CSE 574 - Machine Learning

Logistic Regression and Support Vector Machines

Group – 77

Sathish Kumar Deivasigamani
Sankara Vadivel Dhandapani
Kaushik Raj Palanichamy

Table of Contents

Table of Figures

Logistic Regression:

Performance of logistic regression is poor when compared to support vector machines as the mnist dataset is not linearly separable and has higher dimensions.

Training set Accuracy:	92.27%
Validation set Accuracy:	91.52%
Testing set Accuracy:	91.87%

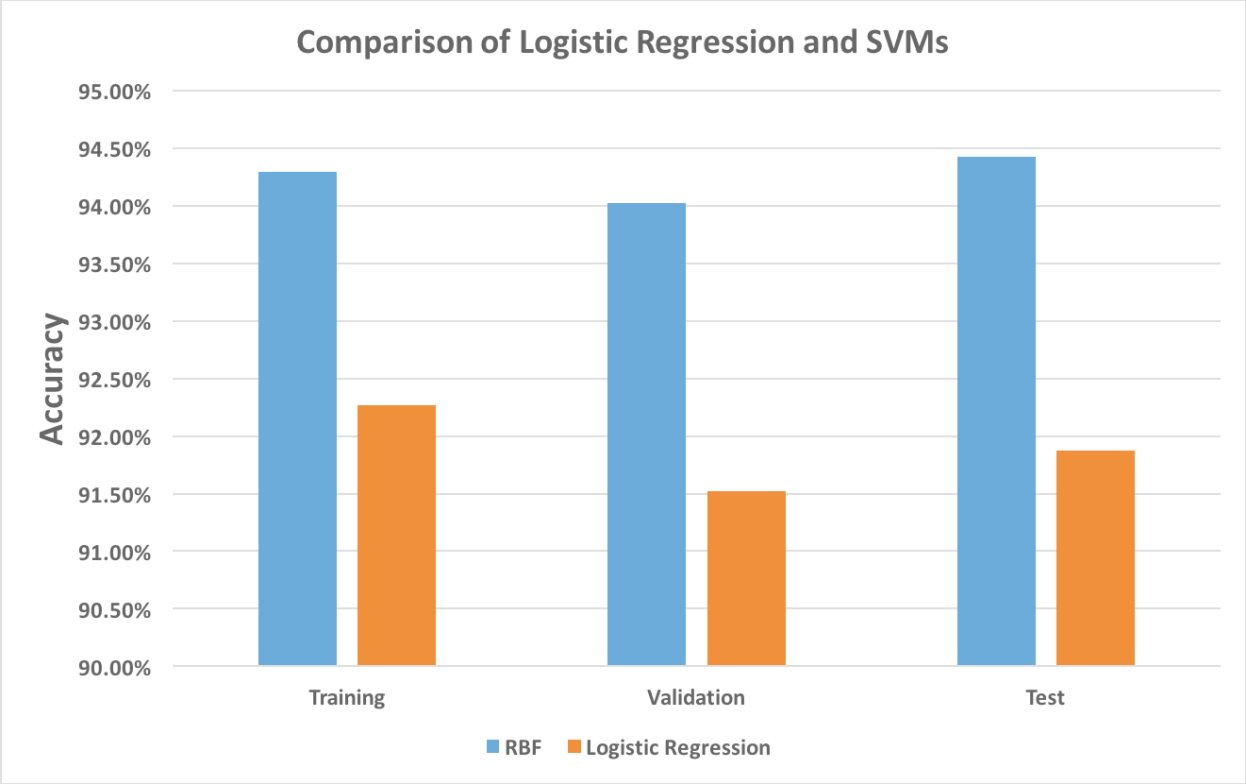

Figure 1: Comparison of Logistic Regression and Support Vector Machines

Support Vector Machines:

DIFFERENT KERNEL FUNCTIONS:

Different kernel functions perform differently depending upon the data. Linear kernel is a degenerate version of RBF[1]. So linear kernel will not be accurate in general when compared to RBF. But when the number of dimensions is high (like in this case, 715) non-linear mapping like RBF does not necessarily outperform RBF, as there is no need to map data into higher dimension [2].

