University at Buffalo
CSE 574 Introduction To Machine Learning, Spring 2017

Programming Assignment 3
Classification and Regression

Project Report

Submitted by

Group 49

Anurag Devulapalli (5020 8153)
Sandeep Shenoy (5020 5705)
Vipin Kumar (5020 8397)
April 11, 2017

Problem 1: Logistic Regression

Requirement:

You are asked to implement Logistic Regression to classify hand-written digit images into correct corresponding labels

a. In your report, record and discuss classification results and accuracy.

Training set Accuracy	: 84.944%
Validation set Accuracy	: 83.71%
Testing set Accuracy	: 84.19%

Analysis:

Logistic regression considers all the data points to construct the hyperplane separating the data as per the classes. Hence Logistic regression works well on data having low input features.

The data provided is not linearly separable and has high input features (dimensions). Our experiment resulted in approximately 84% accuracy across all three data sets.

Problem 2: Direct Multi-class Logistic Regression

Requirement:

In this part, you are asked to implement multi-class Logistic Regression. Traditionally, Logistic Regression is used for binary classification. However, Logistic Regression can also be extended to solve the multi-class classification. With this method, we don’t need to build 10 classifiers like before. Instead, we only need to build 1 classifier that can classify 10 classes at the same time.

a. In your report, record and discuss classification results and accuracy.

Training set Accuracy	: 93.162%
Validation set Accuracy	: 92.46%
Testing set Accuracy	: 92.51%

Our experiment resulted in approximately 92% accuracy across all three data sets.

b. Discuss the performance of multi-class logistic regression compared to the performance of logistic regression when using the one-vs-all strategy.

Regression Type	Training Accuracy	Validation Accuracy	Test Accuracy
Binary	84.944	83.71	84.19
Multinomial	93.162	92.46	92.51

We observe that multinomial regression provides a better accuracy compared to binomial regression. In our data set, each input belongs to exactly one class. Thus, multinomial regression performs better.

Problem 3: Support Vector Machines

Requirement:

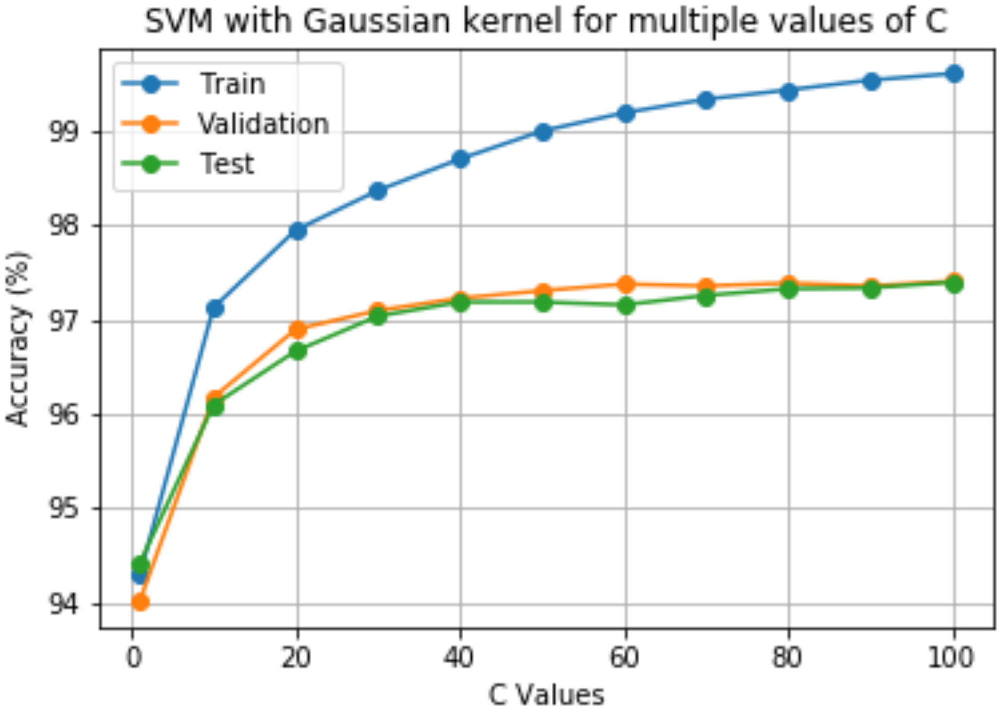
In this part of assignment, you are asked to use the Support Vector Machine tool in sklearn and svm. SVM to perform classification on our data set.

a. **In your report provide justification for the above selection of hyper parameters, as well as plots and discussion of your results.**

Kernel	Gamma	Training Accuracy	Validation Accuracy	Test Accuracy
Linear	0	97.286%	93.64%	93.78%
RBF	1	100%	15.48%	17.14%
RBF	0	94.294%	94.02%	94.42%

We observe that RBF kernel with gamma 0 provides an accuracy of 94.42% which is slightly better than Linear kernel which provides 93.78% accuracy. RBF kernel with gamma 1 gives 100% accuracy on training data but performs badly on validation and test which shows overfitting.

C	**Training Accuracy %**	**Validation Accuracy %**	**Test 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



The value of C, called penalty factor, in Gaussian Kernel model controls the impact of margin and the margin error. From the above plot we can observe that as the value of C increases the value of accuracy increases too. The impact of setting lower C value is the weight of each error term is low and a larger margin hyperplane is created. Similarly, the impact of setting higher C value is the increase in the weight of each error term and a smaller margin hyperplane. Thus, accuracy increases with higher C values. But again, having too high value of C value will cause over-fitting and having smaller C value will cause under fitting. Hence we should carefully select the C value.

b. Discuss results, comparing the selections of linear kernel and radial basis function kernel.

Kernel	Gamma	Training Accuracy	Validation Accuracy	Test Accuracy
Linear	0	97.286%	93.64%	93.78%
RBF	1	100%	15.48%	17.14%
RBF	0	94.294%	94.02%	94.42%

Using SVM with the Gaussian kernel gives higher accuracy because of the flexibility to transform the data into a space of any dimension.

And because of this reason when we have higher number features than observations we prefer Linear Kernel, but when we have higher number of observations than number of features we prefer Gaussian Kernel.

Gamma value controls the influence of each training data on the learned hyperplane. When Gamma = 1, despite getting an accuracy of 100% on the Training dataset, we have observed a very low accuracy for both Validation and Test dataset. Clearly a case of over-fitting.

When Gamma = 0, the results improve on both Validation and Test set as it closely resembles the Linear kernel model.