Homework 4

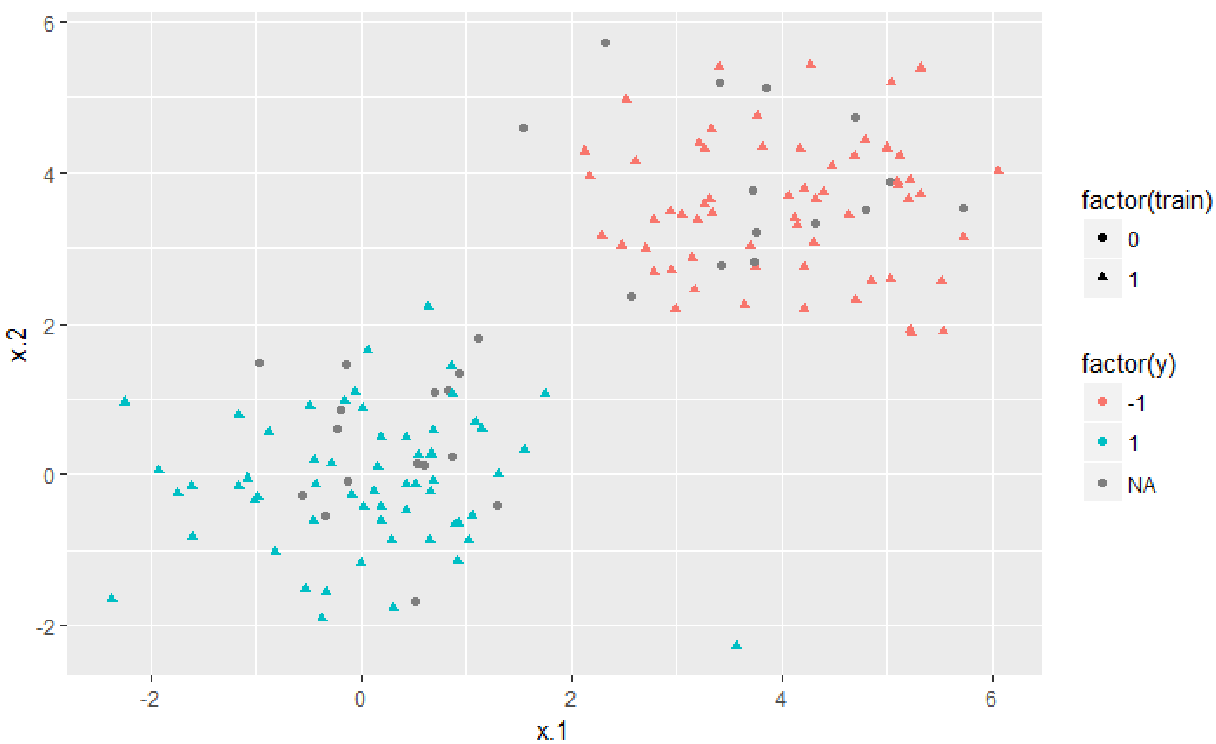
Part A: Linear dataset

A.1/A.2) Visualization of the first dataset:

linear1.RData

Question 1: Save your plot. How can you characterize this dataset?

This dataset can be characterized as linearly separable.

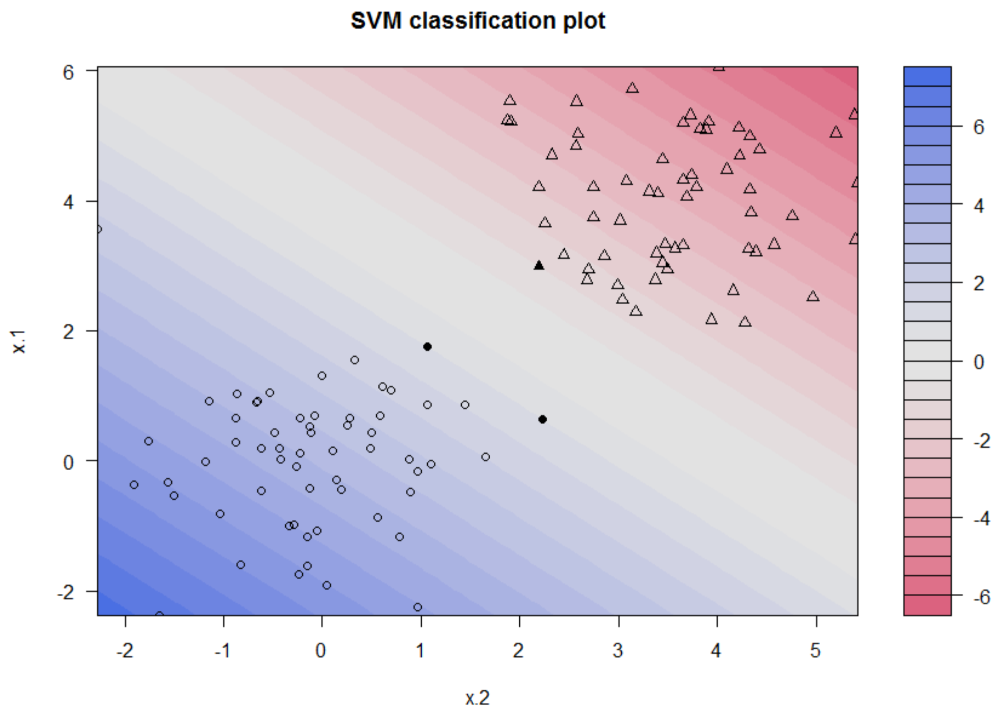


A.3/A.4/A.5) Training the SVM:

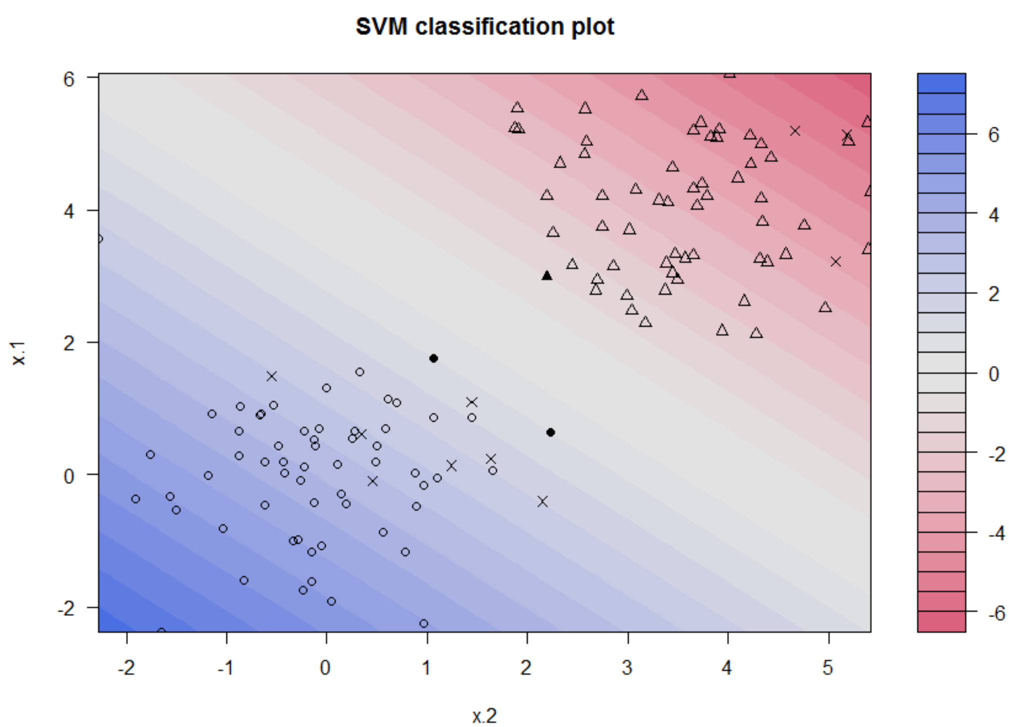
A.3) Train a linear SVM

```
SV type: C-svc  (classification)
 parameter : cost C = 100

Linear (vanilla) kernel function.

Number of Support Vectors : 3

Objective Function Value : -3.3078
Training error : 0
```

A.4) Plot the model



A.5) Adding points of test on the graph



Question 2: Save your plot. What are the black points in the figure?

The black points are the support vectors.

Question 3: What is the parameter C in the svm?

C is the parameter for the margin cost function.

A.6/A.7 Testing predictions on test set

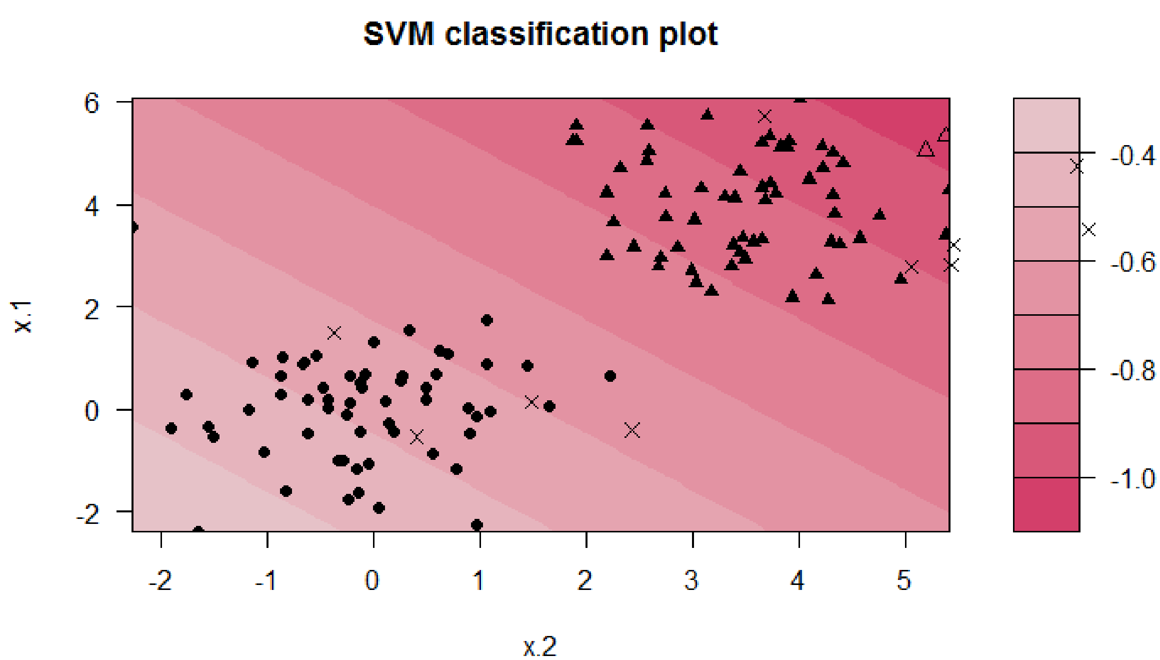
```
"Accuracy: 100%"
```

Question 4: Try different values for the parameter C in the svm. How is the accuracy of the model affected by varying C?

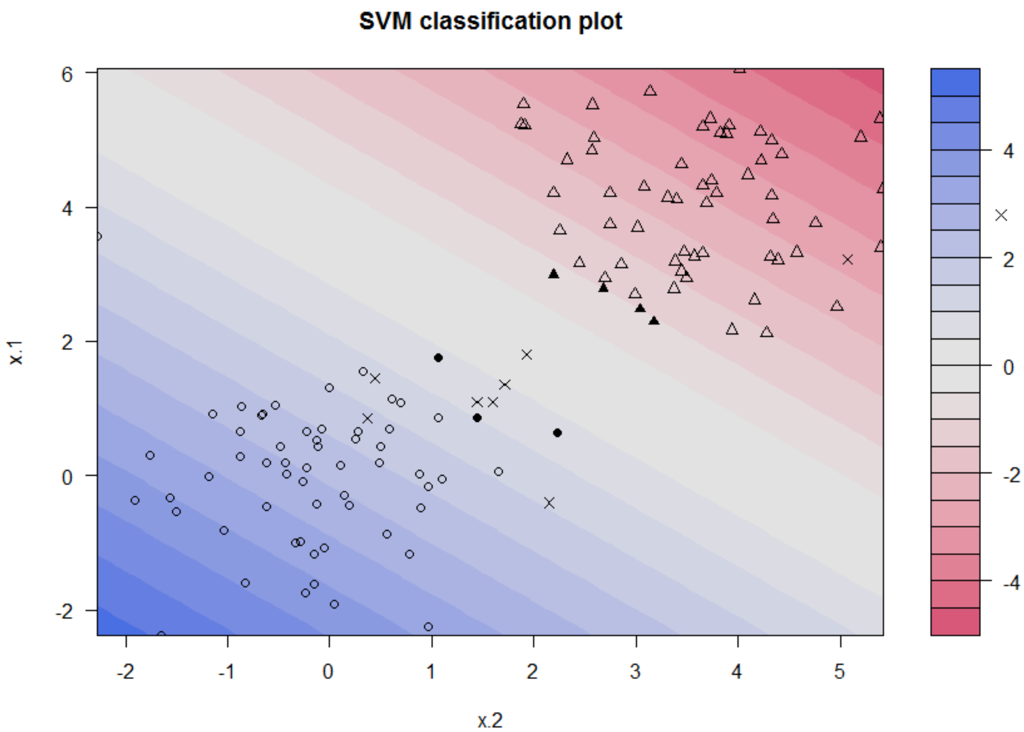
A large C usually gives low bias and high variance and a small C gives you a higher bias and lower variance. The accuracy didn't change for this dataset.

C	Accuracy	# of SVM vectors	Margin
1000	100%	3	narrower
100	100%	3	
10	100%	3	
2	100%	4	
1	100%	7	
2^-9	56.67%	118	
2^-10	46.67%	118	Wider

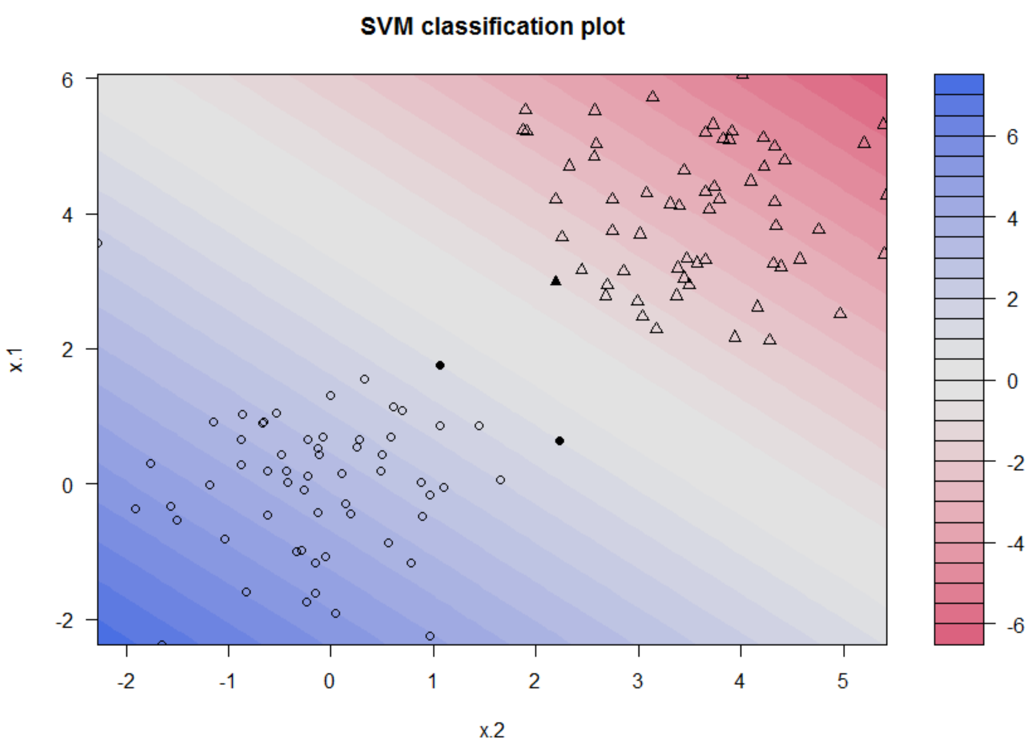
For C=1 (Margin is wider)



For C=1



For C=100 (Margin is narrower)

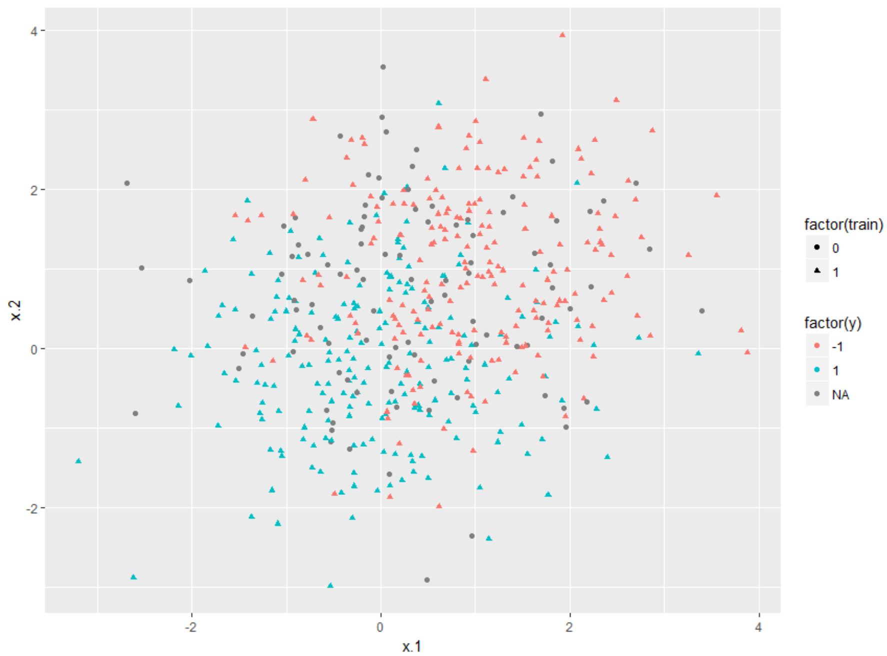


A.9 to A.14) Do the exact same thing as the first part on the linear2 dataset (nonseparable dataset).

