Report: Programming Assignment 3 (CSE 574)
Group Number: 58

Team:

Roopali Vij (roopaliv)

Shreeju Jayesh Tanna (shreejuj)

Ranganatha Poola Narayana Swamy (rpoolana)

Report 1

Below are the accuracies obtained for classification of handwritten digits using the *'one-vs-all'* strategy:

Logistic Regression (One vs All):

Training set Accuracy: 86.194%
Validation set Accuracy: 85.33%
Testing set Accuracy: 85.42%

Report 2

Direct Multi-Class Logistic Regression

Logistic Regression (Multi-Class)

Training set Accuracy: 93.4%
Validation set Accuracy: 92.2%
Testing set Accuracy: 92.5%

Binary logistic regression is preferred when data is classified into two classes; whereas multi-class logistic regression is preferred when data is classified into multiple classes.

The *'one-vs-all'* strategy uses 10 binary logistic classifiers, but the multi-class logistic classifier can classify 10 classes at the same time.

Because the given dataset is classified over 10 classes, multi-class logistic regression has a higher accuracy when classifying it, compared to binary logistic regression.

Report 3

The accuracies for SVM are as follow:

Kernel = Linear
Training set Accuracy:97.286%
Validation set Accuracy:93.64%
Testing set Accuracy:93.78%

RBF with gamma = 1
Training set Accuracy:100.0%
Validation set Accuracy:15.48%
Testing set Accuracy:17.14%

RBF with default gamma
Training set Accuracy: 94.294%
Validation set Accuracy: 94.02%
Testing set Accuracy: 94.42%

RBF with default gamma and C = 1
Training set Accuracy:94.294%
Validation set Accuracy:94.02%
Testing set Accuracy:94.42%

RBF with default gamma and C = 10
Training set Accuracy:97.132%
Validation set Accuracy:96.18%
Testing set Accuracy:96.1%

RBF with default gamma and C = 20
Training set Accuracy:97.952%
Validation set Accuracy:96.9%
Testing set Accuracy:96.67%

RBF with default gamma and C = 30
Training set Accuracy:98.372%
Validation set Accuracy:97.1%
Testing set Accuracy:97.04%

RBF with default gamma and C = 40
Training set Accuracy:98.706%
Validation set Accuracy:97.23%
Testing set Accuracy:97.19%

RBF with default gamma and C = 50
Training set Accuracy:99.002%
Validation set Accuracy:97.31%
Testing set Accuracy:97.19%

RBF with default gamma and C = 60
Training set Accuracy:99.196%
Validation set Accuracy:97.38%
Testing set Accuracy:97.16%

RBF with default gamma and C = 70
Training set Accuracy:99.34%
Validation set Accuracy:97.36%
Testing set Accuracy:97.26%

RBF with default gamma and C = 80
Training set Accuracy:99.438%
Validation set Accuracy:97.39%
Testing set Accuracy:97.33%

RBF with default gamma and C = 90
Training set Accuracy:99.542%
Validation set Accuracy:97.36%
Testing set Accuracy:97.34%

RBF with default gamma and C = 100
Training set Accuracy:99.612%
Validation set Accuracy:97.41%
Testing set Accuracy:97.4%

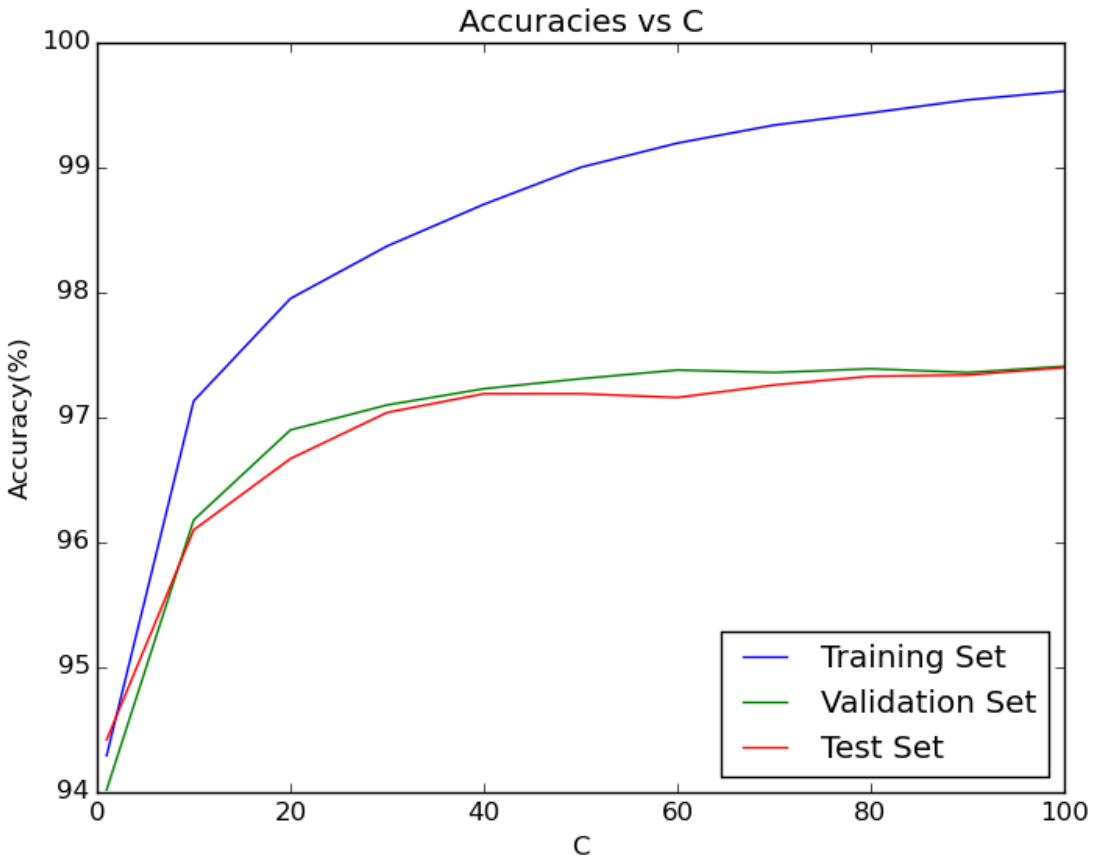
Linear Kernels: They tend to perform very well when the decision boundary is linear. However, when the boundary is not linear, as in our case, the performance was moderate with accuracy of 85% on test data.

RBF Kernels: They perform better with non-linear boundaries. We observed accuracies of over 97% on test data with C set to 100 and gamma set to default.

Gamma: When Gamma is set to This parameter defines how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close’. When gamma was set to low (gamma = 1), we observed that the classification accuracy was 100% on training data, but performing bad with 17% accuracy on test data

C: The C parameter trades off misclassification of training examples against simplicity of the decision surface. 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.

We observed that as the value of C increased from 1 to 100, accuracy of classification improved from 94% at C=1 to 97% at C=100



Based on our observations, the linear kernel has a marginally higher accuracy while classifying training set, compared to radial basis function (with low C values and default gamma). However, the radial basis function has a better accuracy when classifying validation and test data. As we increase the value of C, the accuracy of RBF kernel increases.