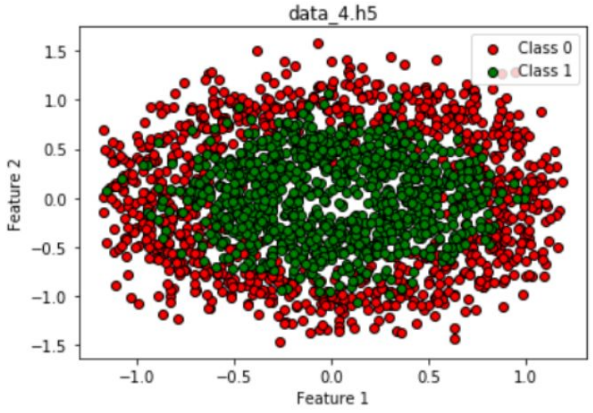
```
Original No of Examples:  100
No of Examples after removing outliers:  100
```



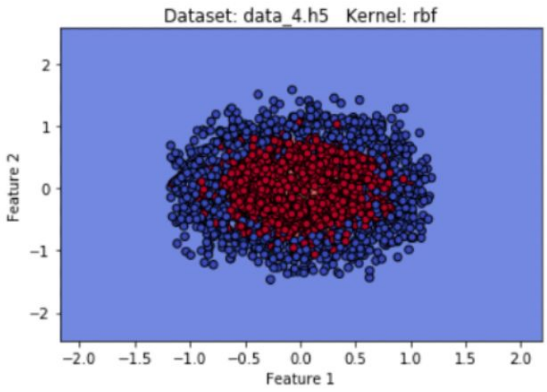

```
Original No of Examples:  100
No of Examples after removing outliers:  100
```



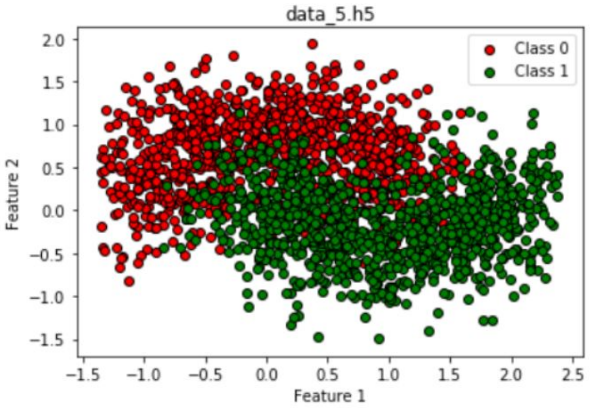
```
Original No of Examples:  2000
No of Examples after removing outliers:  1946
```



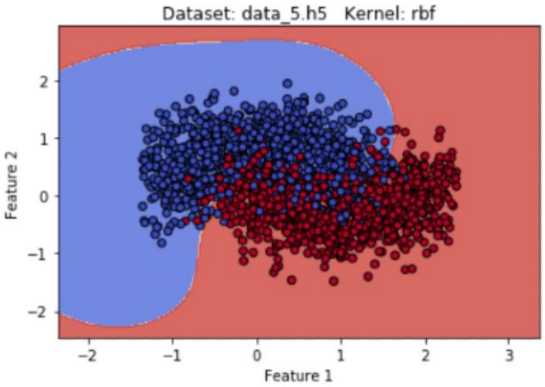

```
Original No of Examples:  2000
No of Examples after removing outliers:  1946
```




```
Original No of Examples:  2000
No of Examples after removing outliers:  1938
```




```
Original No of Examples:  2000
No of Examples after removing outliers:  1938
```



Outliers are removed using normal distribution considering the fact that most of the data which are **not** outliers lie between mean-3*standard_deviation to mean+3*standard_deviation.

