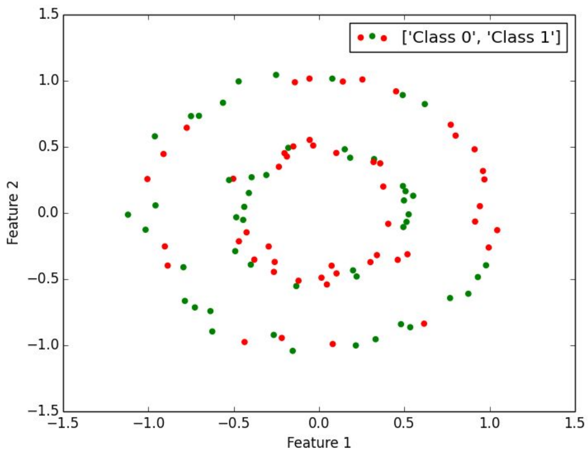
Analysis Report

Vasisht Duddu

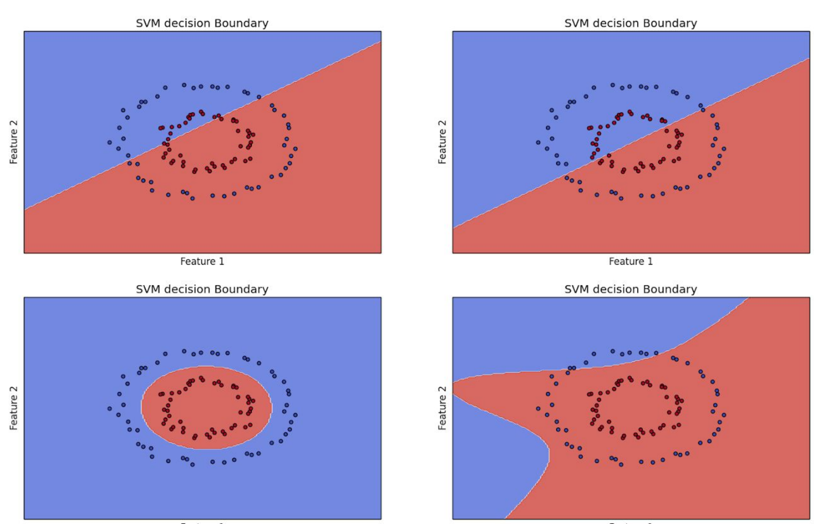
2015137

Question 1

Data 1

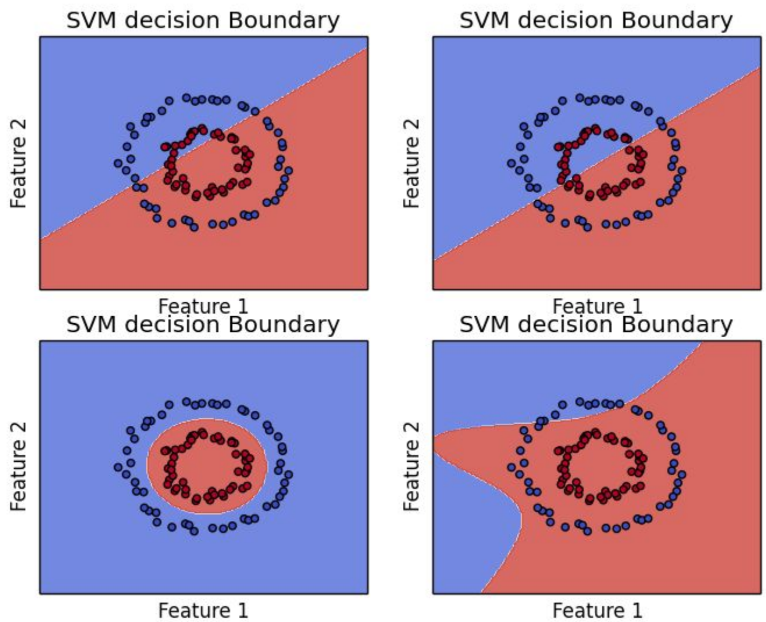


The data is not linearly separable and hence we need to use kernel to separate them.



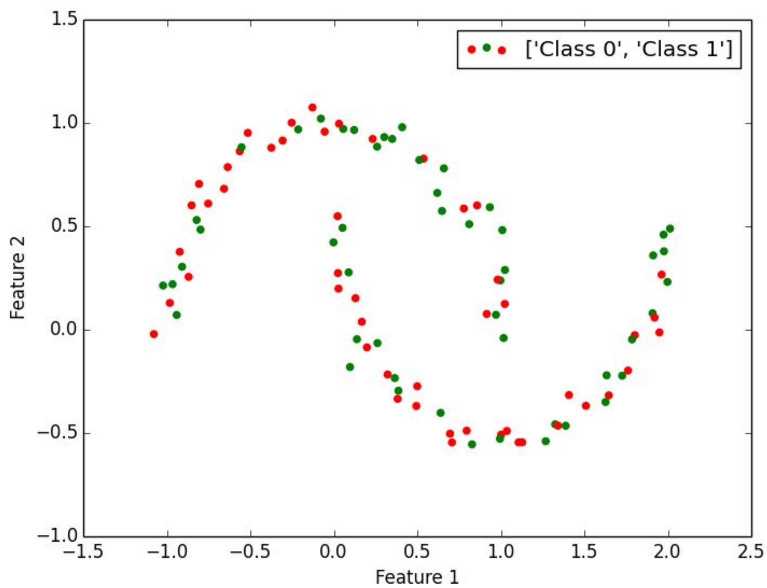
The top two images show the linear kernel and Linear SVM classifier while the bottom left is decision boundaries corresponding to rbf kernel while the image in the bottom right is polynomial kernel with degree 3.