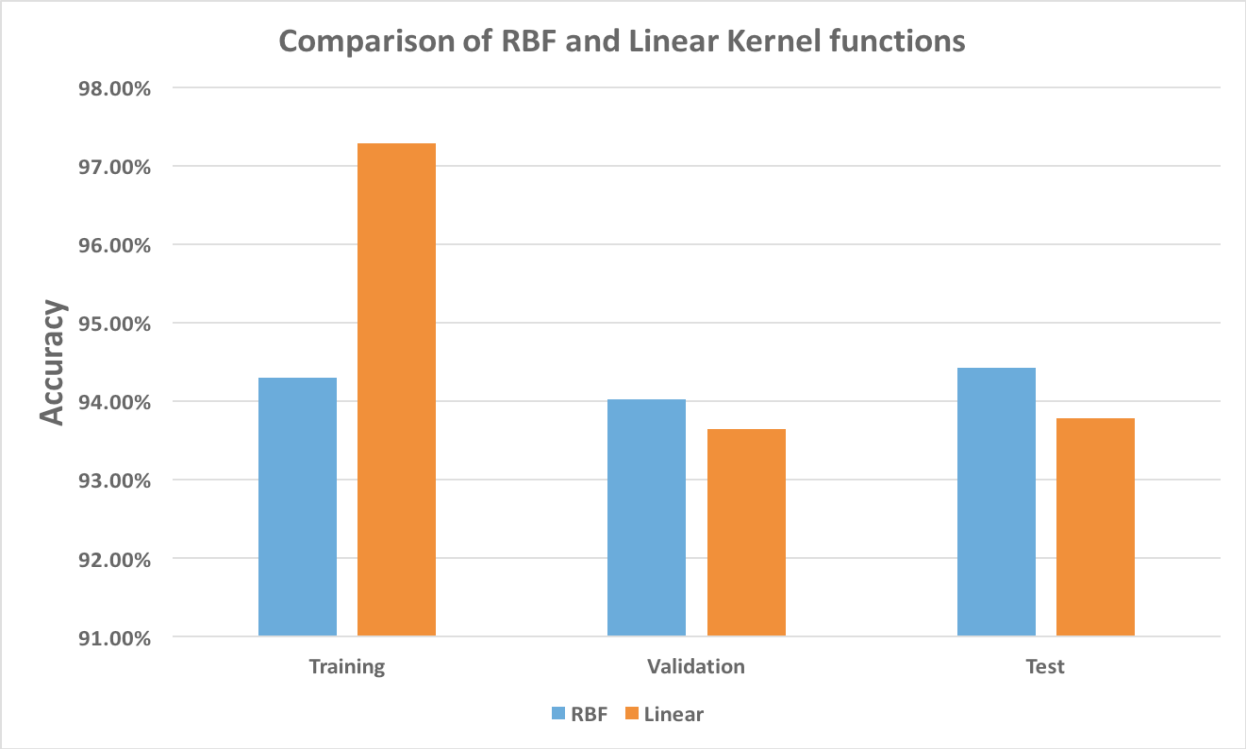

Figure 2: Comparison of RBF and Linear Kernel functions

DIFFERENT GAMMA VALUES:

The gamma parameter defines how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close’. So Gamma value 1 means, every single training example is important and it’s essential to classify every single example correctly. Not only it takes a lot of time to classify all the training examples, the impact of noise present in the training data is also high. This results in the poor performance in validation and test data.

This is a classic over fitting example. The SVM basically memorizes all the points.