Q1 - Part 4

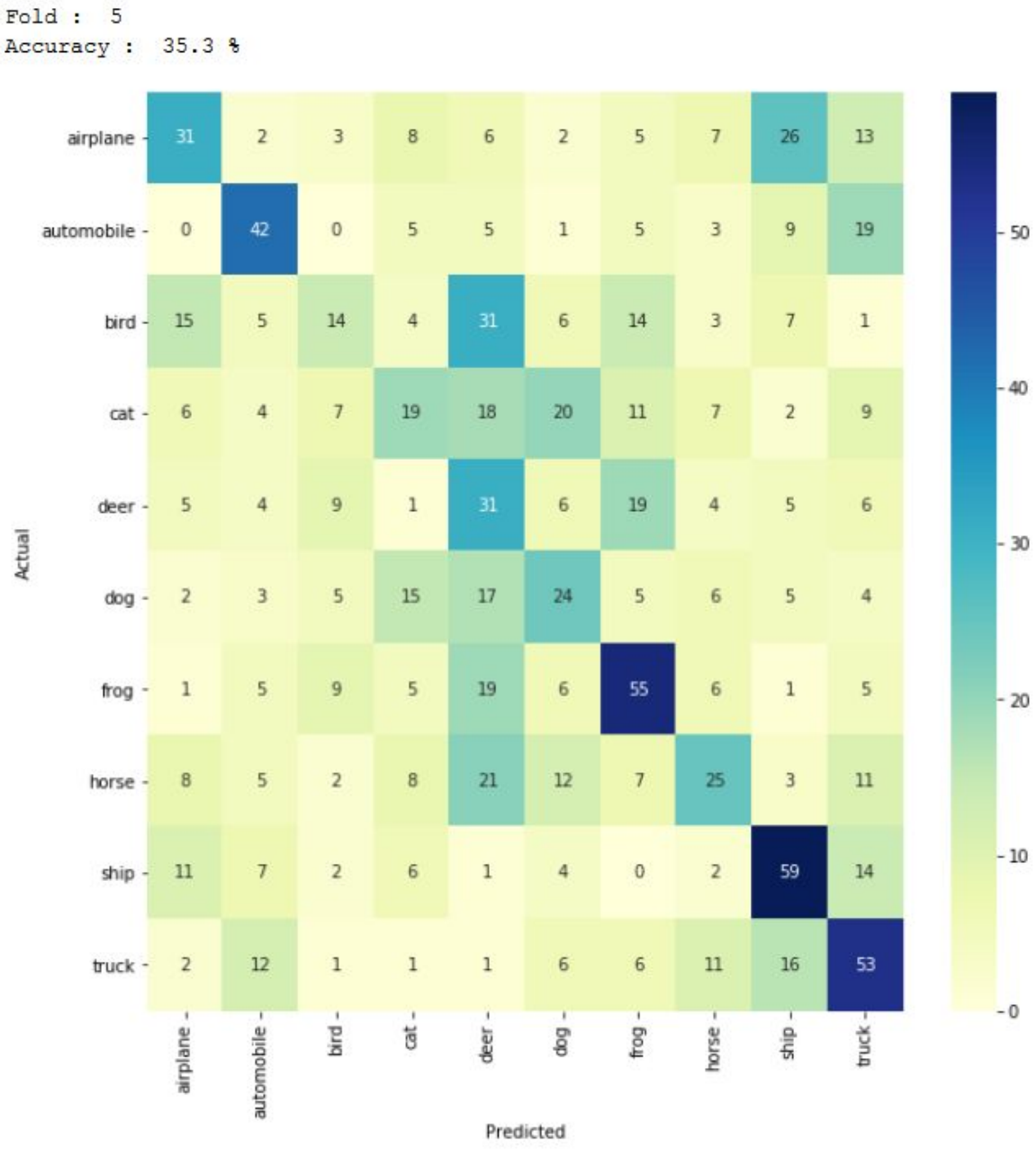
```
DATASET 4
--------------------------
Accuracy using Linear Kernel:
Accuracy using SVM prediction function =  53.58974358974359
Accuracy using own prediction function =  50.0
Accuracy using RBF Kernel:
Accuracy using SVM prediction function =  88.71794871794872
Accuracy using own prediction function =  87.6923076923077

DATASET 5
--------------------------
Accuracy using Linear Kernel:
Accuracy using SVM prediction function =  82.21649484536083
Accuracy using own prediction function =  81.95876288659794
Accuracy using RBF Kernel:
Accuracy using SVM prediction function =  82.9896907216495
Accuracy using own prediction function =  77.83505154639175
```

Q2 : SVM ON CIFAR 10 DATASET

RBF kernel performs the best.

