**CSE 574: Introduction to Machine Learning**

**Programming Assignment 3**

**Group 79**

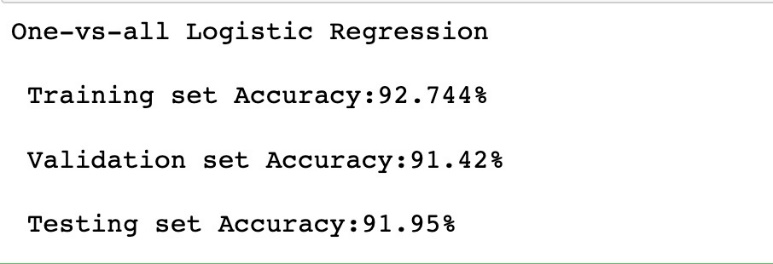
**Chihuangji Wang (50130295)**

**Wei Zhang (50240861)**

**Sonal Singh (50290784)**

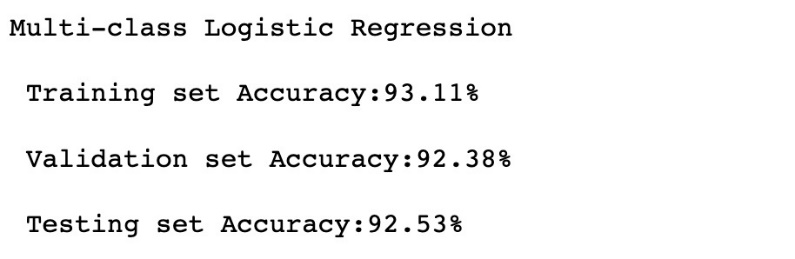
1. **One-vs-all Logistic Regression**

The overall accuracy for training, validation, and testing data are presented as below:



1. **Multi-class Logistic Regression**

The overall accuracy for training, validation, and testing data are presented as below:

