CSE574 Introduction to Machine Learning
Programming Assignment 3
Classification and Regression

yuxiang liu <yliu268@buffalo.edu>
shaoming xu < shaoming@buffalo.edu >

Problem 1: Implementation of Logistic Regression

1. In your report, you should train the logistic regression using the given data X (Preprocessed feature vectors of MNIST data) with labels y. Record the total error with respect to each category in both training data and test data. And discuss the results in your report and explain why there is a difference between training error and test error.
    a. After training the logistic regression using the given data X and comparing the predicted labels with the true labels, we get this result: Training set Accuracy:86.094%, Validation set Accuracy:85.17%, Testing set Accuracy:85.39%.
    b. For MINIST data in our experiment, the accuracy of Testing set is smaller than the Training set for 1%. And this difference is made by the overfitting.

Problem 2: Multi-class Logistic Regression

1. In your report, you should train the logistic regression using the given data X (Preprocessed feature vectors of MNIST data) with labels y. Record the total error with respect to each category in both training data and test data. And discuss the results in your report and explain why there is a difference between training error and test error. Compare the performance difference between multi-class strategy with one-vs-all strategy.
    a. After training the multi-class logistic regression using the given data X and comparing the predicted labels with the true labels, we get this result: Training set Accuracy:93.484%, Validation set Accuracy:92.35%, Testing set Accuracy:92.6%.
    b. the accuracy of Testing set is smaller than the Training set for 1%. And this difference is made by the overfitting.
    c. The performance of multi-class logistic regression is better than that of traditional logistic regression. This is because the multi-class logistic regression takes all weights into consideration but the traditional only use one weight in training for each digit.

Problem 3: Support Vector Machines

1. Performance differences between linear kernel and radial basis, different gamma setting.
    a. First we compare the linear kernel, RBF kernel with gamma=1, and RBF kernel with gamma=default.
        i. First we consider the accuracy of training data, in the Figure3.1 we can see the RBF(gamma=1) > Linear kernel >> RBF(gamma=default). And accuracy of the RBF(gamma=1) even reaches100%.
        ii. Then we consider validation data and Testing data, we can see the RBF(gamma=default) > Linear kernel >> RBF(gamma=1). And now both the accuracies of Validation and Testing data of RBF(gamma=1) are around 20%, which are much lower than those of linear kernel and RBF(gamma=default). So it shows RBF(gamma=1) has serious overfitting problem. And the accuracy of RBF(gamma=default) on validation and testing data has around 1% greater than those of linear kernel. If consider the lower accuracy of training data in RBF(gamma=default), the RBF(gamma=default) is better than linear kernel because it suffers less overfitting problem.
        iii. When gamma=default, the gamma will be set as 1/n_features which is much smaller than gamma=1. The larger gamma means the Gaussian function with a small variance. So larger gamma tends to make overfitting in training phase just as showed in Figrue3.1.

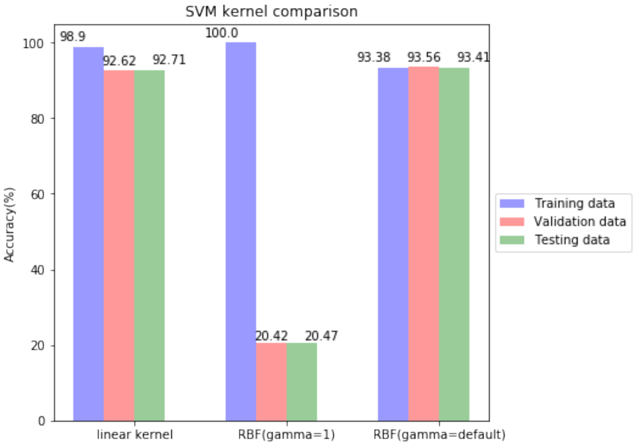

Figure3.1

b. The figure3.2 shows the accuracy with respect to values of C in the radial basis function with value of gamma setting to default.
    i. In the figure we can see with the C increasing from 1 to 100, the the accuracy for the training data keep increasing and closing to 100%. However, both the accuracies for the validation and testing data become stable after C is greater than 20. The accuracies are between 95.5% to 96%. It means the overfitting problem happens when C becomes greater than 20.
    ii. C is the penalty parameter for error term. The large C means the high penalty which will cause overfitting in training phase of SVM model.

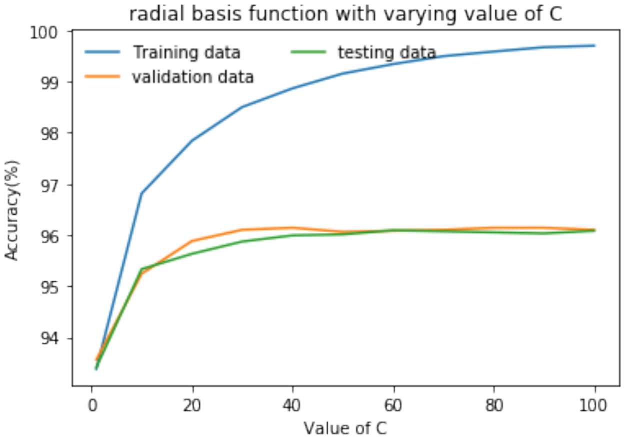

Figure3.2