CSE 474 – Programming Assignment 3
Classification and Regression

Group 28
Team Members: Ankit Nerlakanti, Vikram Garu, and Veronica Ng

Comparing Logistic Regression and SVM

Logistic Regression:

	Training Set Accuracy	Validation Set Accuracy	Testing Set Accuracy
2-class	86.166%	85.26%	85.23%
Multi-class	93.39%	92.43%	92.67%

In Logistic Regression, it is a discriminative classifier such that it tries to learn a linear boundary. Traditionally, logistic regression is used for binary classification. However, logistic regression can be extended to solve multi-class classification. This means that instead of having to build many classifiers in a 2-class scenario, in a multi-class scenario we will only need to build one classifier to classify multiple classes at the same time. This also implies that with a multi-class classification, it will be better and faster than the binary classification. The above chart displays the accuracies and exhibits that the multi-class classification has higher accuracies.

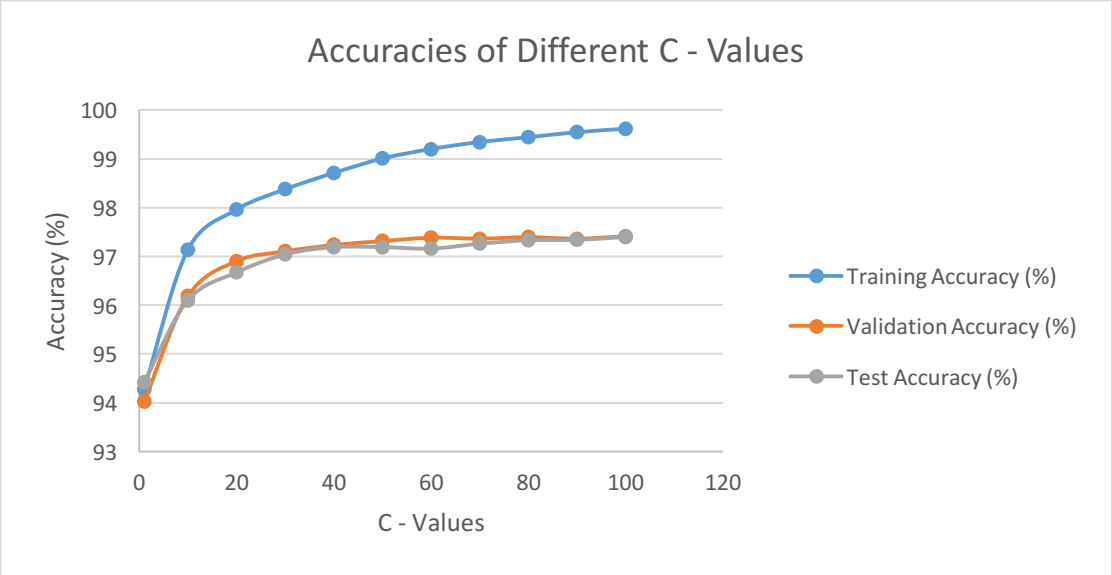
SVM:

	Training Set Accuracy	Validation Set Accuracy	Testing Set Accuracy
Linear Kernel	97.286%	93.64%	93.78%
Radical Basis Function with $\gamma = 0.1$	99.992%	94.76%	94.96%
Radical Basis Function with γ = Default	94.294%	94.02%	94.42%
Radical Basis Function with γ = Default, C = 1	94.274%	94.02%	94.42%
Radical Basis Function with γ = Default, C = 30	98.372%	97.1%	97.04%
Radical Basis Function with γ = Default, C = 50	99.002%	97.31%	97.19%
Radical Basis Function with γ = Default, C = 70	99.34%	97.36%	97.26%
Radical Basis Function with γ = Default, C = 100	99.612%	97.41%	97.4%

In SVM, it is a hyperplane based classifier. It tries to find the hyperplane with the maximum separation margin on the data. Thus, SVM would be better as compared to Logistic Regression. This can be seen when comparing the SVM chart and the Logistic Regression chart where SVM has higher accuracies.

As shown on the chart above, when Radical Basis Function with $\gamma = 0.1$ and when Radical Basis Function with γ = Default, we can see that as you slightly increase gamma the accuracy increases. However, if the gamma is too high there is a possibility of over fitting.

C - Values	Training Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)
1	94.274	94.02	94.42
10	97.132	96.18	96.1
20	97.952	96.9	96.67
30	98.372	97.1	97.04
40	98.706	97.23	97.19
50	99.002	97.31	97.19
60	99.196	97.38	97.16
70	99.34	97.36	97.26
80	99.438	97.39	97.33
90	99.542	97.36	97.34
100	99.612	97.41	97.4



Comparing Linear Kernel and Radical Basis Function with $\gamma = 0.1$ and variations of C, as C increases the accuracies increases. Thus, the Radical Basis Function with $\gamma = 0.1$ with increasing values of C is more accurate than the Linear Kernel.

The variations of C displays how it affects the error on training data. Therefore, as C value increases, the margin between the the hyperplane is smaller and vice versa.