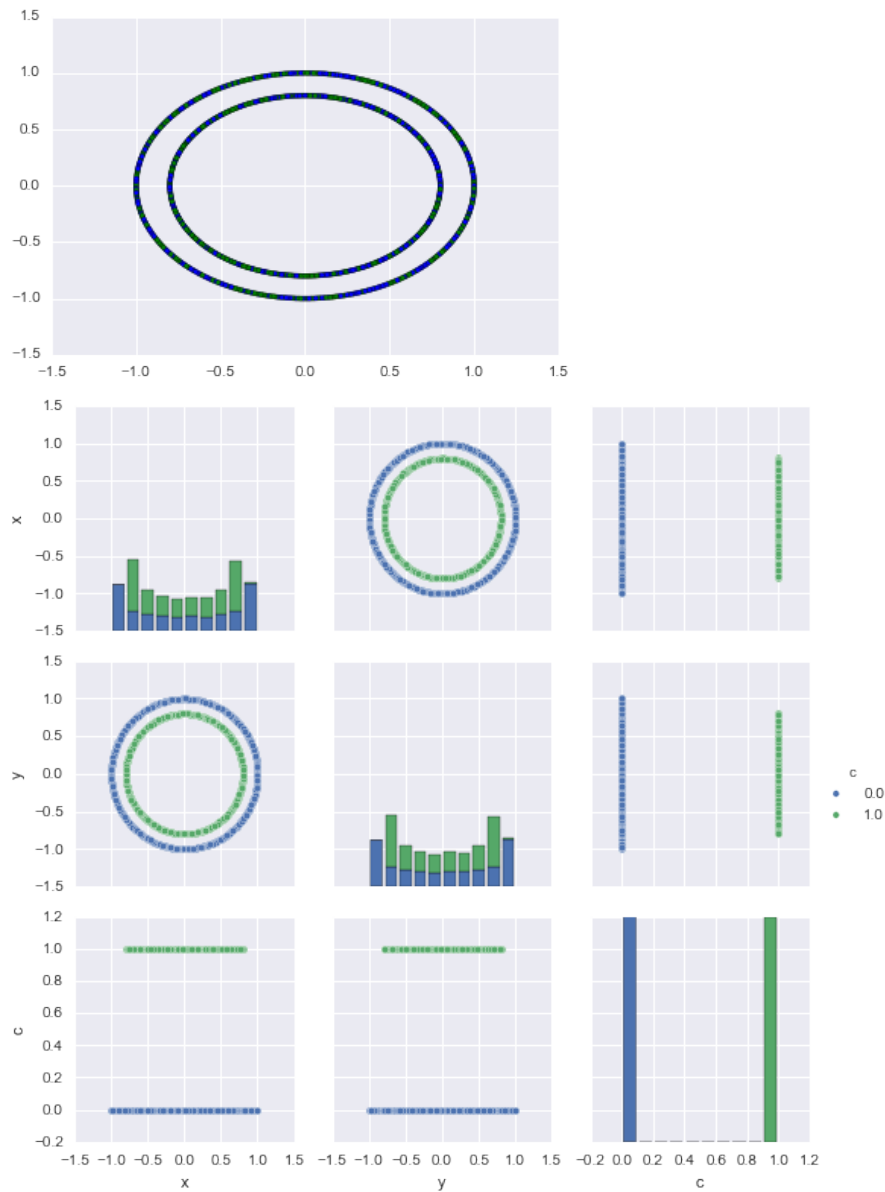


Question 1 Part A



Source Code: Question 1 Part A.ipynb

Question 1 Part B

Source Code: Question 1 Part B.ipynb

Performance

Accuracy: 0.44 (+/- 0.06)

Question 1 Part C

The Logistic Regression fails because there is no linearity between the feature. Also we know that the data is in the form of two concentric circles (almost circles, although they are ellipse, but this generalization will work in this case), we can separate them out based on which circle they lie and that will be a linear classification. For doing so we create a

new dataset, which is the $\sqrt{f_1^2 + f_2^2}$. Using these features we are able to classify them properly with 100% accuracy.