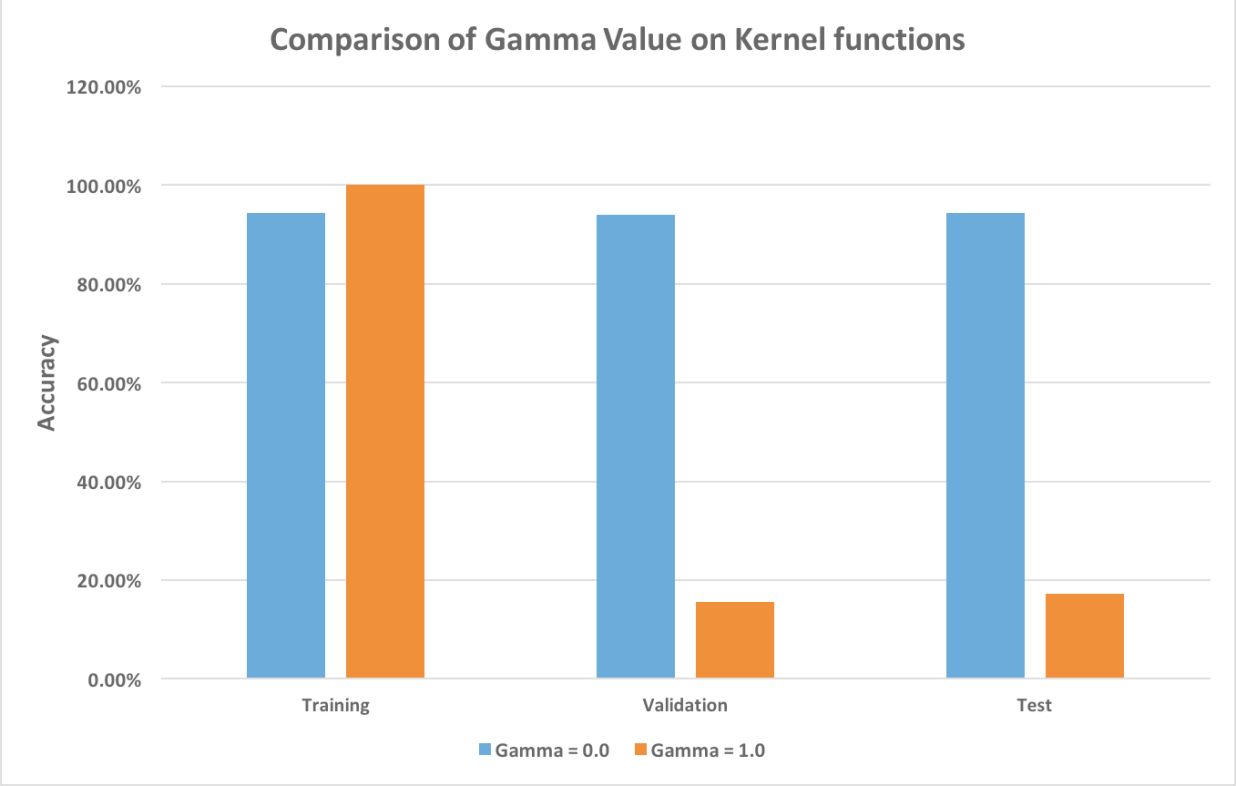

Figure 3: Comparison of Gamma Value on Kernel functions

DIFFERENT C VALUES:

C value specifies the tolerance level in misclassifying the training data. So smaller value of C causes SVM to look for a larger margin separating hyper plane even if it means misclassifying some training points. A low C makes the decision surface smooth, while a high C aims at classifying all training examples correctly by giving the model freedom to select more samples as support vectors.

Higher C value indicates smaller margin separating hyper plane. In higher C values, the training error will be mostly zero.

From the data shown below, it is clear that higher the value of C, higher the training accuracy but there is no such direct correlation in the test data. Choosing proper C, depends upon the data and it should be done experimentally as there is no clear way to choosing the right C for any data.

C	Training	Validation	Test
1	94.29%	94.02%	94.42%
10	97.13%	96.18%	96.10%
20	97.95%	96.90%	96.67%
30	98.37%	97.10%	97.04%
40	98.71%	97.23%	97.19%
50	99.00%	97.31%	97.19%
60	99.20%	97.38%	97.16%
70	99.34%	97.36%	97.26%
80	99.44%	97.39%	97.33%
90	99.54%	97.36%	97.34%
100	99.61%	97.41%	97.40%
1000	100.00%	97.36%	97.33%
10000	100%	97.35%	97.30%
100000	100%	97.35%	97.30%

