CSE 574 INTRODUCTION TO MACHINE LEARNING

MAY 3, 2017

PA3: CLASSIFICATION AND REGRESSION

SUBMITTED BY: ABHISHEK YEOLA

ROSHAN BABY REJI

CSE574 Machine Learning Programming Assignment 3

Group 6
Abhishek Yeola
Roshan Baby Reji

May 2, 2017

Contents

1 Introduction

In this project, the popular problem of classifying handwritten digits using logistic Regression and Support Vector Machines.

2 Handwritten Digits Classification and Regression

2.1 Comparison between BLR and MLR

Binary Logistic Regression(BLR)
Multinomial Logistic Regression(MLR)

Logistic Regression	*Accuracy*		
	Training Data	*Validation Data*	*Test Data*
BLR	84.898	83.73	84.22
MLR	93.448	92.48	92.55

We can observe from the above figure that the MLR method gives better accuracy on training, test and validation data as compared to BLR method.

2.2 Comparison between SVM and Logistic Regression

SVM	*Accuracy*		
	Training Data	*Validation Data*	*Test Data*
BLR	84.898	83.73	84.22
MLR	93.448	92.48	92.55
Linear Kernel	97.286	93.64	93.78
RBF,Gamma = 1	100.0	15.48	17.14
RBF,Gamma = 0	94.294	94.02	94.42
RBF,C=1	94.294	94.02	94.42
RBF,C=30	98.372	97.1	97.04
RBF,C=50	99.002	97.31	97.19
RBF,C=80	99.438	97.36	97.34
RBF,C=100	99.612	97.41	97.4

As we can see from the data, SVM gives better performance than Logistic Regression on all 3 data sets.

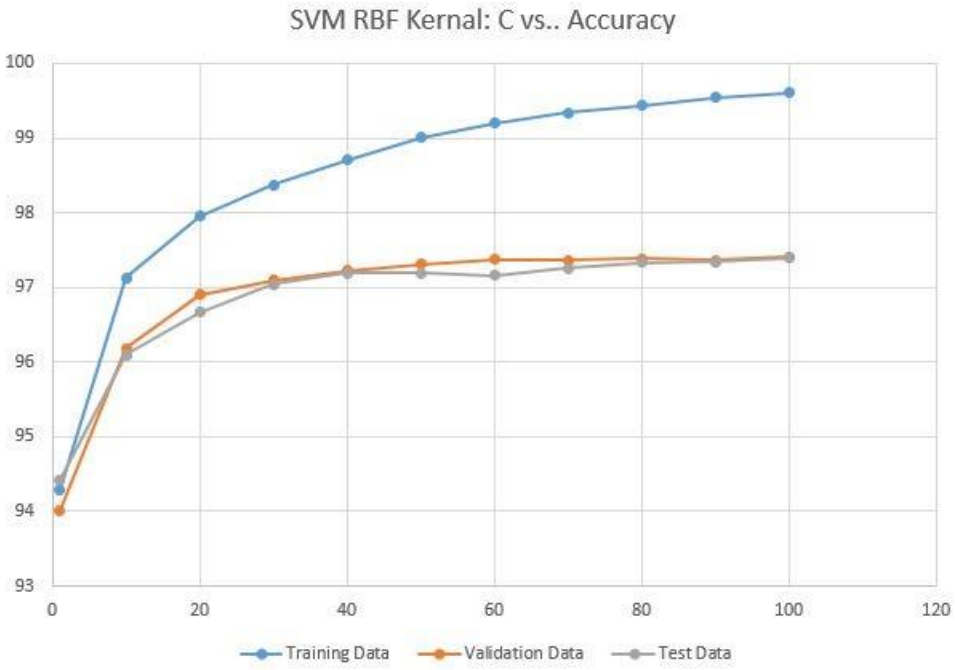
This is because Logistic Regression learns any hyperplane that separates the data to maximize the posterior class probability, while SVM tries to maximize the margin between the closest support vectors.

SVM is more robust with outliers as compared with Logistic Regression.

Logistic Regression performs well on low dimension spaces. On the other hand, SVM gives better results where the number of dimensions is significantly greater than the number of samples.

Gamma defines how far the influence of a single training example reaches, which a low value meaning far and high values meaning close. We can see from the table that for gamma=1 we have overfitting and very low accuracy on both validation and training data sets even though we have 100% accuracy on the training data set. For gamma=0 we have comparatively better results for accuracy on the validation and test data sets.

2.3 Accuracy of SVM with RBF Kernel for different values of 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 training examples correctly by giving the model freedom to select more samples as support vectors. In short,it controls the complexity of the hyperplane.

We can see from the plot that as the value of C increases, the accuracy of the solution also increases.
However, as the value of C increases, there is a risk of over-fitting the data on the training set.
Therefore we select the value of C between 20 and 40 which gives a good fit on the training, test and validation data sets.