Source Code: Question 1 Part C.ipynb

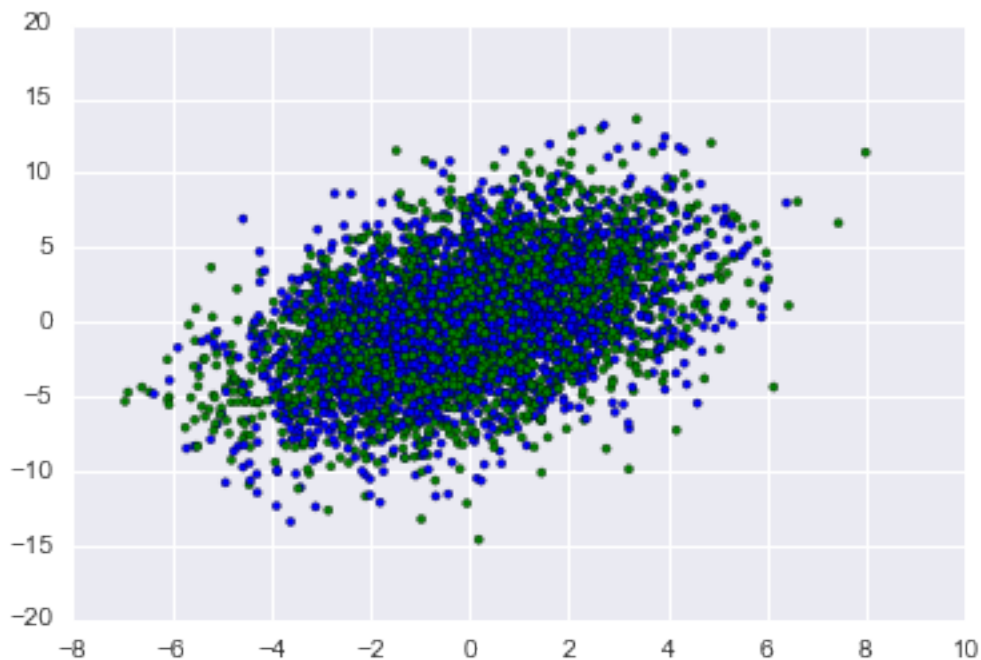
Question 1 Part D

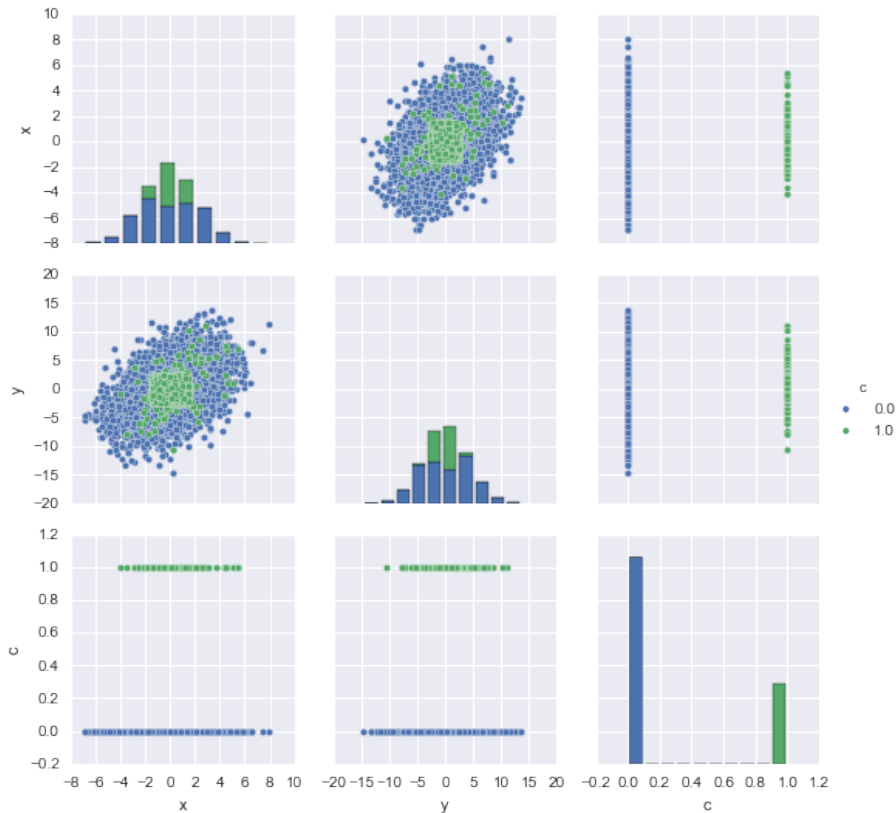
In the question the data consists of two concentric circles, as seen from the graph, the decision tree approach of random forest performs better as compared to linear model logistic regression. Random forest develops lot of decision trees based on random selection of data. It uses these large number of decision trees to make a classification. These large of decision trees are independent of each other. The data is run through all the decision trees and the class with majority decision is given as the final decision. As we can see this is not a linear model, and since it is not a linear model the results produced by it are better than that of a linear model for a non-linear data.

Accuracy: 1.00 (+/- 0.01)

Source Code: Question 1 Part D.ipynb

Question 2 Part A





The data as per the graph is not linearly separable. (Source Code: Question 2 Part A.ipynb)

Question 2 Part B

Source Code: Question 2 Part B.ipynb

Accuracy: 0.72 (+/- 0.00)

Question 2 Part C

Source Code: Question 2 Part C.ipynb

Accuracy: 0.97 (+/- 0.01)

Question 2 Part D

Source Code: Question 2 Part D.ipynb

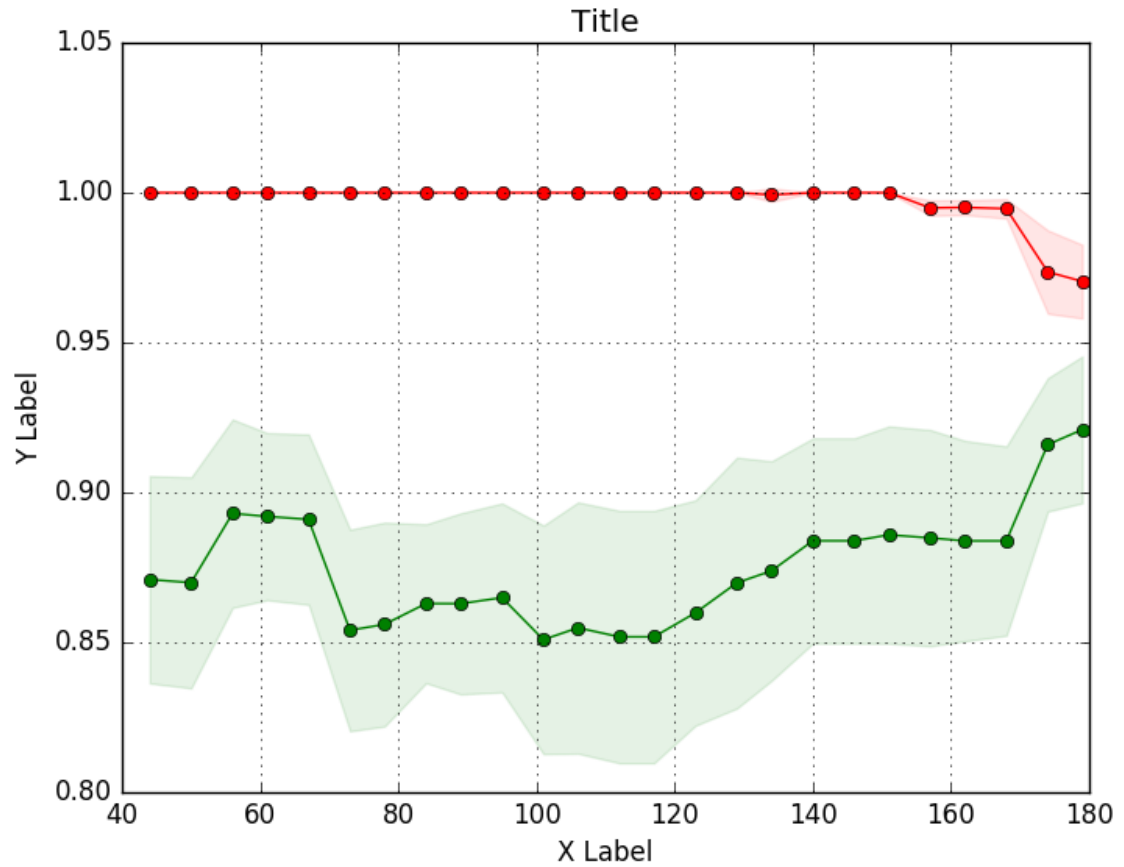
Accuracy: 0.98 (+/- 0.01)

Random Forest Classifier performs better than the Linear SVM and performs equally good as the Non-Linear(RBF) Classifier.

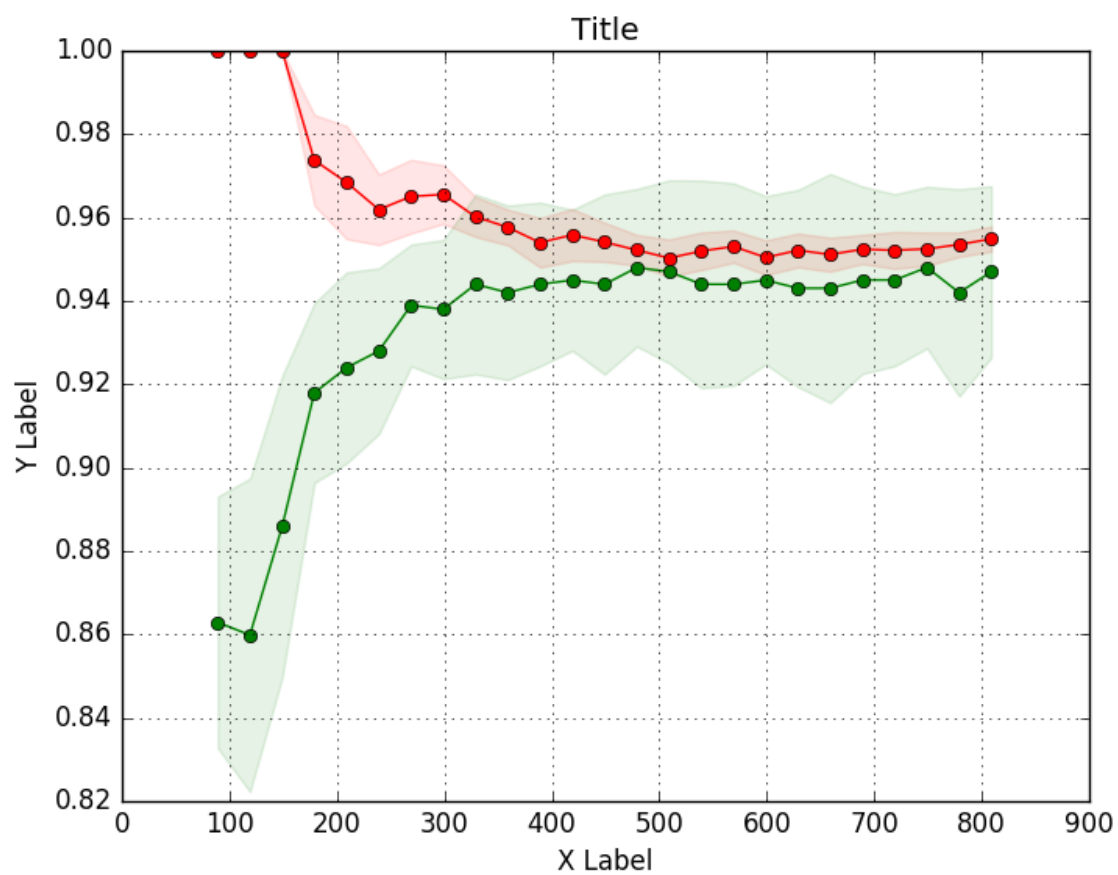
The accuracy shown above is with the default number of trees in the forest which is 10. The accuracy remains almost the same in the above case even if I decrease the number

of trees to 2 or increase it to 20. In the case with decision trees as 2 it changes to 97% and remains 98% for 20.

Question 3



OVERFITTING (SOURCE CODE: Question 3 Part A.ipynb)



NON OVERFITTING (SOURCE CODE: Question 3 Part B.ipynb)