CSE 574

PROGRAMMING ASSIGNMENT 3

Project Report

Classification and Regression

Group 24

Ayush Jitendra Sanghvi – 50132376

Avani Chandrashekar Atre - 50134720

Priyankar Nandi – 50134010

Logistic Regression:

Training set Accuracy: 92.328%
Validation set Accuracy: 91.46%
Testing set Accuracy: 91.92%

Above are the accuracies observed for Training set Validation set and Test set for logistic regression.

Support Vector Machines:

1. **Linear Kernel**
   Training set Accuracy: 97.286%
   Validation set Accuracy: 93.64%
   Test set Accuracy: 93.78%

   Linear kernel is useful given that the data is multi-dimensional and the features are informative. MNIST data set consist of digit image data set which are high dimensional but the pixels are very less informative hence the accuracy is not as high as it would be for non-linear models.

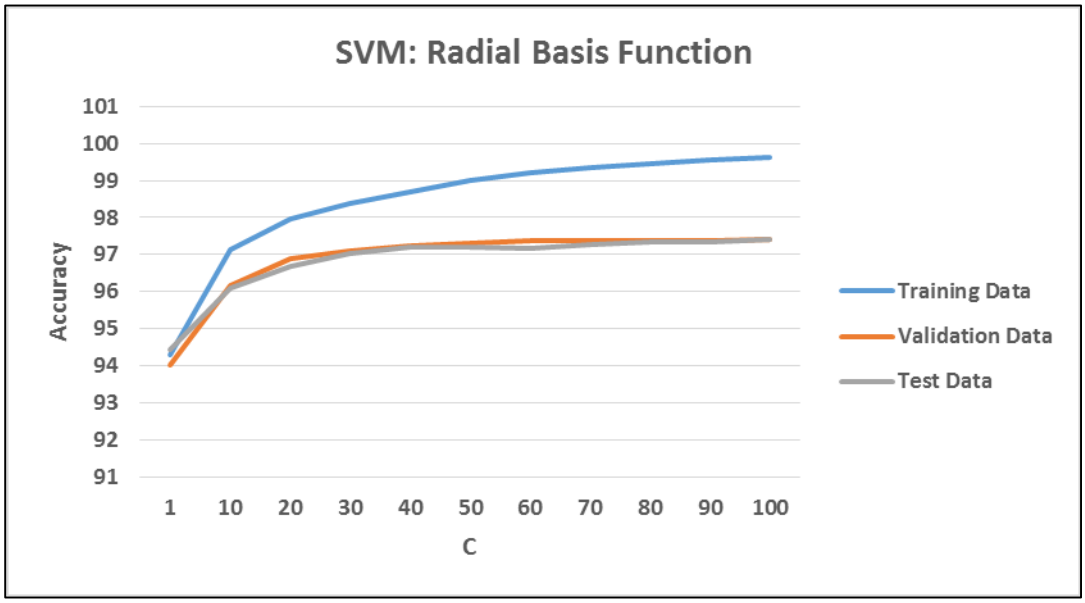
2. **Radial basis function , gamma 1**
   Training set Accuracy: 100.0%
   Validation set Accuracy: 15.48%
   Test set Accuracy: 17.14%

3. **Radial basis function , gamma default**
   Training set Accuracy: 94.294%
   Validation set Accuracy: 94.02%
   Test set Accuracy: 94.42%

   Low accuracies are observed in case of gamma = 1 compared to gamma default which shows us a clear case of over fitting.

4. **Radial basis function , gamma default , C varied from 1 to 100**

C	Training data accuracy(%)	Validation data accuracy(%)	Test data accuracy(%)
1	94.294	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



SVM observation

- As seen from the graph, it can be concluded that with increase in C accuracy increases.
- In the support-vector networks algorithm one can control the trade-off between complexity of decision rule and frequency of error by changing the parameter C.
- In any regularization scheme, it is important to choose proper value for C, also known as the penalty factor. If it is too large, we have a high penalty for non-separable points and we may store many support vectors and overfit. If it is too small, we may have underfitting.