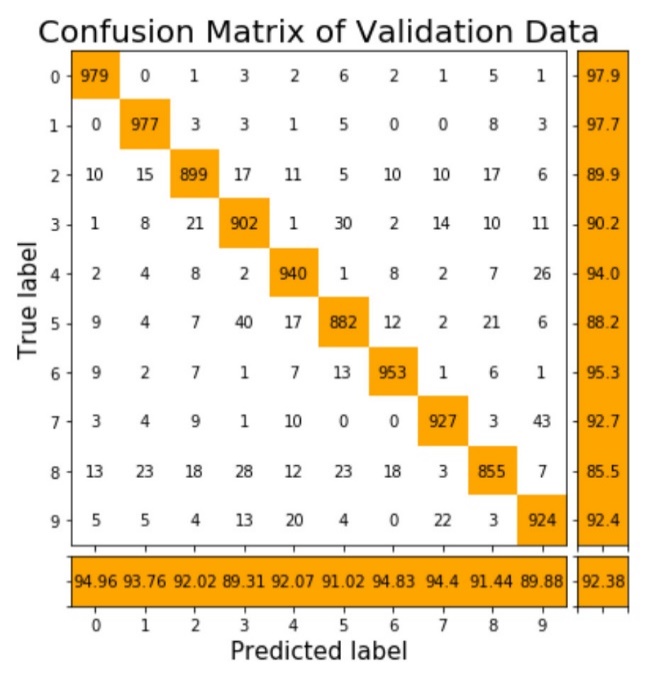
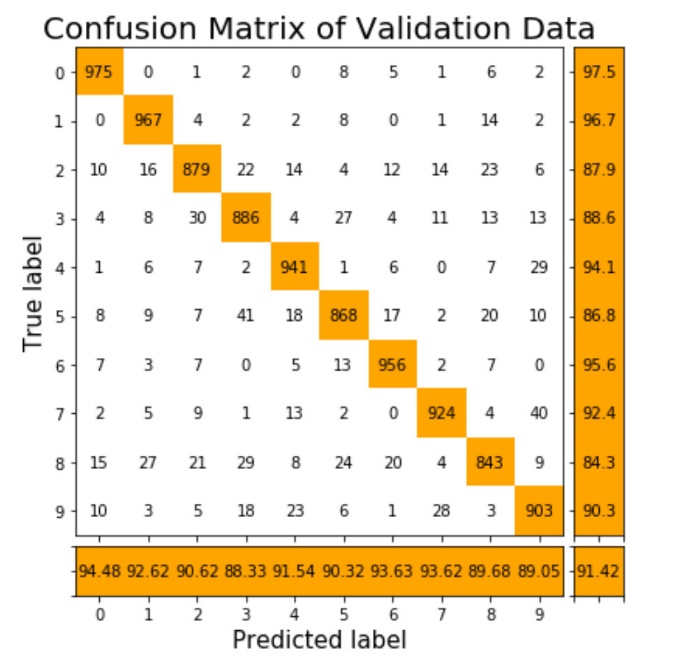
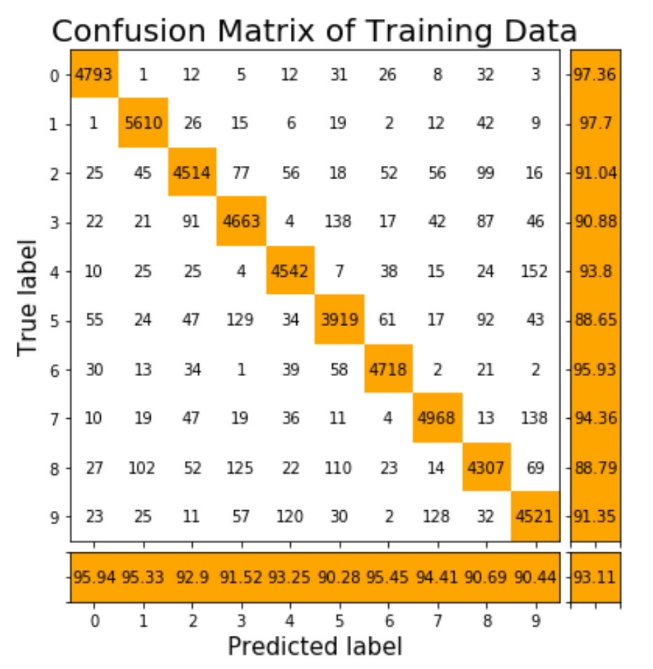