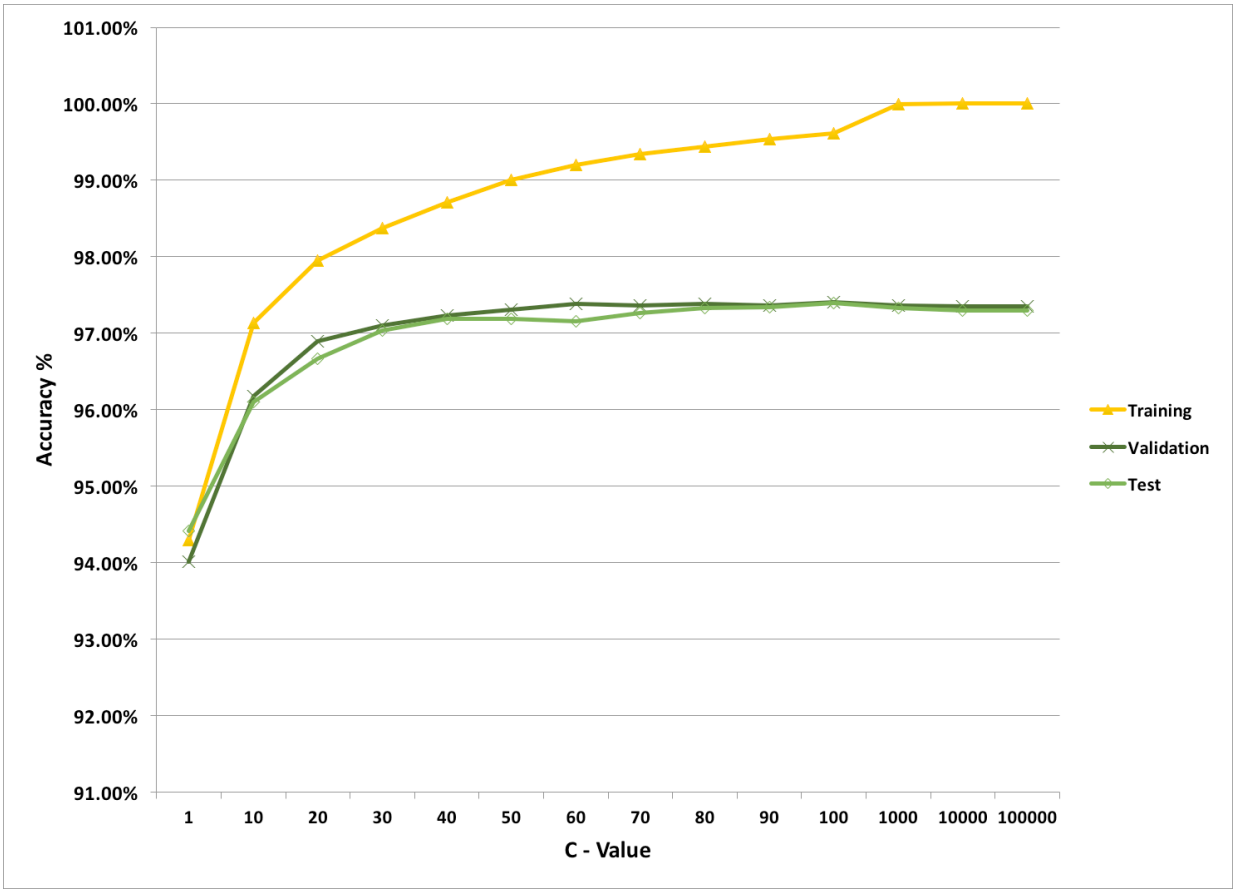

Figure 4: Training, Validation and Test Accuracy Vs C-value

Conclusion:

Logistic regression is great when the training data has fewer features ie. lesser dimensions. SVMs do better when there's a higher number of dimensions, and especially on problems where the predictors determine the responses. In SVM only points near the boundary influence the decision boundary but in logistic regression, all the points influence it. In this mnist data set, SVM is the clear winner because the data is not linearly separable has low noise and higher dimension of data.

References:

[1] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.141.880&rep=rep1&type=pdf
[2] http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf