[redacted]

Part 1 2 3: Observation on the value of the C on linear model

Initially, the model divides the data into train-validation and test sets. The train-validation data is then converted into train and validation set by k-fold cross validation method. Thus, everytime the training would be done on a different dataset giving us some diversity. The process goes on as follows:

1. The data is prepared.
2. 10 folds are made for the data into train and validation.
3. At the end, of every fold the score is calculated and appended.
4. The standard score is the mean of every score value in the appended array.
5. The best value is then tested on the test set.

Different values of C are tried and the best average score is chosen. The test results are as follows:

On first Run:

Average Score for **C=0.1**
0.9666666666666666
Average Score for **C=1**
0.9666666666666666
Average Score for **C=2**
0.9666666666666666
Average Score for **C=100**
0.9583333333333333
Average Score for **C=1000**
0.95
Average Score for **C=10000**
0.95
Average Score for **C=10000000**
0.9416666666666667

On second Run:

Average Score for **C=0.1**
0.95
Average Score for **C=1**
0.9666666666666666
Average Score for **C=2**
0.9749999999999999
Average Score for **C=100**
0.9833333333333332
Average Score for **C=1000**
0.9583333333333333
Average Score for **C=10000**
0.9749999999999999
Average Score for **C=10000000**
0.9749999999999999

On third Run:

Average Score for **C=0.1**
0.9583333333333333
Average Score for **C=1**
0.975
Average Score for **C=2**
0.975
Average Score for **C=100**
0.9583333333333333
Average Score for **C=1000**
0.9749999999999999
Average Score for **C=10000**
0.9749999999999999
Average Score for **C=10000000**
0.9666666666666666

On Fourth Run:

Average Score for **C=0.1**
0.95
Average Score for **C=1**
0.975
Average Score for **C=2**
0.9583333333333333
Average Score for **C=100**
0.9666666666666666

Average Score for **C=1000**
0.9666666666666666
Average Score for **C=10000**
0.975
Average Score for **C=10000000**
0.95

In the end it can be inferred from the values that the model gives best score at the regularization parameter **C = 2**

Part 4: Observation on the value of the C and gamma on non linear model

The model used the K-fold cross validation method. Dividing the data into 10 folds of data containing training and validation data. Non-linear kernel known as Gaussian rbf was used. Different combinations of gamma and Regularization parameter C were tried.
Following were the results:

C=1
Average Score for C=1 and **gamma=1**
0.9666666666666666
Average Score for C=1 and **gamma=5**
0.9583333333333333
Average Score for C=1 and **gamma=10**
0.9166666666666666
Average Score for C=1 and **gamma=20**
0.8
Average Score for C=1 **gamma=40**
0.6833333333333333
Average Score for C=1 and **gamma=80**
0.5583333333333333

C=2
Average Score for C=2 and **gamma=1**
0.9749999999999999
Average Score for C=2 and **gamma=5**
0.95
Average Score for C=2 and **gamma=10**
0.9166666666666667
Average Score for C=2 and **gamma=20**
0.8416666666666666

Average Score for C=2 **gamma=40**
0.7416666666666666
Average Score for C=2 and **gamma=80**
0.6
Average Score for C=2 and **gamma=80**
0.5583333333333333

C=5
Average Score for C=5 and **gamma=1**
0.975
Average Score for C=5 and **gamma=5**
0.9666666666666666
Average Score for C=5 and **gamma=10**
0.9166666666666666
Average Score for C=5 and **gamma=20**
0.8583333333333332
Average Score for C=5 **gamma=40**
0.7
Average Score for C=5 and **gamma=80**
0.5

C=1000

Over here unlike previous runs gamma = 5 gave a higher result than gamma=1 in all runs

Average Score for C=1000 and **gamma=1**
0.9333333333333333
Average Score for C=1000 and **gamma=5**
0.95
Average Score for C=1000 and **gamma=10**
0.9166666666666667
Average Score for C=1000 and **gamma=20**
0.8416666666666666
Average Score for C=1000 **gamma=40**
0.6583333333333334
Average Score for C=1000 and **gamma=80**
0.5

C=10000

Average Score for C=10000 and **gamma=1**
0.9416666666666667
Average Score for C=10000 and **gamma=5**

0.95
Average Score for C=10000 and **gamma=10**
0.9583333333333334
Average Score for C=1000000 and **gamma=20**
0.8166666666666667
Average Score for C=10000 **gamma=40**
0.6083333333333332
Average Score for C=10000 and **gamma=80**
0.5750000000000001

C=100000

Average Score for C=100000 and **gamma=1**
0.9499999999999998
Average Score for C=100000 and **gamma=5**
0.9583333333333334
Average Score for C=100000 and **gamma=10**
0.925
Average Score for C=100000 and **gamma=20**
0.8166666666666668
Average Score for C=100000 **gamma=40**
0.725
Average Score for C=100000 and **gamma=80**
0.625

C=100000000

Average Score for C=100000000 and **gamma=1**
0.9333333333333332
Average Score for C=100000000 and **gamma=5**
0.95
Average Score for C=100000000 and **gamma=10**
0.9083333333333334
Average Score for C=100000000 and **gamma=20**
0.8416666666666668
Average Score for C=100000000 **gamma=40**
0.6916666666666667
Average Score for C=100000000 and **gamma=80**
0.5833333333333333

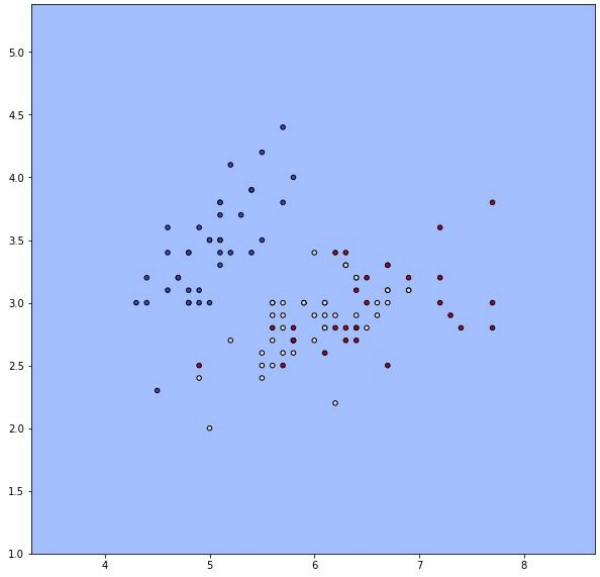
The above given data is extracted after multiple runs on the same data and same variables. Then the most common score is written over here.

It can be inferred from the above given data that the value of **c=5** and **gamma= 1** will yield an optimal result.

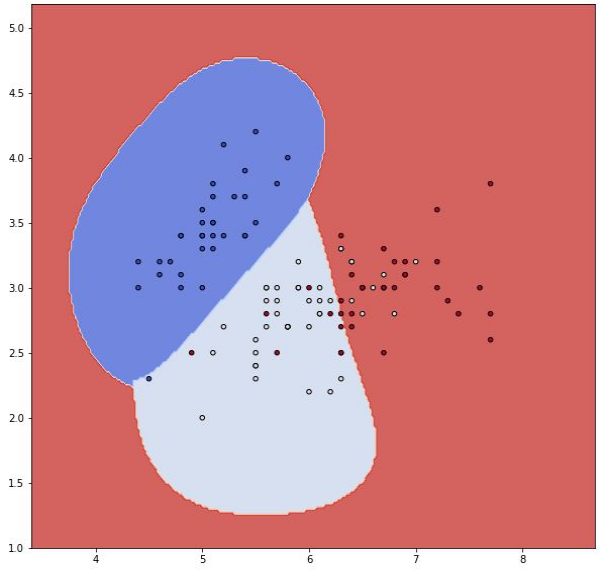
Part 5: Observation of gamma and C on the regions figure

Values of C and gamma had a large effect on the boundary of the svm. Firstly, the value of C will be observed. It is already known that a large value of a model leads toward the overfitting of the model.

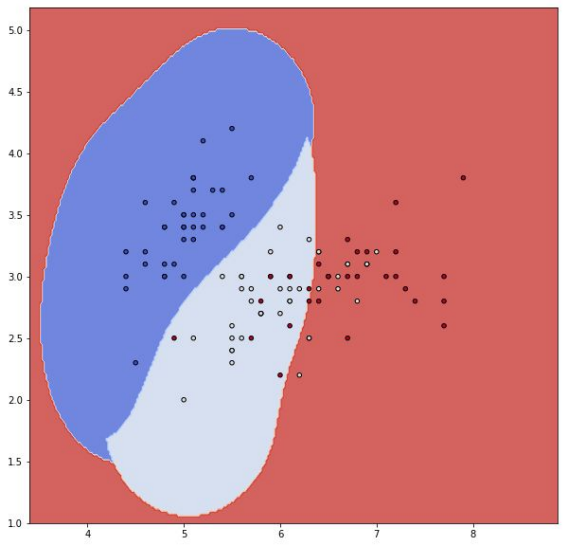
If the value of C is kept low the amount of slack would be higher and the misclassification would be higher. However, if the value of C is kept too much the model starts heading towards overfitting. Omitting the genericity of the classifier.



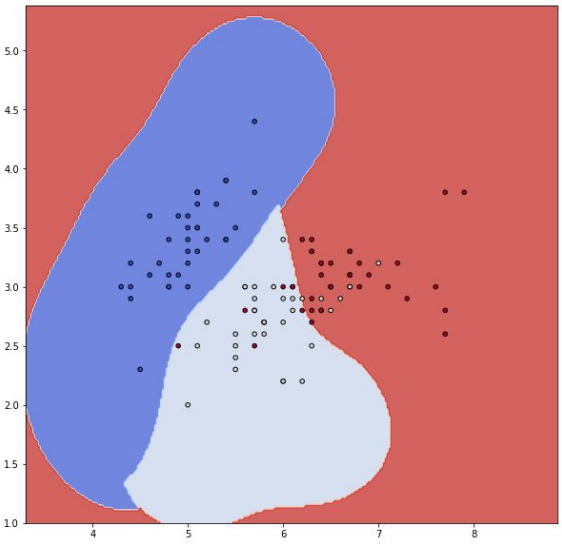
C=0.01



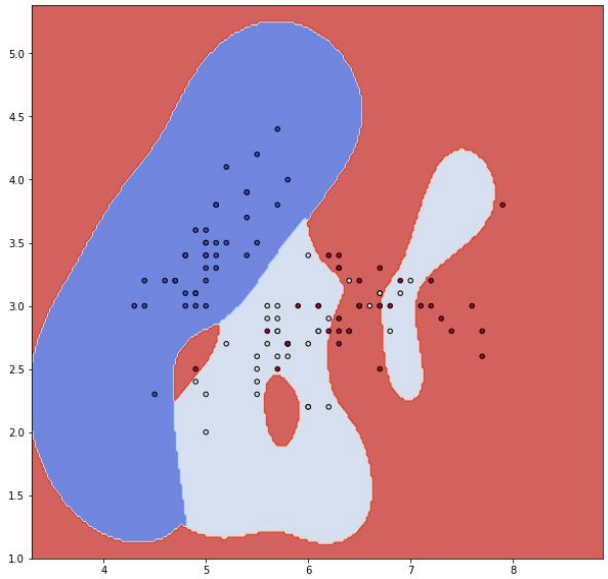
C=0.1



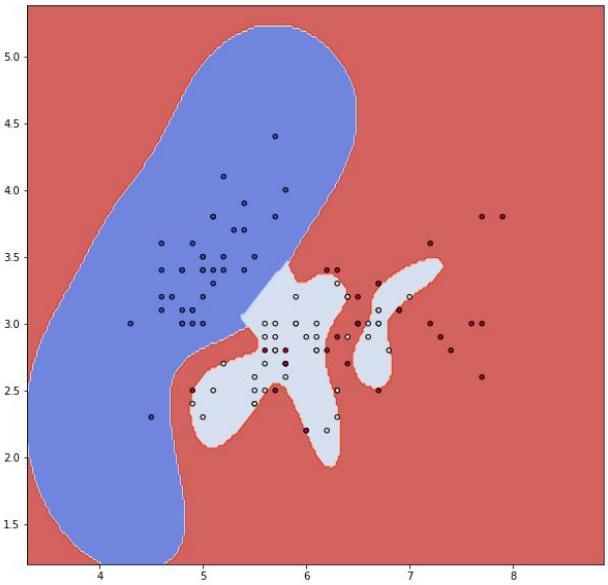
C=1



C=5



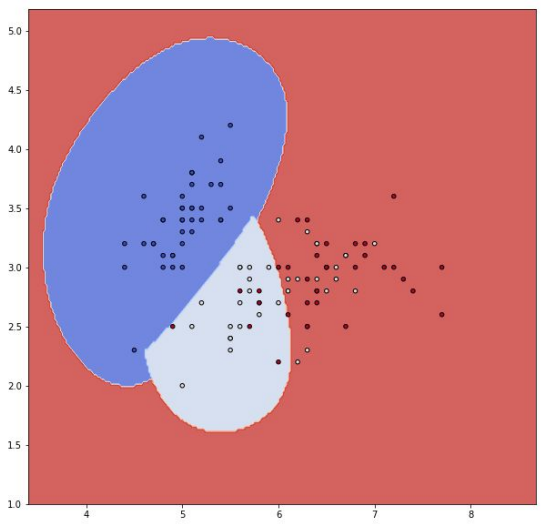
C=1000



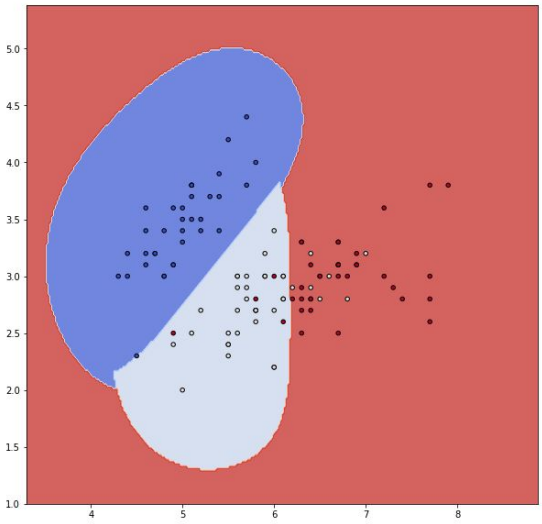
C=100000

Looking at all the figures, the best value seems like (one with the least error) **5**. As the misclassifications are least and the classifier looks like a generic one. In all the above examples the value of gamma is kept at 5.

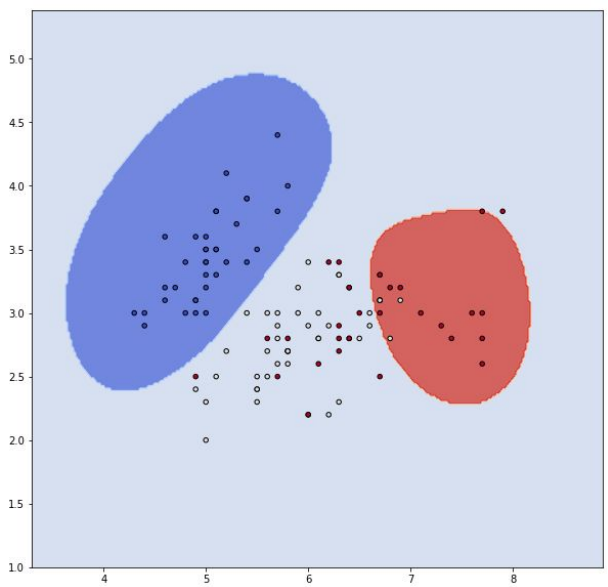
Assuming for the value of gamma that the optimal value should be in between 3/2 and 6/2. Assuming the denominator as the value of the number of input fields. Hence, the numbers of 1.5, 1.75, 2, 2.25, 2.5, 2.75 and 3 were tried keeping the value of C at 5 and the resulting models looked like as follows:



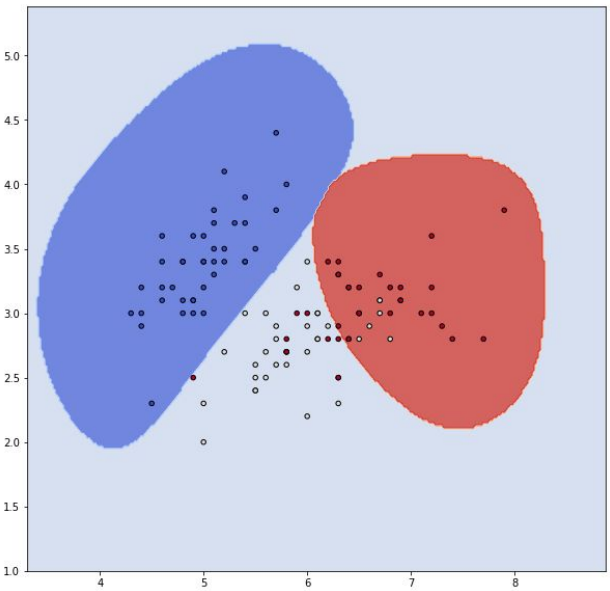
C=1.5



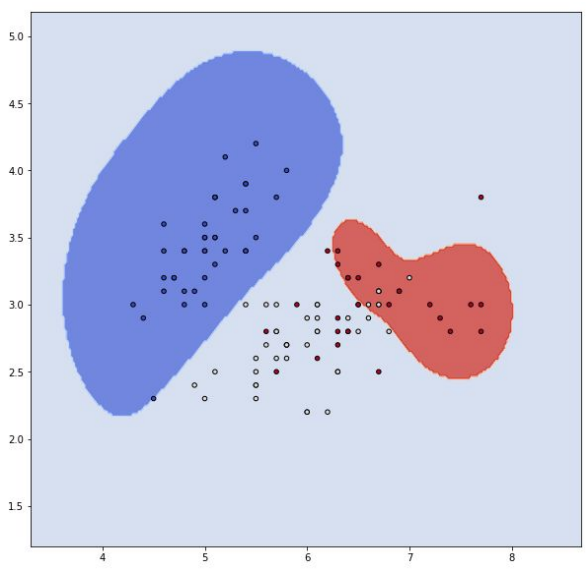
C=1.75



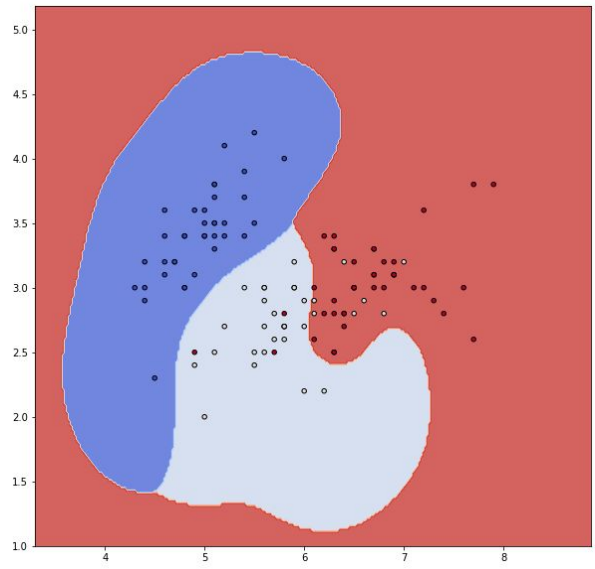
C=2



C=2.25



C=2.5



C=3

The best value for gamma would be **1.75** in the end giving overall score of 0.95 on iris_test dataset.