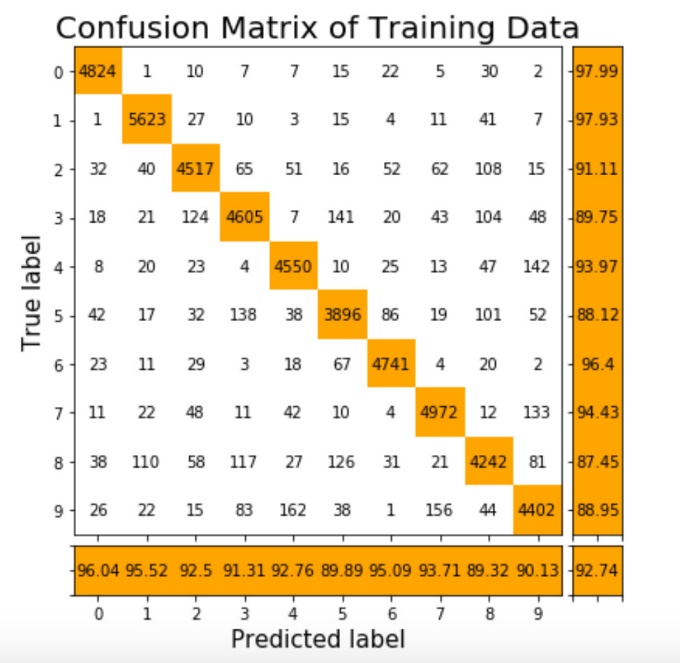
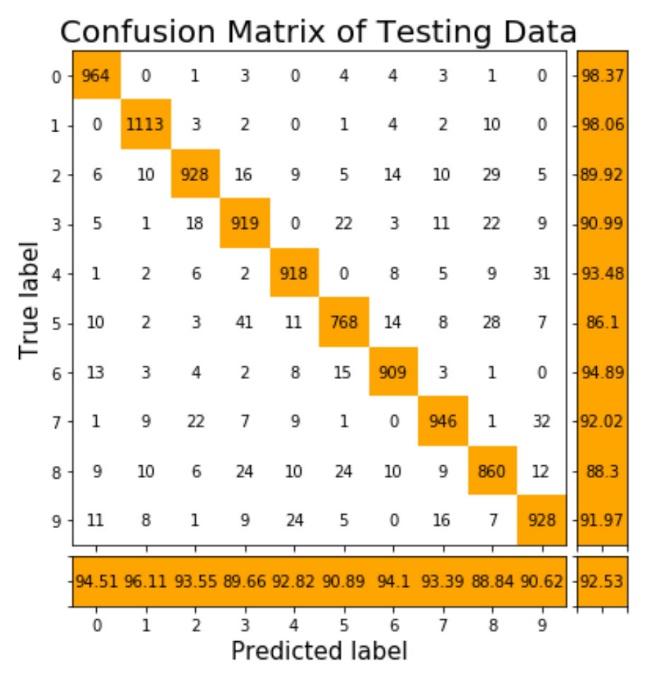
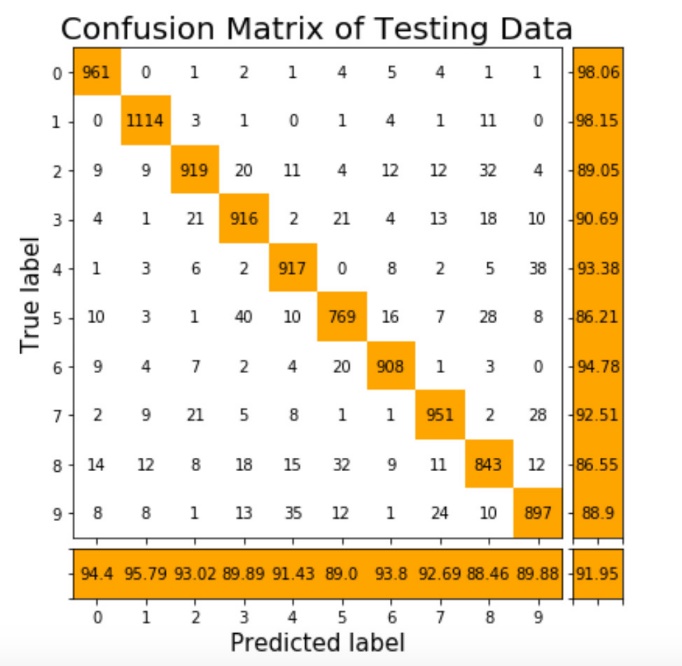