linear2.RData

A.10) Visualize the data

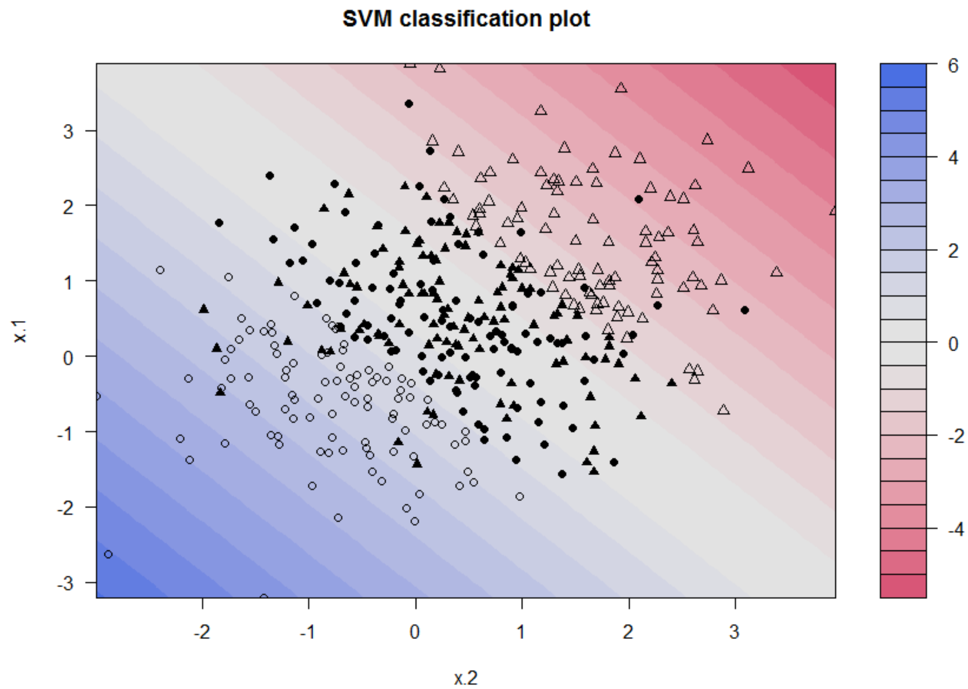
This dataset can be characterized as non-separable linearly.



A.11) Train a linear SVM

```
SV type: C-svc  (classification)
 parameter : cost C = 100

Linear (vanilla) kernel function.

Number of Support Vectors : 212

Objective Function Value : -21053.68
Training error : 0.23
```

A.12) Plot the model

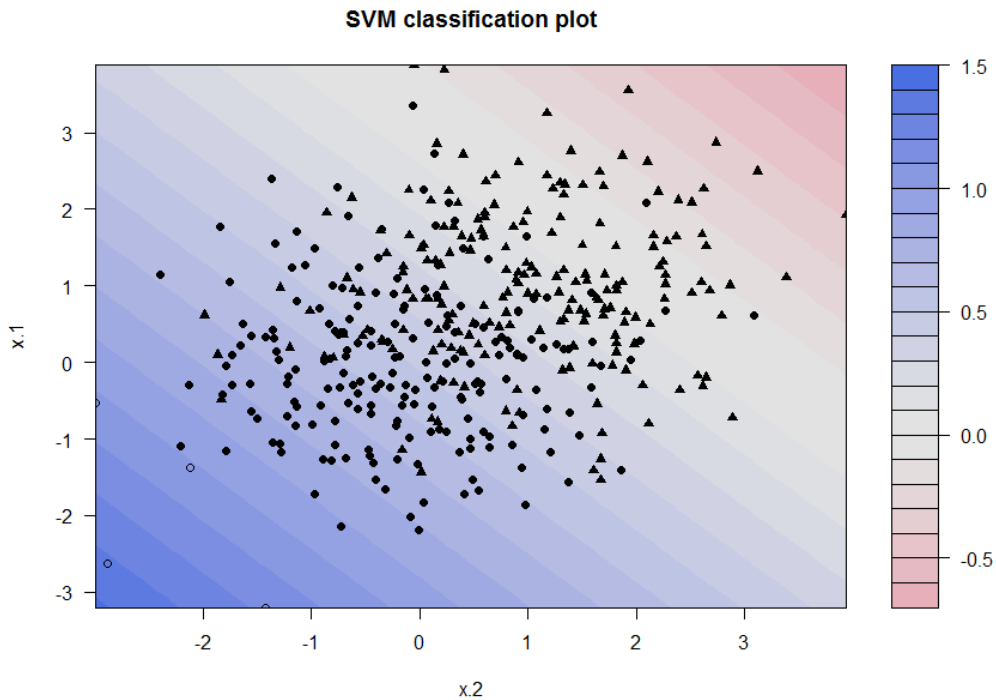


A.14) Look at accuracy

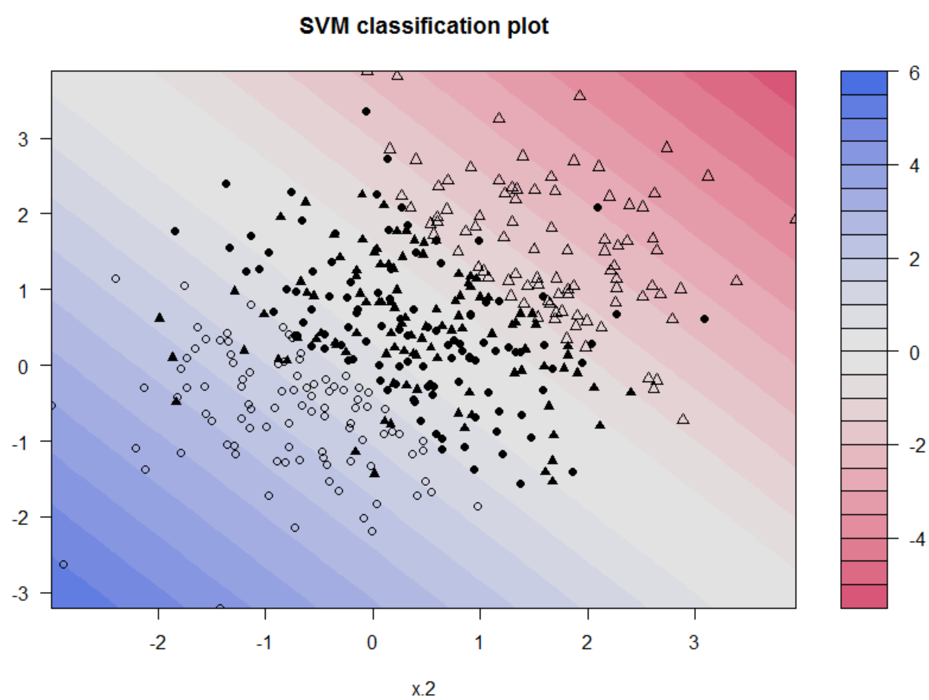
```
"Accuracy: 67%"
```

C	Accuracy	# of SVM vectors	Margin
1000	67%	212	narrower
100	67%	212	
10	67%	212	
2	67%	213	
1	67%	214	
2^-5	67%	252	
2^-10	58%	396	Wider

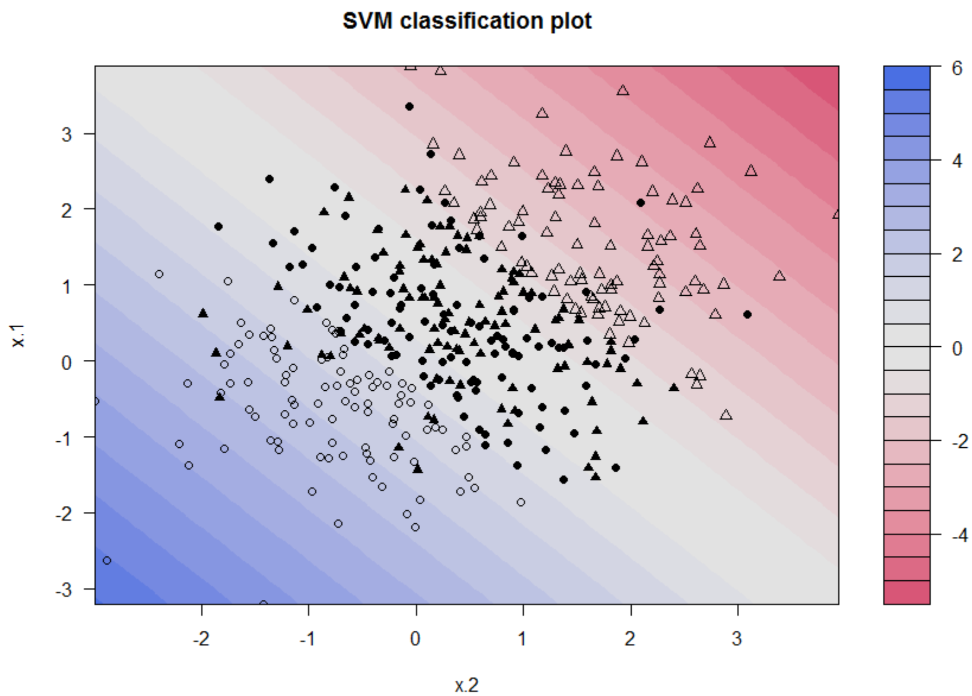
For C=2^-10 (Accuracy=58%)



For C=1 (Accuracy=67%)



For C=100 (Accuracy=67%)



Question 5: Is the accuracy a sufficient method to assess the performance of a model?

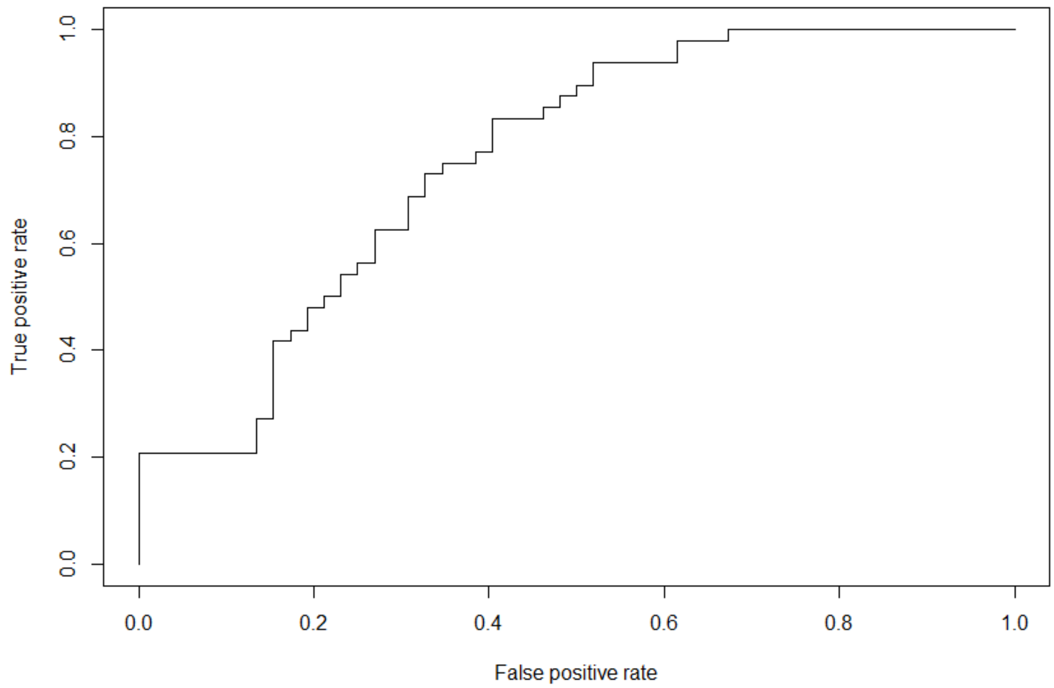
No, because of the skewed classes; the model that always predicts positive labels will have a high accuracy. We can do better by finding the confusion matrix, which gives us an indication if the model is confusing two classes. Also, ROC curve and performance curves can provide us with a better perspective of the model's accuracy.

A.15) A confusion matrix gives more information than just accuracy

Confusion Matrix:

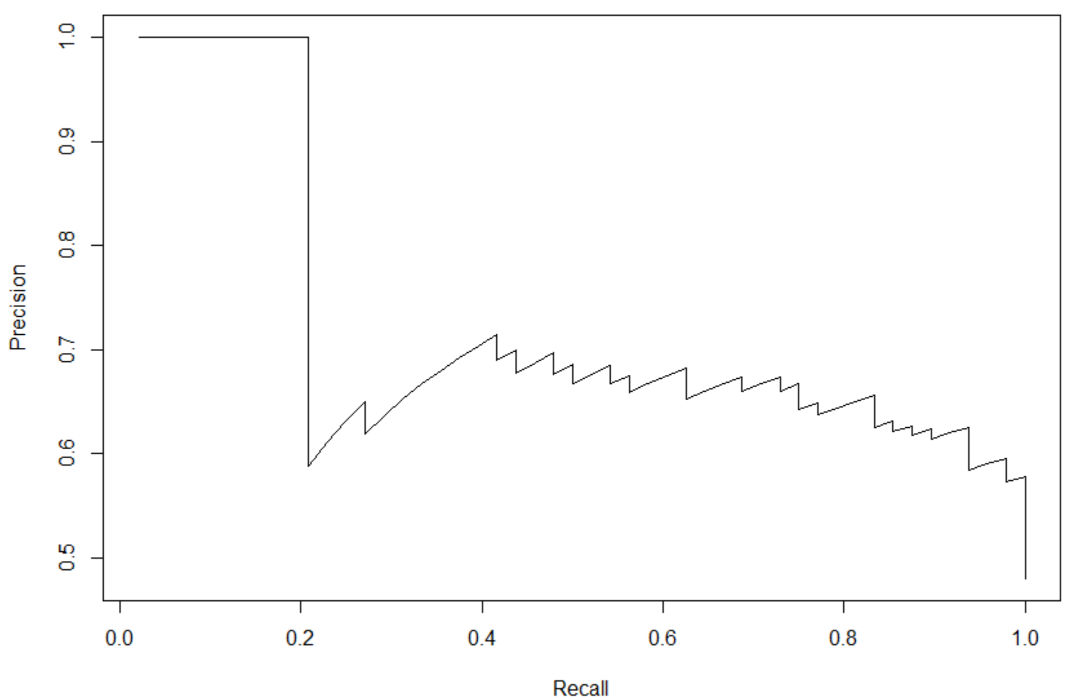
	Reality	
Prediction	-1	1
-1	36	17
1	16	31

A.16) ROC Curves



The area under the curve gives us an indication about the accuracy of the model. Since the area under the curve is more than 0.5, the accuracy can be interpreted as somehow good.

Performance Curve



Question 6: How would you characterize the performance of this model? Is this good? bad? surprising?

The performance of the model can be characterized as average. There is a big difference between the accuracy of this dataset and the accuracy of the previous set.

Part B: Non-Linear dataset:

B.1/B.2/B.3) Trying linear SVM on a dataset where it is not appropriate

B.1) Load the data

B.2) Visualize the data

Question 7: How can you characterize this dataset?

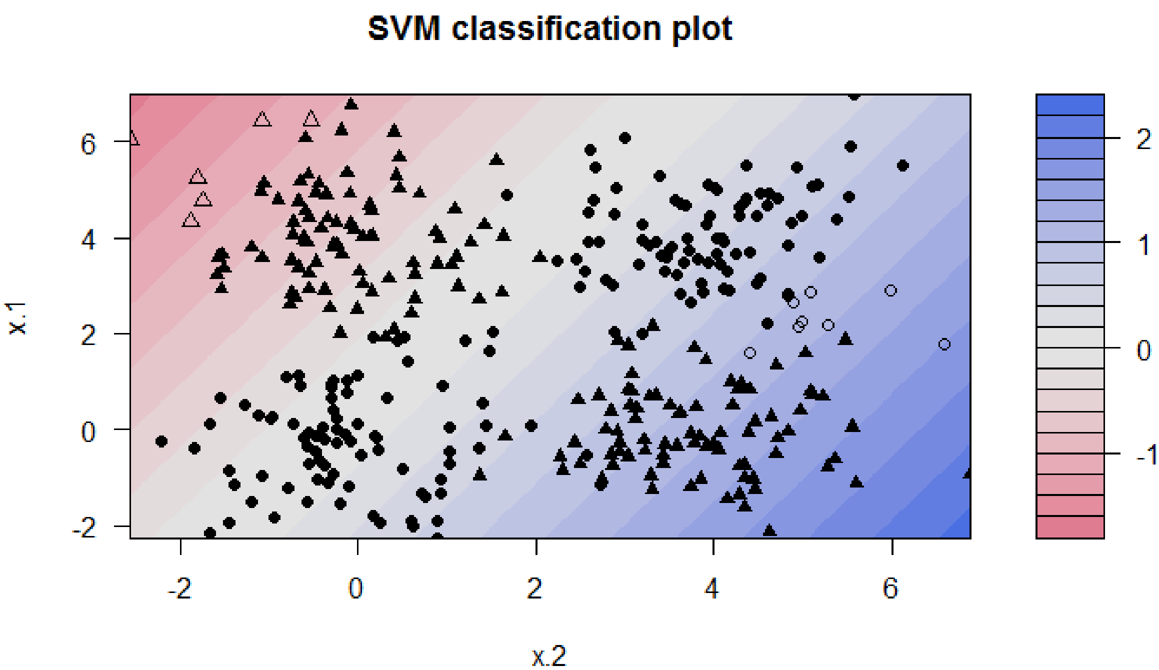
This dataset can be characterized as non-linear separable. It requires binomial parameters.

B.3) Let's still try a linear svm

```
SV type: C-svc  (classification)
 parameter : cost C = 100

Linear (vanilla) kernel function.

Number of Support Vectors : 336

Objective Function Value : -33487.83
Training error : 0.365714

"Accuracy: 54%"
```

Question 8: How would you characterize the performance of this model? Is this good? bad? surprising?

