CSE 474 Project 3
Group- 21

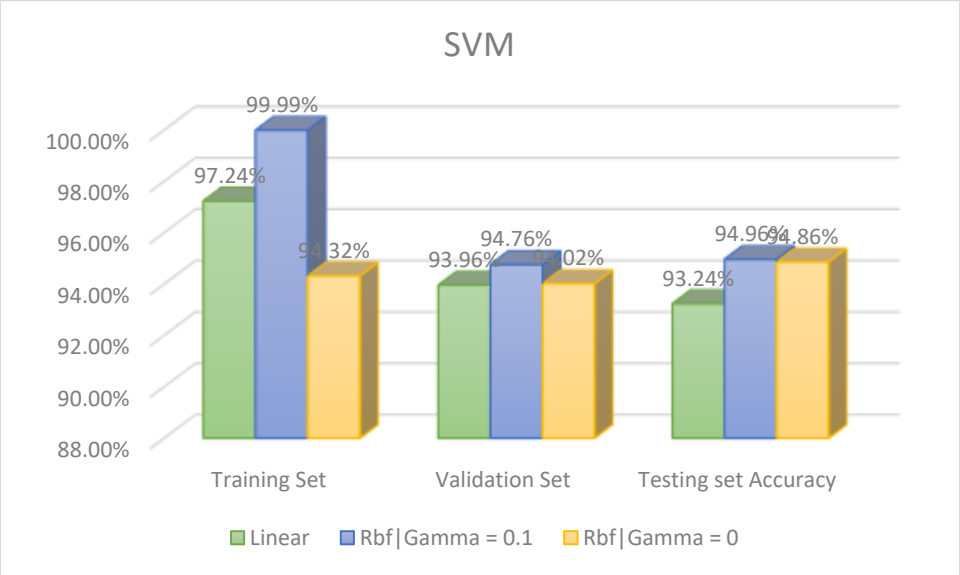
Jasmeet Kaur Chawla

Joshua Suttenberg

May 9th, 2018

Logistic Regression:

As we know Logistic regression is named for the function used at the core of the method, logistic function. Logistic regression models the probability of the default class. Here our classifier not just learns the linear boundary but also calculates the accuracies using radial basis function with different gamma properties. All of the accuracies we attained from the binary-class are good and prove that calculations found a strong fit to the data given (train and test data).



The accuracies received are:
Linear Kernel:

Training set Accuracy:97.236%
Validation set Accuracy:93.96%
Testing set Accuracy:93.238%

RBF Kernel (Gamma = 1.0):

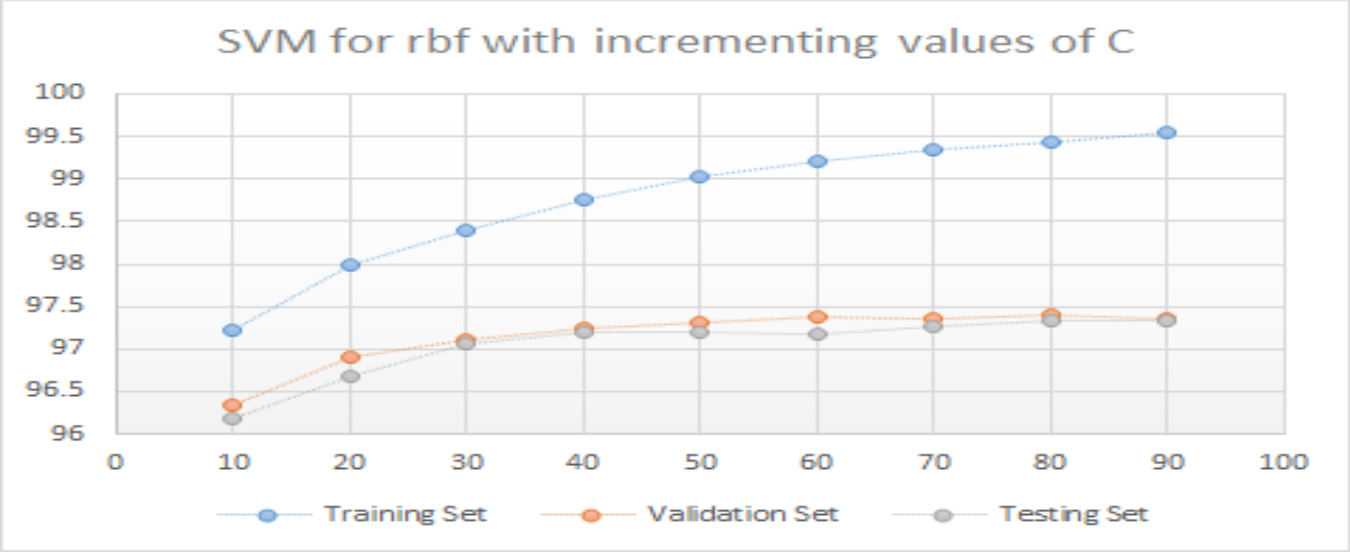
Training set Accuracy:99.992%
Validation set Accuracy:94.76%
Testing set Accuracy:94.96%

RBF Kernel (Gamma = 0.0):

Training set Accuracy:94.32%
Validation set Accuracy:94.02%
Testing set Accuracy:94.863%

SVM for incrementing Values of C

C	Training Set	Validation Set	Testing Set
10	97.218	96.34	96.179
20	97.997	96.91	96.67
30	98.398	97.119	97.07
40	98.762	97.25	97.2
50	99.026	97.31	97.2
60	99.21	97.38	97.17
70	99.346	97.36	97.26
80	99.444	97.399	97.33
90	99.556	97.36	97.33



Comparing the different kernels

With Gamma = 0.1, it has a very high Training set Accuracy. It has over fit the data, as in practice it only gets ~95%.. The kernel with default gamma = 1/n has much less overfitting and has generalized better. For any n >10, 1/n is much smaller a value than 0.1, so it over fits less. Our larger gamma of 0.1 has a larger bias with lower variance, while our smaller gamma has less bias, with more variance. In practice

for this problem, The performed close to the same, but with default gamma much less overfitting in training