Since the points are in a circular shape, radial basis functions perfectly separate the different points.

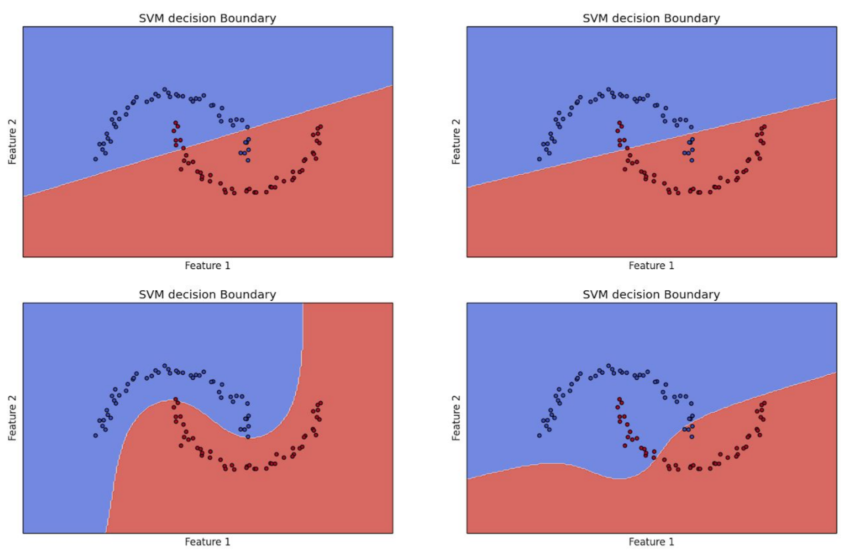


The above image shows the data after removing the outliers. The outliers are considered the points not between +-3*(mean-std. deviation).

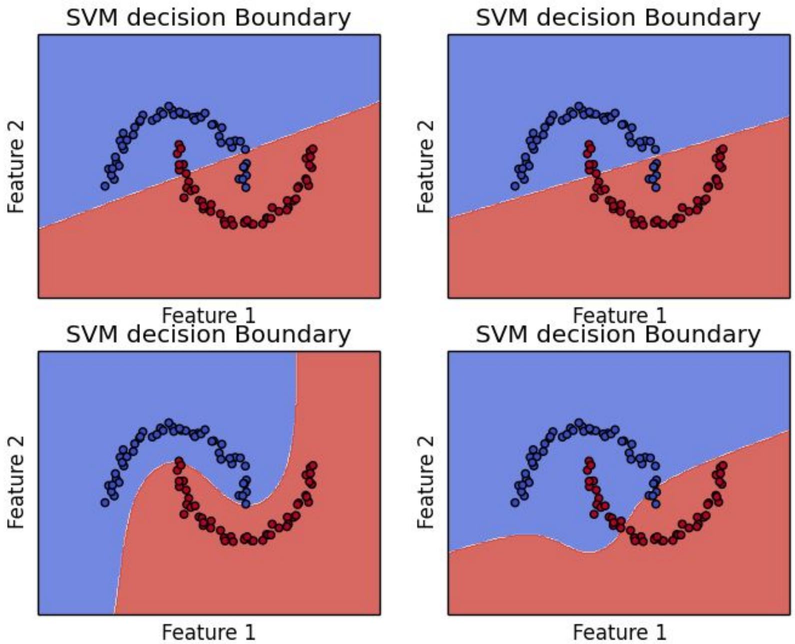
Data 2



The data is not linearly separable and hence we need to use kernel functions to separate them.

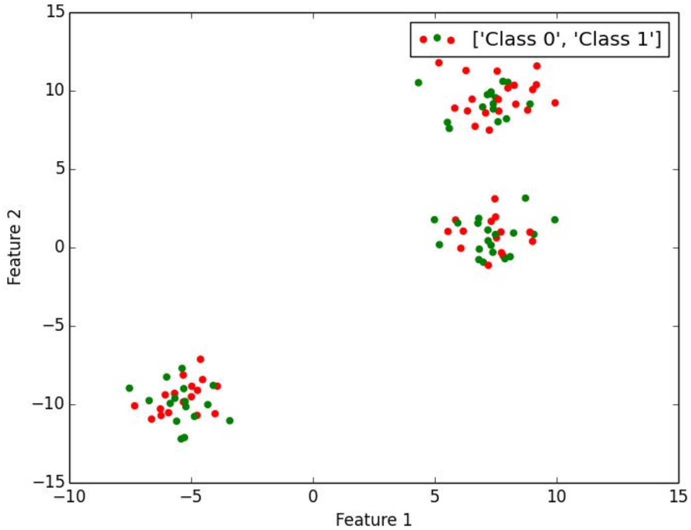


The top two images show the linear kernel and Linear SVM classifier while the bottom left is decision boundaries corresponding to rbf kernel while the image in the bottom right is polynomial kernel with degree 3.
The rbf kernel perfectly fits the data and it is better to use rbf kernel for seperating the data.

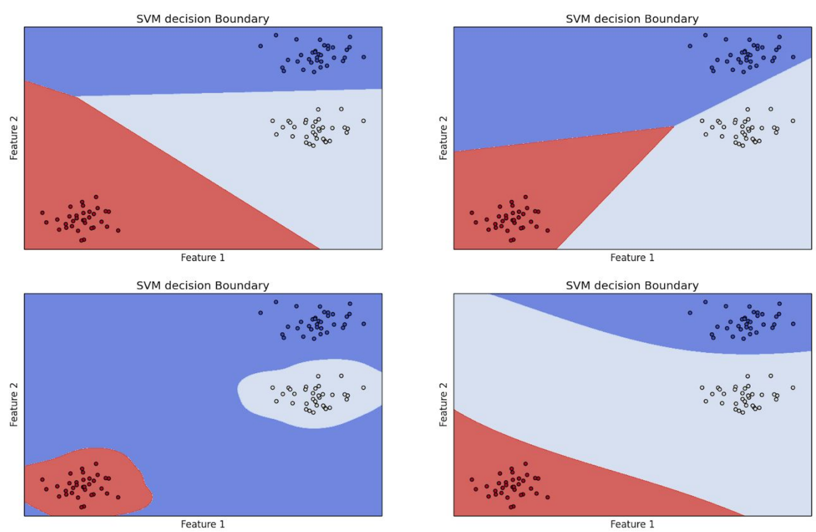


The above image shows the data after removing the outliers. The outliers are considered the points not between +-3*(mean-std. deviation).

Data 3

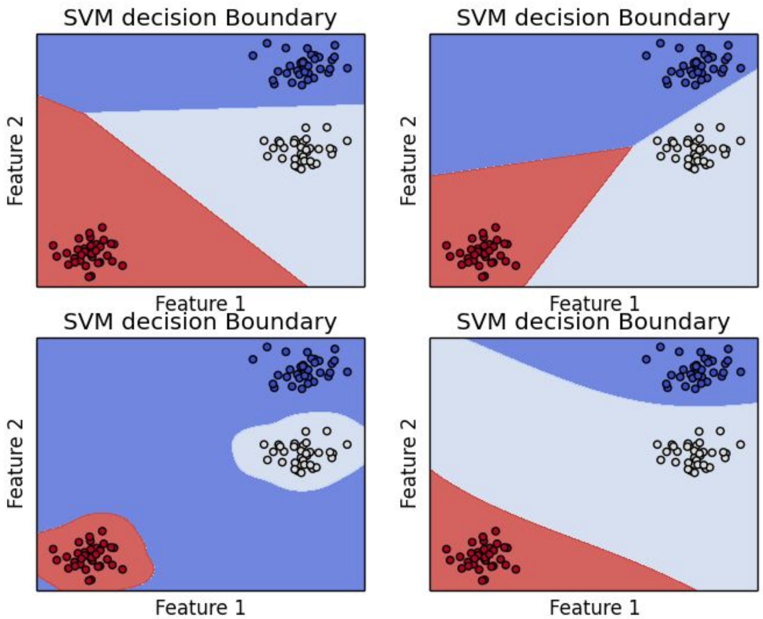


The data can be easily separated and different kernels and corresponding decision boundaries are shown below.



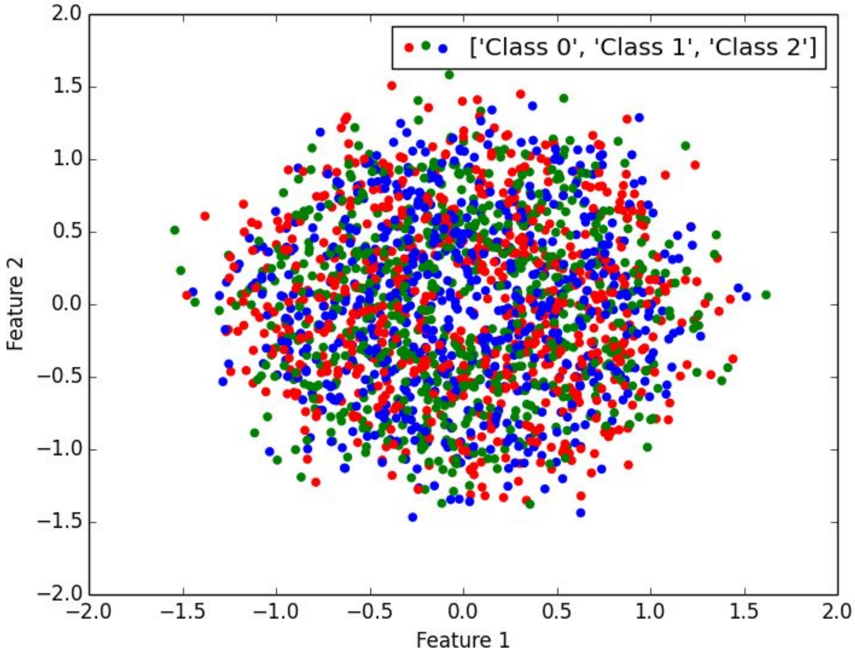
The top two images show the linear kernel and Linear SVM classifier while the bottom left is decision boundaries corresponding to rbf kernel while the image in the bottom right is polynomial kernel with degree 3.

Clearly the linear kernel will be the most optimal as the overall boundaries are most optimal.

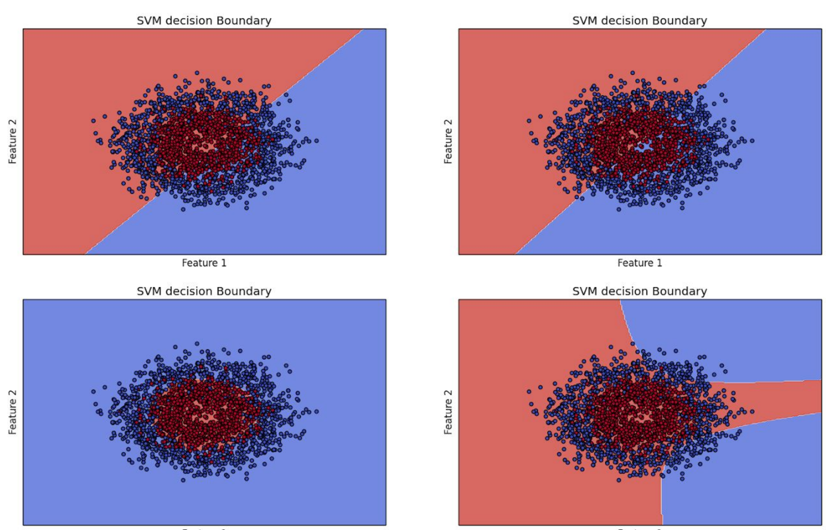


The above image shows the data after removing the outliers. The outliers are considered the points not between +-3*(mean-std. deviation).

Data 4

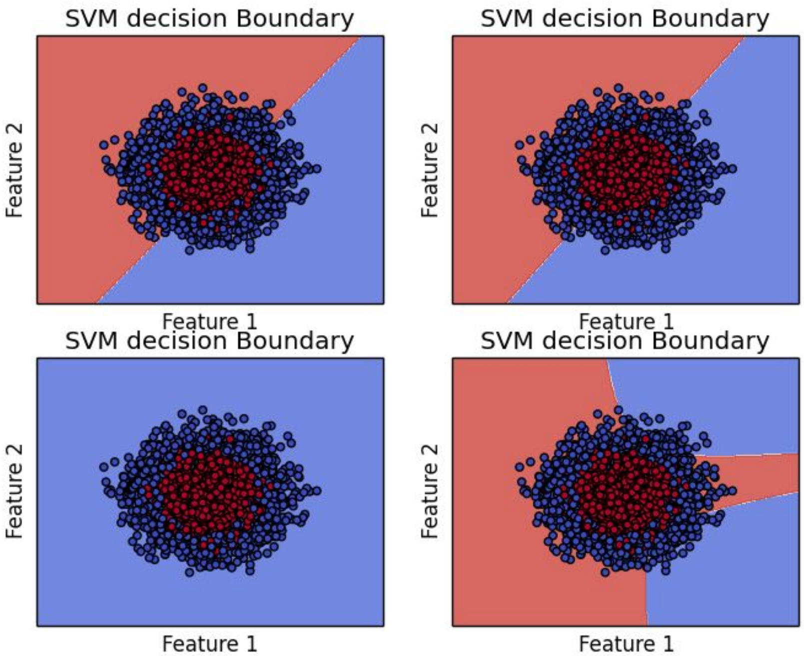


The data is not linearly separable and we need to use the



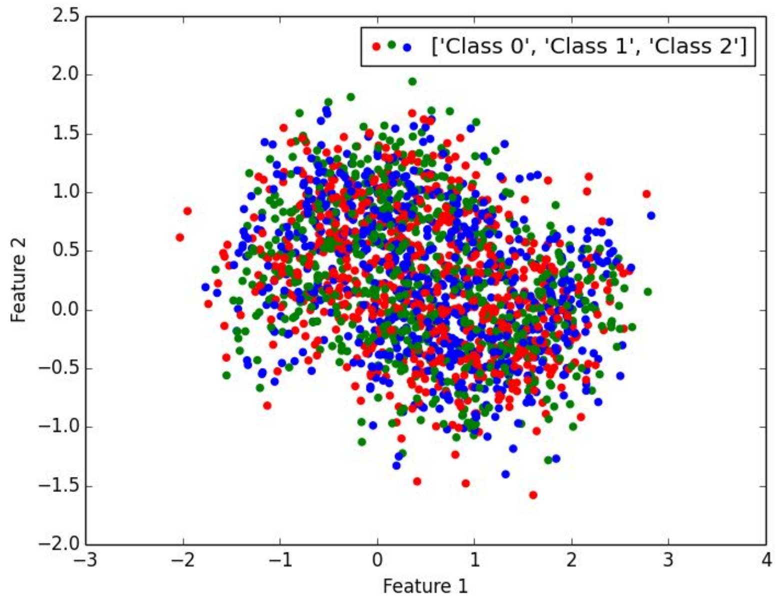
The top two images show the linear kernel and Linear SVM classifier while the bottom left is decision boundaries corresponding to rbf kernel while the image in the bottom right is polynomial kernel with degree 3.

The rbf seems to separate the data much better than other kernels.

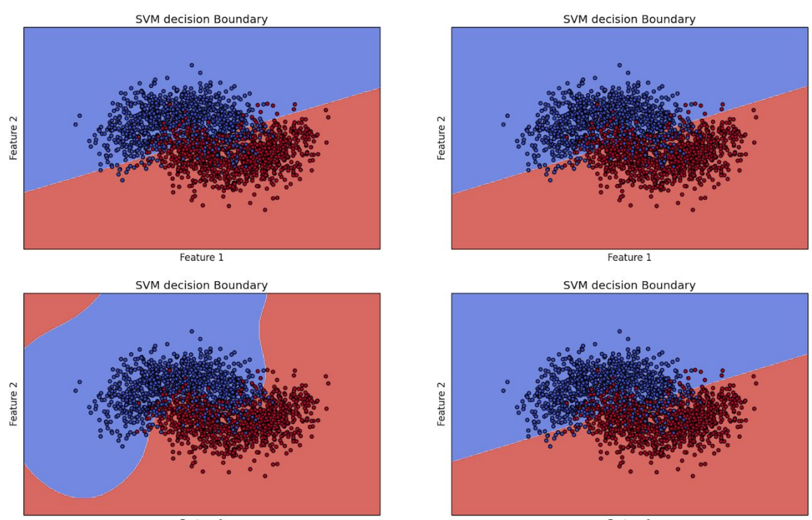


The above image shows the data after removing the outliers. The outliers are considered the points not between +-3*(mean-std. deviation).

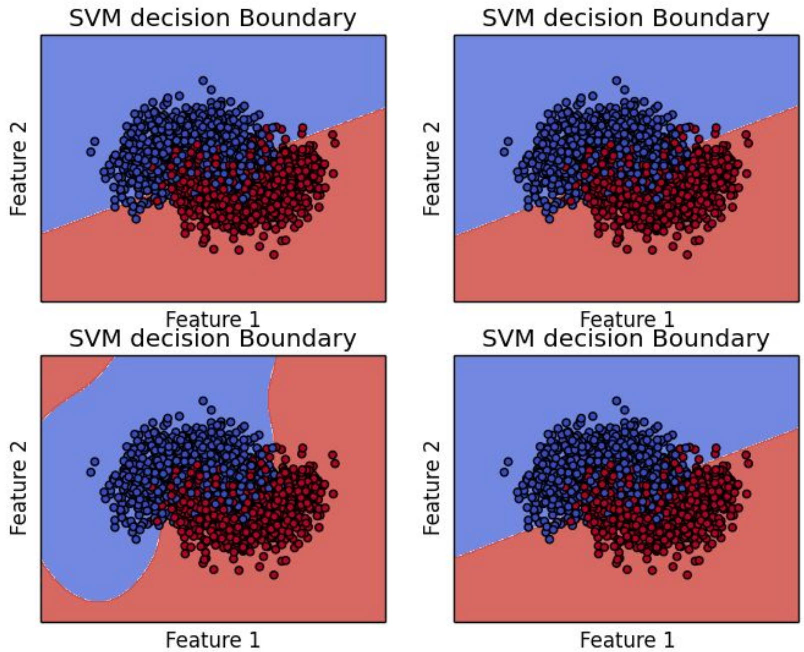
Data 5



The data is not linearly separable and hence we need to use kernels to move them to a higher dimensional space.



The top two images show the linear kernel and Linear SVM classifier while the bottom left is decision boundaries corresponding to rbf kernel while the image in the bottom right is polynomial kernel with degree 3.



The above image shows the data after removing the outliers. The outliers are considered the points not between +-3*(mean-std. deviation).

Question 2

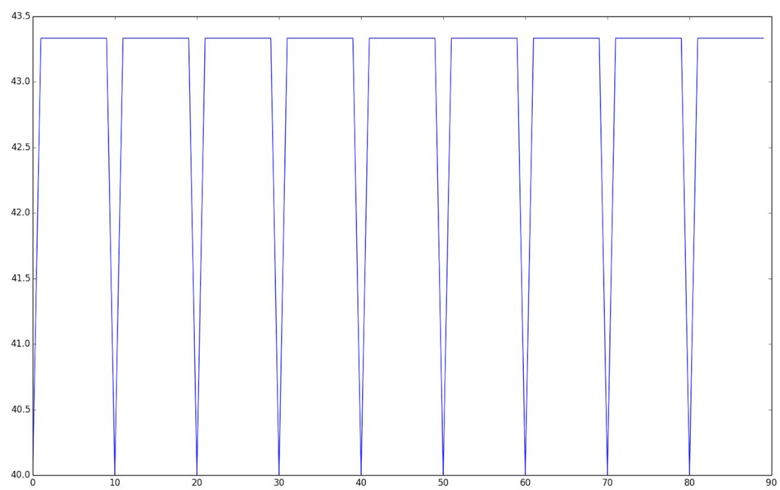
Data 1

Confusion Matrix

```
[[17  0]
 [13  0]]
```

TP=confusion[1,1]
TN=confusion[0,0]
FP=confusion[0,1]
FN=confusion[1,0]

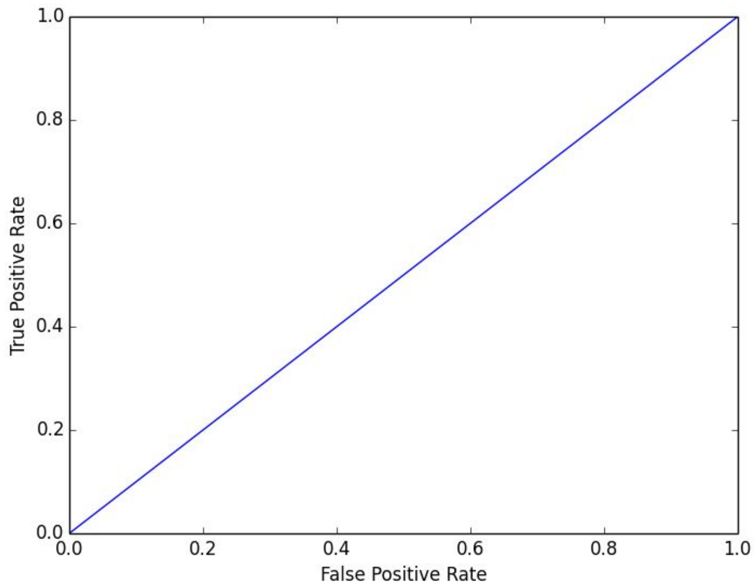
Soft margin with Linear Kernel



The plot shows different accuracies achieved by passing different parameters. To select the right parameters:
-Randomized method to train and test data was used
-Grid Search for different values of C and gamma

The right set of parameters can be chosen corresponding to the maximum parameters.

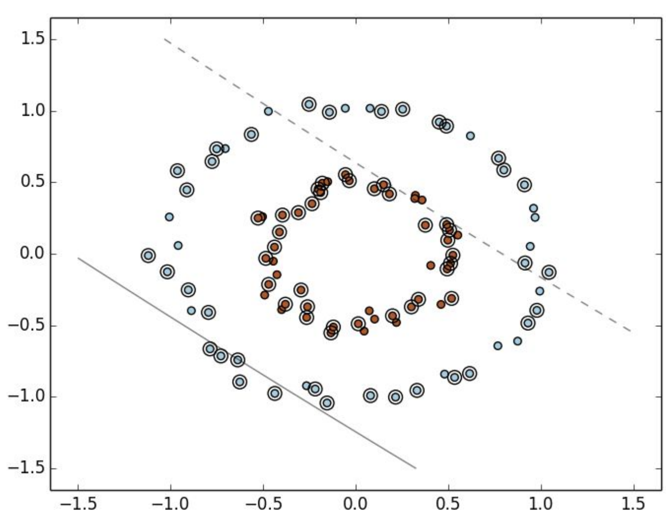
ROC Curve



It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity)

The area under the curve is a measure of text accuracy.

Plot showing Support Vectors and Classifier



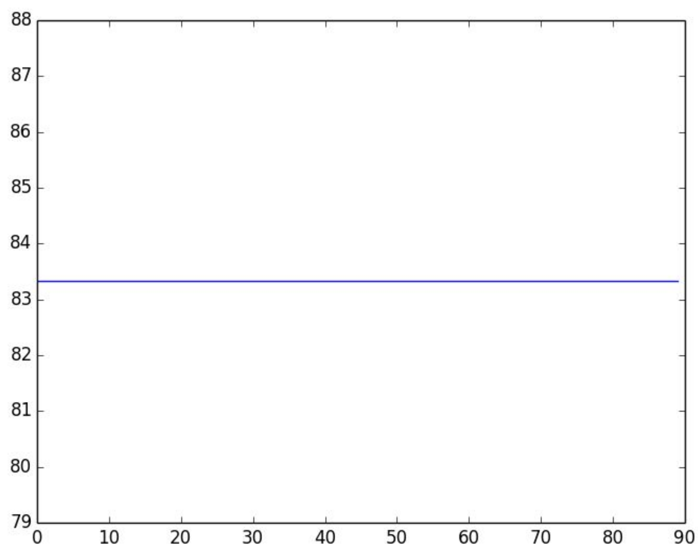
Data 2

Confusion Matrix

```
[[ 1 14]
 [10  5]]
```

TP=confusion[1,1]
TN=confusion[0,0]
FP=confusion[0,1]
FN=confusion[1,0]

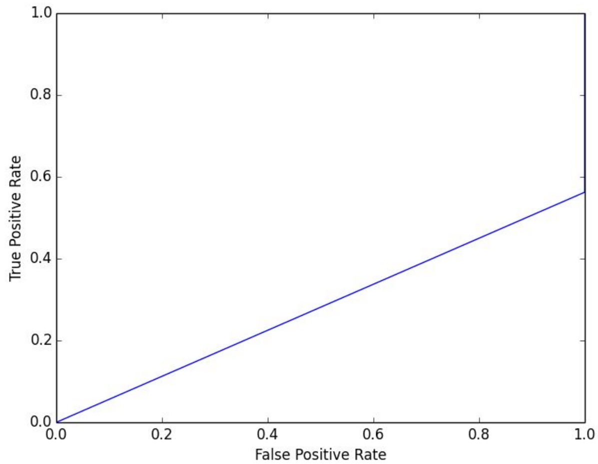
Soft margin with Linear Kernel



The plot shows different accuracies achieved by passing different parameters. To select the right parameters:
-Randomized method to train and test data was used
-Grid Search for different values of C and gamma

The right set of parameters can be chosen corresponding to the maximum parameters.

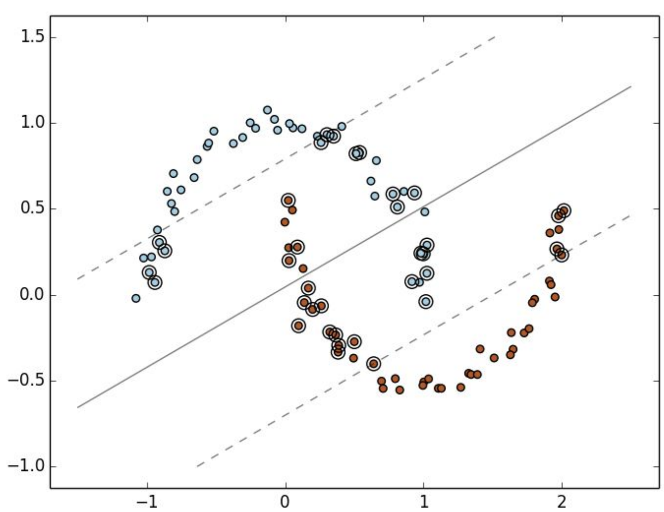
ROC Curve



It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity)

The area under the curve is a measure of text accuracy

Plot showing Support Vectors and Classifier



Data 3

Confusion Matrix

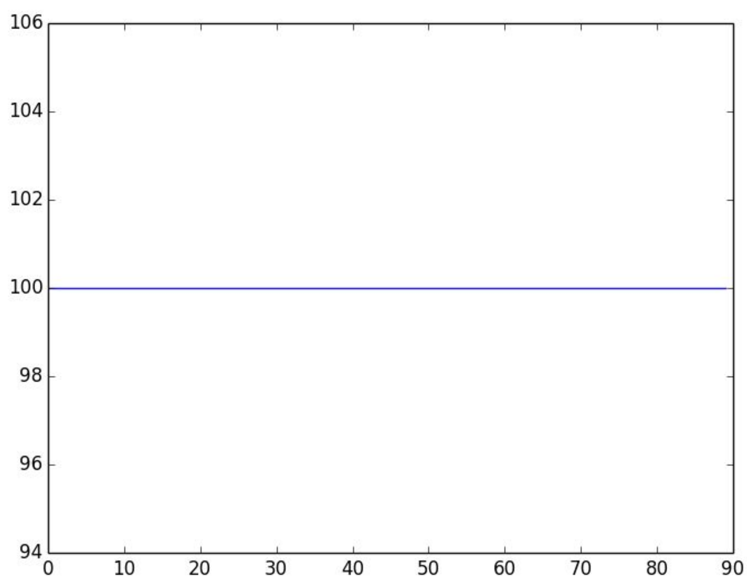
```
[[11  0  0]
 [ 0  8  0]
 [ 0  0 11]]
```

The confusion matrix is 3 by 3 as the dataset has three different classes. The corresponding values can be determined by the image given below.

Along with a high accuracy, we want to ensure a low false positive.

		predicted class		
		class 1	class 2	class 3
actual class	class 1	True positives		
	class 2		True positives	
	class 3			True positives

Soft margin with Linear Kernel



The plot shows different accuracies achieved by passing different parameters. To select the right parameters:
-Randomized method to train and test data was used
-Grid Search for different values of C and gamma

The right set of parameters can be chosen corresponding to the maximum parameters. The accuracy across all the different parameters is 100% as the SVM separates the data perfectly.

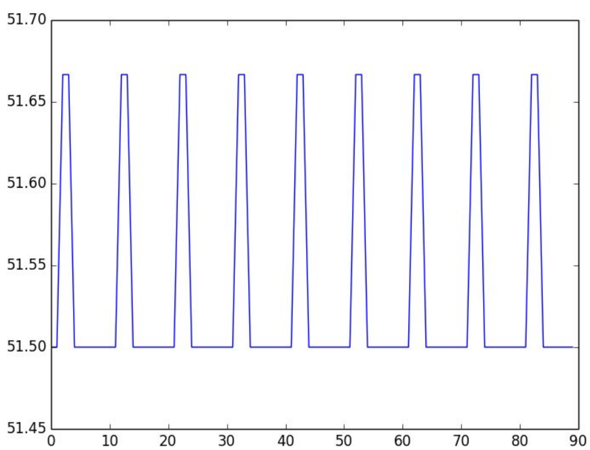
Data 4

Confusion Matrix

```
[[305   0]
 [295   0]]
```

TP=confusion[1,1]
TN=confusion[0,0]
FP=confusion[0,1]
FN=confusion[1,0]