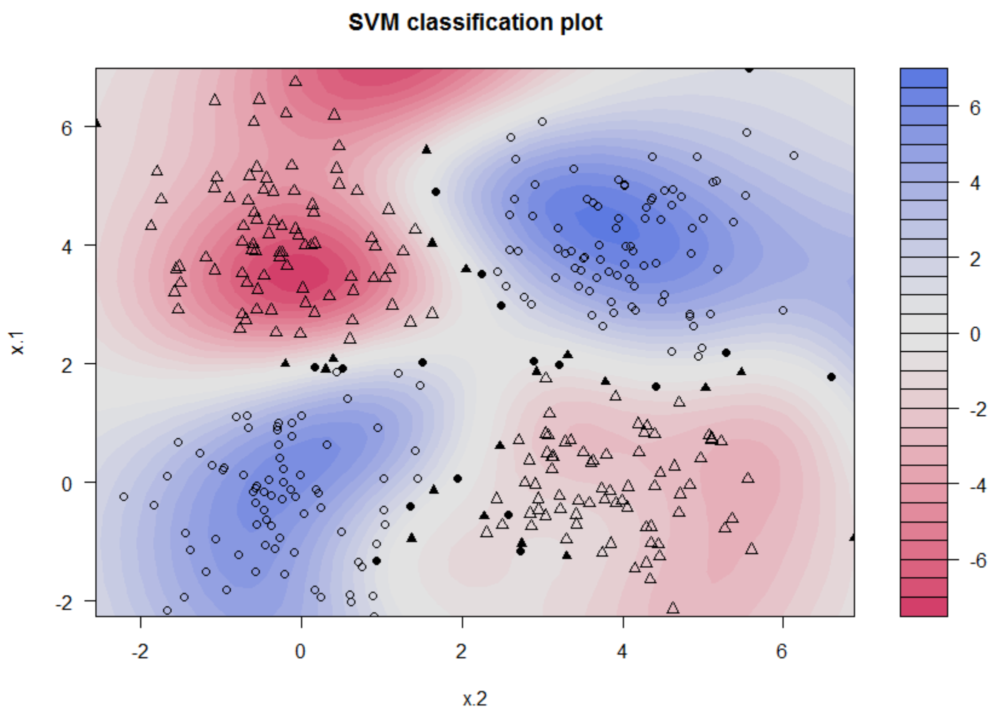
Not so great! The linear SVM performs badly with this dataset.



B.4) Ok try another one (RBF)

```
SV type: C-svc  (classification)
 parameter : cost C = 100

Gaussian Radial Basis kernel function.
 Hyperparameter : sigma =  1

Number of Support Vectors : 36

Objective Function Value : -2149.981
Training error : 0.025714


"Accuracy: 96%"
```

Using a different kernel enhanced the accuracy significantly (from 54% to 96%)



It is easier here to see the model from inside and see the errors in classification.

Question 9: How would you characterize the performance of this model? Is this good? bad? surprising?

The performance of the model is very high (good).

B.5) You can interactively see the impact of C and kernel type on the model

Testing on a scale of [-10,10]

Using 'Linear'='vanilladot', the maximum accuracy we get is 56% when c=-5

c.exponent	Accuracy
-10	48%
…	
-7	48%
-6	52%
-5	56%
-4	54%
…	
10	54%

Using Gaussian kernel, the maximum accuracy we get is 96% when c>=-5

c.exponent	Accuracy
-10	48%
…	
-7	48%
-6	94%
-5	96%
…	
10	96%

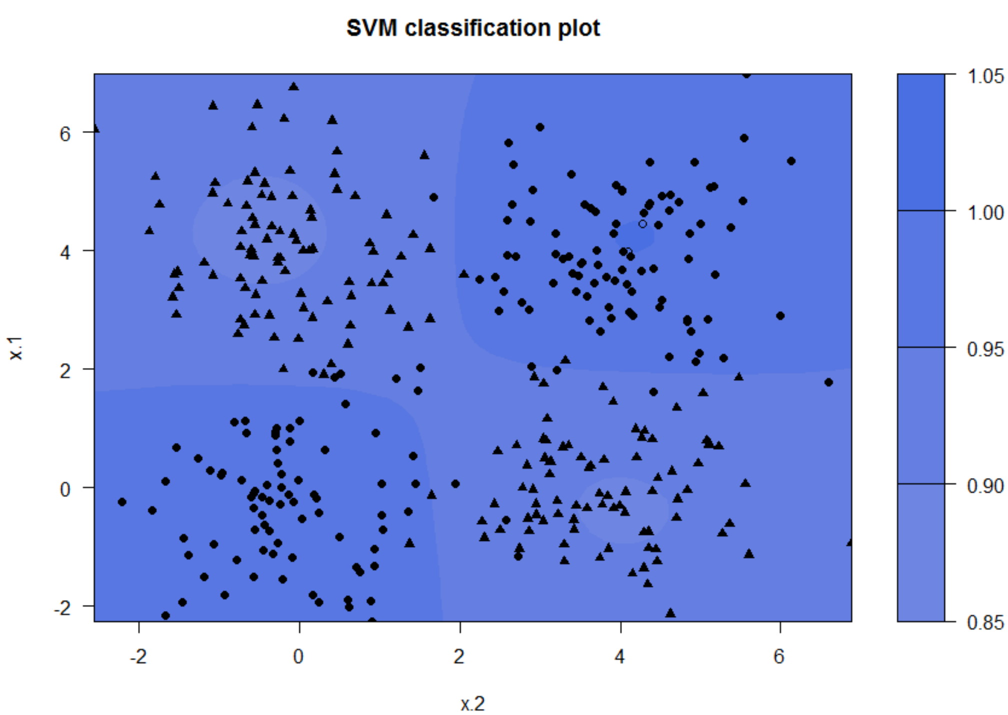
Using 'Hyperbolic'='tanhdot', the maximum accuracy we get is 48% when c=-10