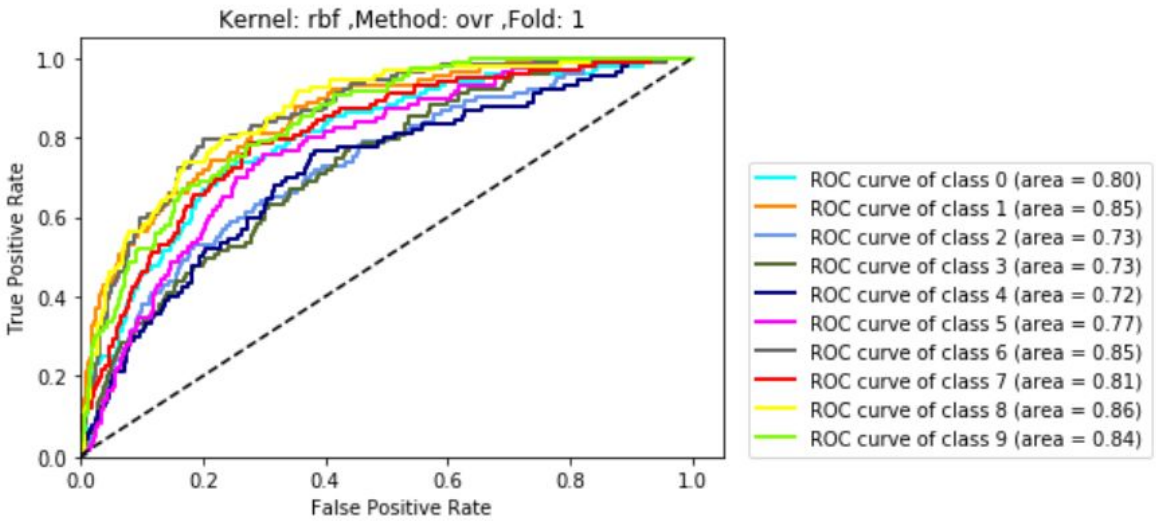
Highest accuracy of this model is: 35.3%

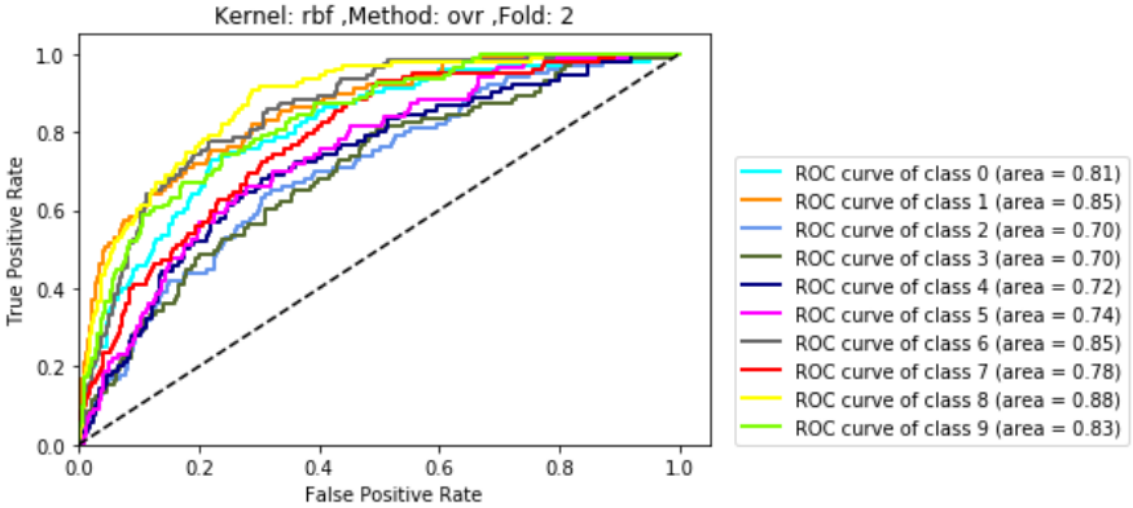
❑ It's Confusion Matrix:

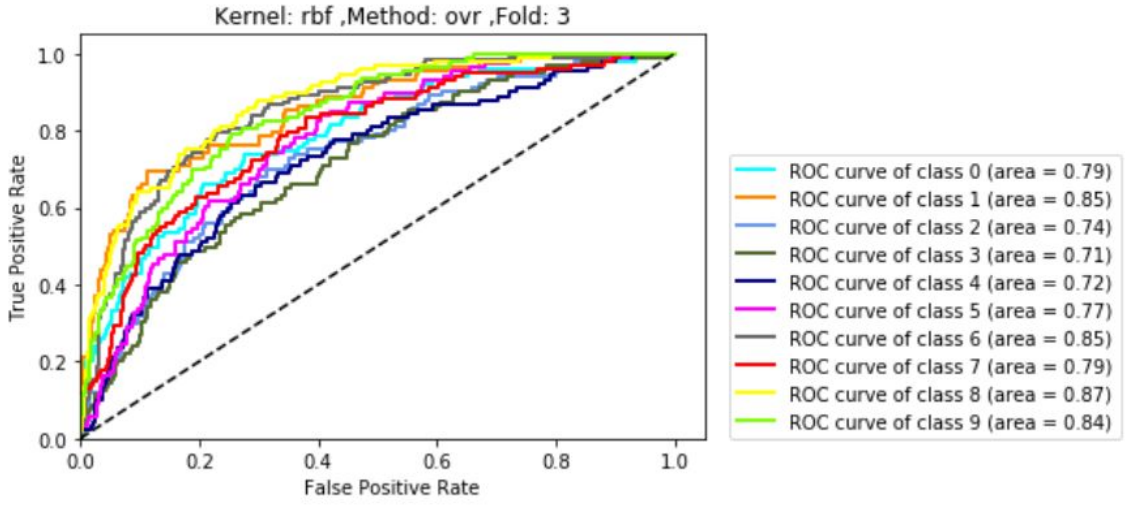


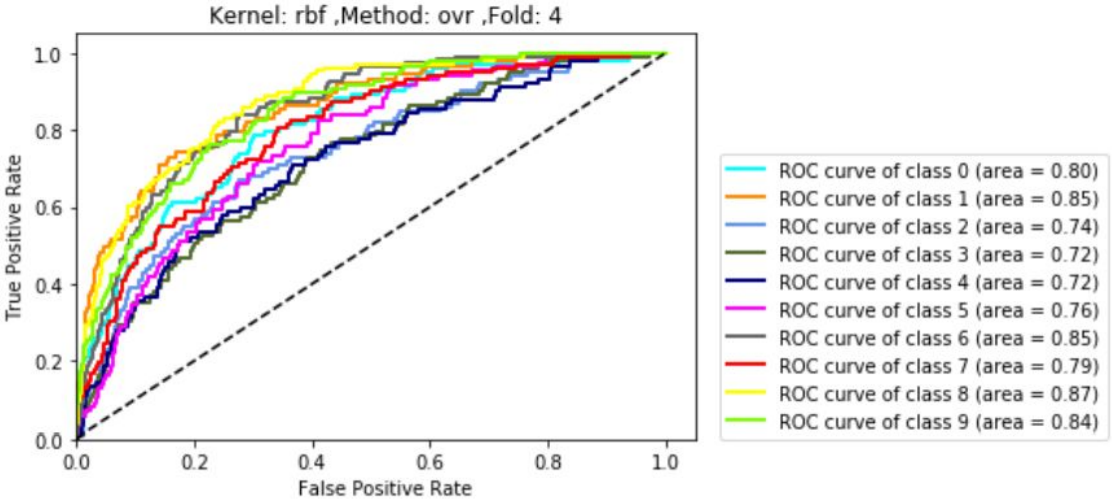
❑ It's ROC Curve: (For one-vs-all classification method over 5 folds)

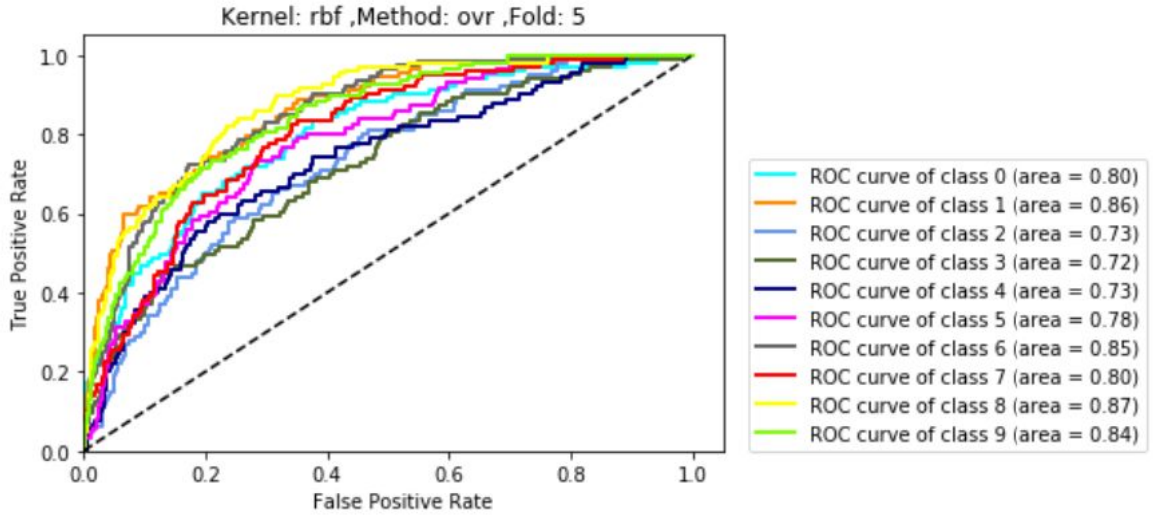
All other ROC curves are submitted in the project folder.











NOTE: ALL OTHER DIAGRAMS FOR ALL FOLDS AND ALL METHODS ARE SUBMITTED IN THE PROJECT FOLDER.

Accuracy report:

i) No Kernel:	29.7%
ii) RBF Kernel:	35.3%
iii) Quadratic Polynomial Kernel:	31.8%

Clearly RBF Kernel performs the best as it gives highest accuracy and area under the curve in ROC.