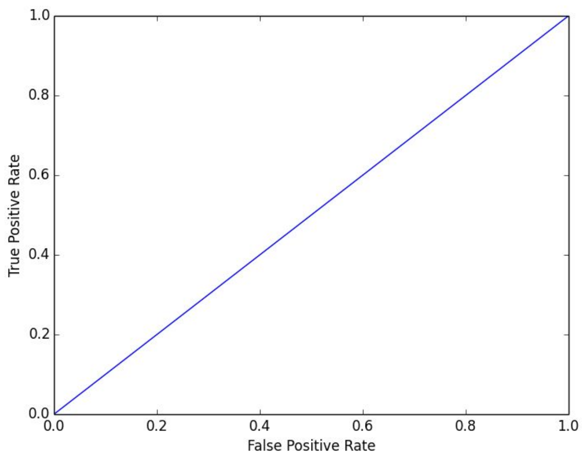
Soft margin with Linear Kernel



The plot shows different accuracies achieved by passing different parameters. To select the right parameters:
-Randomized method to train and test data was used
-Grid Search for different values of C and gamma

The right set of parameters can be chosen corresponding to the maximum parameters. The maximum accuracy is slightly more than 51%.

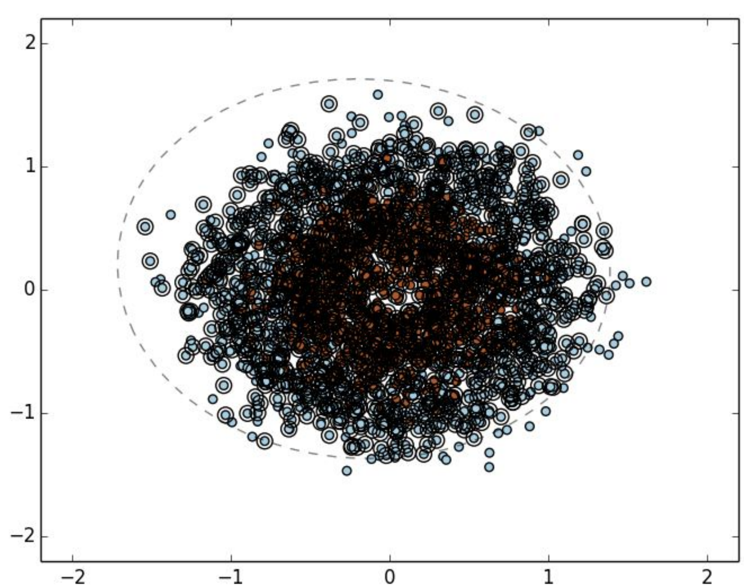
ROC Curve



It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).

The area under the curve is a measure of text accuracy

Plot showing Support Vectors and Classifier



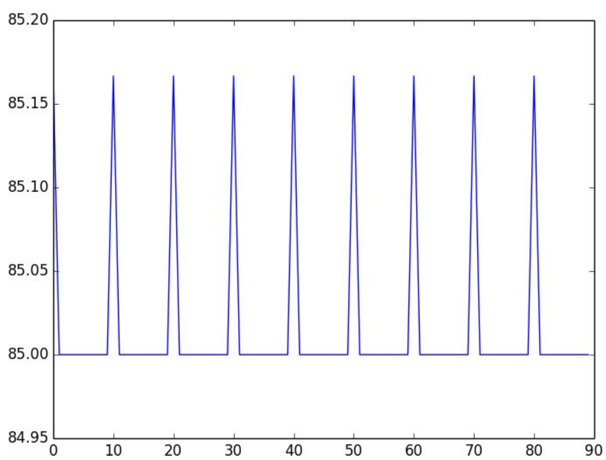
Data 5

Confusion Matrix

```
[[318   0]
 [282   0]]
```

TP=confusion[1,1]
TN=confusion[0,0]
FP=confusion[0,1]
FN=confusion[1,0]

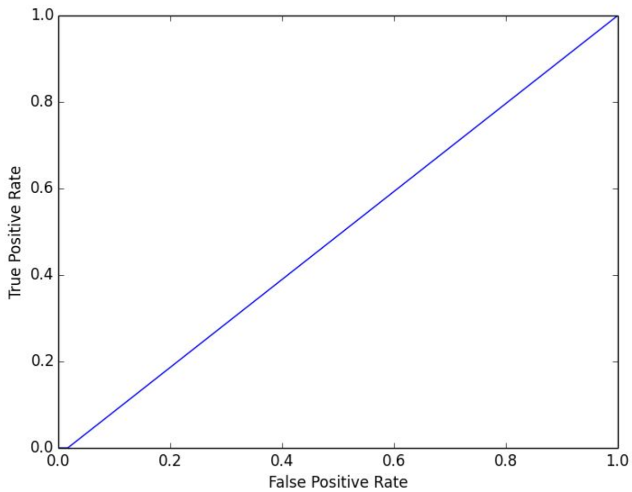
Soft margin with Linear Kernel



The plot shows different accuracies achieved by passing different parameters. To select the right parameters:
-Randomized method to train and test data was used
-Grid Search for different values of C and gamma

The right set of parameters can be chosen corresponding to the maximum parameters. The overall accuracy is a little more than 85%.

ROC Curve



It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity)

The area under the curve is a measure of text accuracy.

Plot showing Support Vectors and Classifier