1. **Comparison of One-vs-all and Multi-class Logistic regressions**

As we can see above, the testing accuracy for Multi-class Logistic regression is a little bit higher than that of One-vs-all Logistic Regression, meaning Multi-class Logistic regression performs better than the other at predicting hand-writing images. Also, it has higher training and validation accuracies as well. The explanation of this can be that, since there are 10 labels/classes, one-vs-all classifier can only use a highly imbalanced dataset (e.g. each class only consists of one tenth of the whole data set) to determine whether an image belongs to the particular class or not, which may degrade performance. While multi-class classifier can run multiple classifiers on one image to choose the most possible “voted” class/label. So, in our case, multi-class logistic regression performs better.

We also plot the confusion matrix for each data set separately as below. From the confusion matrix, averagely speaking, Multi-class Logistic regression has higher accuracies for each label/class.

***One-vs-all Logistic Regression Multi-class Logistic Regression***

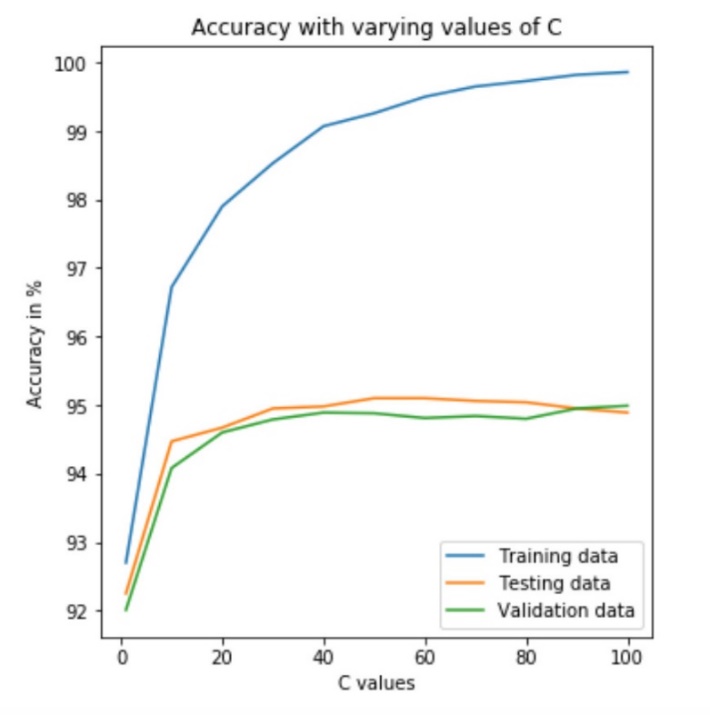
1. **Support Vector Machines (SVM)**

The table below shows the results from SVMs with (1) linear kernel, (2) radial basis function with gamma = 1, and (3) radial basis function with gamma setting and other parameters setting to default. As we can see from the table, when setting gamma = 1, the training accuracy reaches 100%, which indicates overfitting issue, thus testing and validation accuracy are both very low. When comparing the other two SVMs (linear kernel and radial basis function with default setting), we can find that the testing and validation accuracy are pretty similar, while the training accuracy from linear kernel is very high.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear kernel | Radial basis function (gamma = 1) | Radial basis function (setting to default) |
| Training Accuracy | 99.69% | 100.0% | 92.7% |
| Validation Accuracy | 91.87% | 18.4% | 92.01% |
| Testing Accuracy | 92.21% | 19.14% | 92.25% |

We also apply radial basis function with value of gamma setting to default and varying C value of C (from 1, 10,… to 100). The plot below shows the of accuracy with respect to values of C. Training accuracy increases as C value goes up, which is easy to understand as larger C means that a smaller margin in the decision function will be accepted when classifying training points. From the plot and the table below, we decide to use C= 50, since that gives us the highest testing accuracy, as well as validation accuracy.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| C-Value | 1 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
| Training Accuracy | 92.7% | 96.72% | 97.90% | 98.53% | 99.07% | 99.26% | 99.50% | 99.65% | 99.73% | 99.82% | 99.86% |
| Validation Accuracy | 92.01% | 94.08% | 94.60% | 94.79% | 94.89% | 94.88% | 94.81% | 94.84% | 94.80% | 94.95% | 94.99% |
| Testing Accuracy | 92.25% | 94.47% | 94.67% | 94.95% | 94.98% | 95.10% | 95.10% | 95.06% | 95.04% | 94.95% | 94.89% |



So, the final result from the optimal SVM, which is using radial basis function with value of gamma setting to default and C= 50, is presented as below. And we can see that all three accuracies all increase dramatically.

Training Accuracy: 99.612 %

Testing Accuracy: 97.40 %

Validation Accuracy: 97.41 %