c.exponent	Accuracy
-10	48%
…	
-7	48%
-6	94%
-5	96%
…	
10	90%

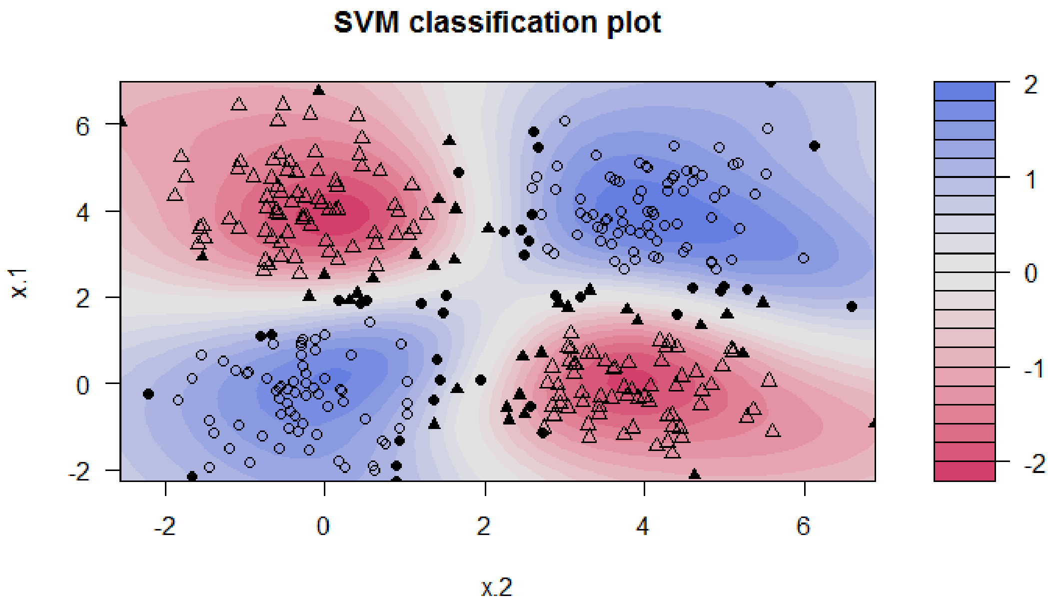
Using 'Laplacian'='laplacedot', the maximum accuracy we get is 96% when c>=-5

Following are the graphs for different c values:

c.exponent= -10 (Accuracy=48%)



c.exponent=-5 (Accuracy=96%)



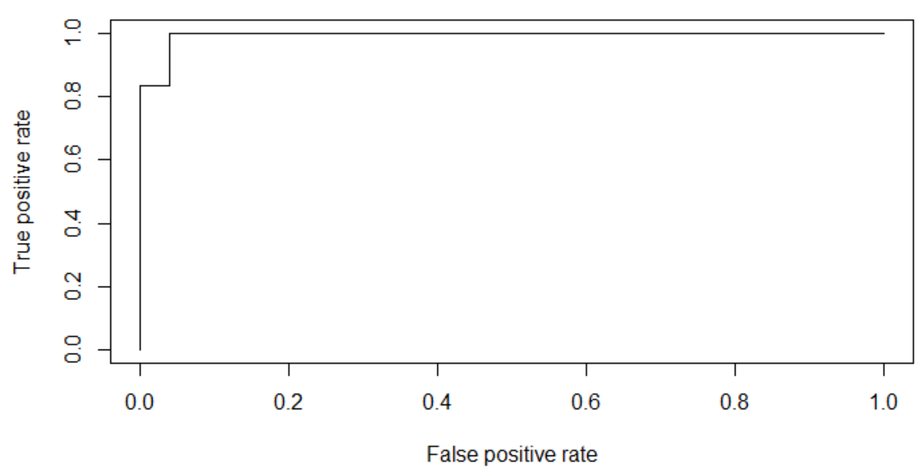
Q10: What is the kernel and C value you set for your model here?

I will choose the Gaussian kernel as it performs the best with an accuracy of 96%, and any C value larger than -5. I will pick c=10; since the ROC curve shows better performance

Confusion matrix:

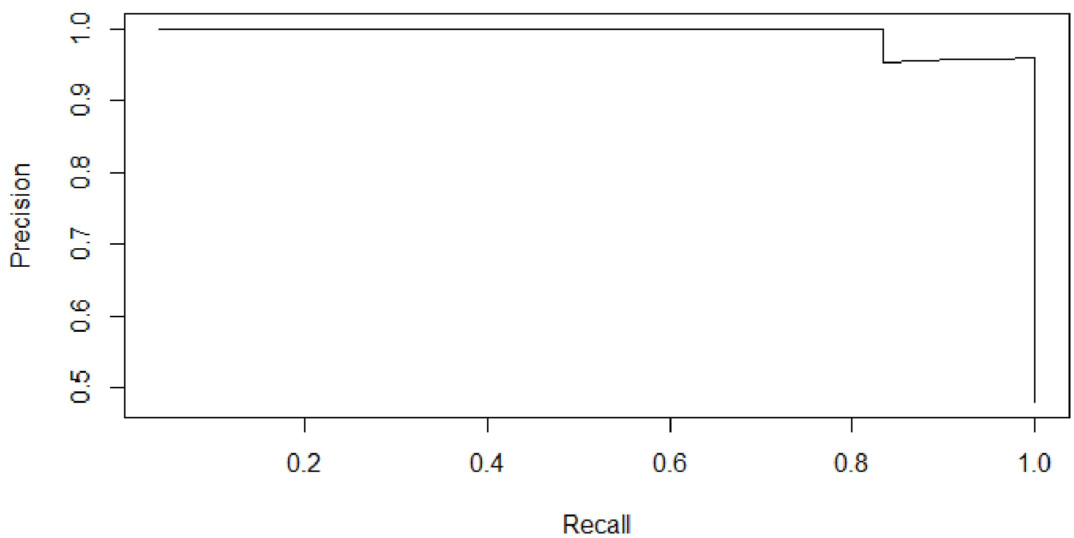
	Reality	
Prediction	-1	1
-1	25	1
1	1	23

ROC Curve:



As we can see in the figure, the area under the curve is almost 1; which means an excellent model.

Performance Curve



B.6) Visualization of the impact of C on the prediction accuracy (Bias-Variance Tradeoff):

Q12: Plot the performance versus C curve for the nonlinear SVM with Guassian kernel. How would you use this plot to optimize the choice for C?

I would choose a value of c when the error is minimum. According to the graph, when the value of c is larger than ~1e-01, the error stabilizes and becomes minimum.

Variance