Assignment 3 - Logistic Regression and SVM
CSE 474 Group 22
David Olsen, Der Shen Tan, Hoan Duc Tran

**Logistic Regression Results:**

Training set Accuracy:92.77%
Validation set Accuracy:91.49000000000001%
Testing set Accuracy:91.96%

As these numbers clearly show, our logistic regression was very successful at predicting labels for the data set we were given.

As a side note, changing the max_iter option down a bit does result in slightly worse result, but overall not noticable.

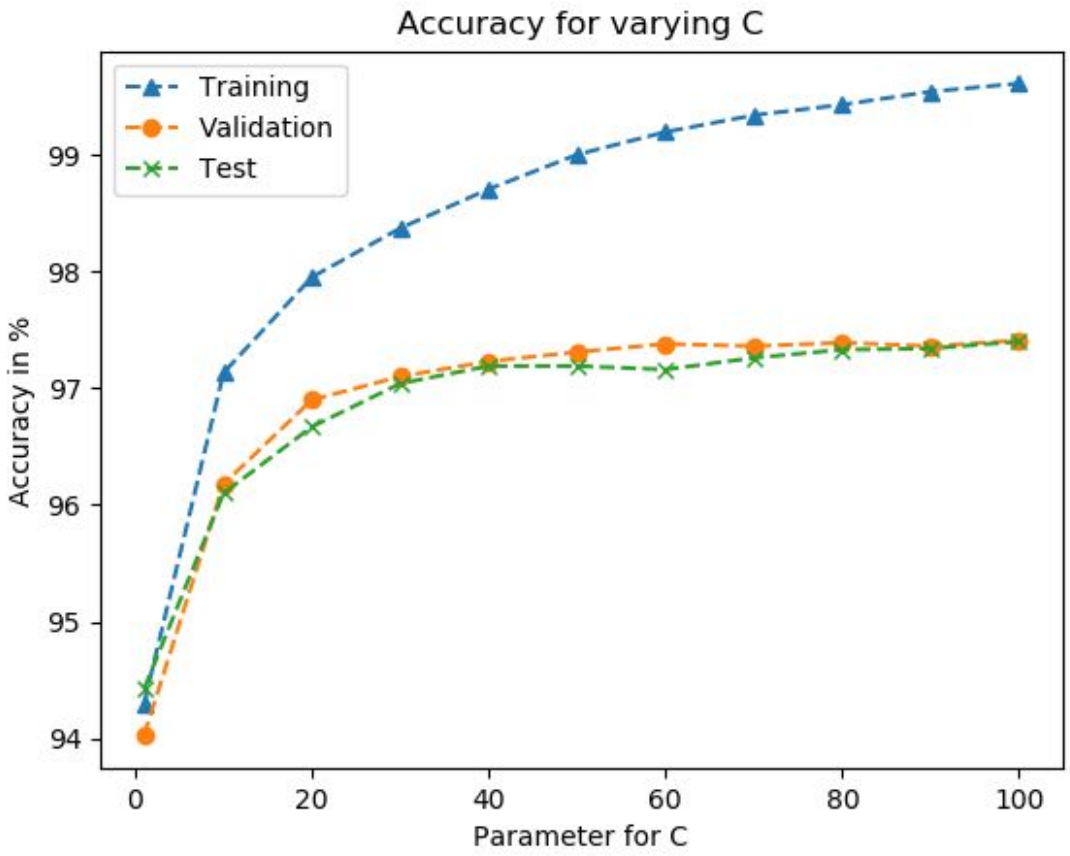
**Multilabel Logistic Regression Results (Extra Credit):**

Training set Accuracy:93.096%
Validation set Accuracy:92.45%
Testing set Accuracy:92.52%

Multilabel regression was similar in accuracy to single label, but slightly better across the board.

Classification time was also better.

**SVM Results:**
For Varying parameter of C, the figure belows illustrate the result:




According to the figure, both training, validation and testing accuracy increase. Testing and validating seems to capped at around 97.4 % at max, while training accuracy approaches 100%. This suggest that higher C means more fitting toward the training data, which would greatly boosts testing and validating accuracy, as the data set are from similar distribution. (similar results with Programming assignment 1).

The accuracy output for the rest of SVM test were as follow:
**Linear SVM Accuracy:**
Training set Accuracy:97.286%
Validation set Accuracy:93.64%
Testing set Accuracy:93.78%

**RADIAL SVM GAMMA 0.1 Accuracy:**
Training set AccurSVM GAMMA 0.1 Accuracy:
Training set Accuracy:99.992%
Validation set Accuracy:94.76%
Testing set Accuracy:94.96%

**RADIAL SVM Accuracy:**
Training set Accuracy:94.294%
Validation set Accuracy:94.02000000000001%
Testing set Accuracy:94.42%

With Gamma = 0.1, the result for all data set seems to be the highest. Also noticably noticed is the time to train Gamma = 0.1 is much much longer, which might contributes to almost perfect training result.

The whole result can be found as follow (std output from our program).

Training set Accuracy:92.77%
Validation set Accuracy:91.49000000000001%
Testing set Accuracy:91.96%

----------SVM----------
Linear SVM Accuracy:
Training set Accuracy:97.286%
Validation set Accuracy:93.64%
Testing set Accuracy:93.78%
RADIAL SVM GAMMA 0.1 Accuracy:
Training set AccurSVM GAMMA 0.1 Accuracy:
Training set Accuracy:99.992%
Validation set Accuracy:94.76%
Testing set Accuracy:94.96%
RADIAL SVM Accuracy:
Training set Accuracy:94.294%
Validation set Accuracy:94.02000000000001%
Testing set Accuracy:94.42%
RADIAL SVM Varying C value Accuracy:
C = 10.0
Training set Accuracy:97.13199999999999%
Validation set Accuracy:96.17999999999999%
Testing set Accuracy:96.1%
C = 20.0
Training set Accuracy:97.952%
Validation set Accuracy:96.89999999999999%
Testing set Accuracy:96.67%

C = 30.0
Training set Accuracy:98.372%
Validation set Accuracy:97.1%
Testing set Accuracy:97.04%
C = 40.0
Training set Accuracy:98.706%
Validation set Accuracy:97.23%
Testing set Accuracy:97.19%
C = 50.0
Training set Accuracy:99.002%
Validation set Accuracy:97.31%
Testing set Accuracy:97.19%
C = 60.0
Training set Accuracy:99.196%
Validation set Accuracy:97.38%
Testing set Accuracy:97.16%
C = 70.0
Training set Accuracy:99.33999999999999%
Validation set Accuracy:97.36%
Testing set Accuracy:97.26%
C = 80.0
Training set Accuracy:99.438%
Validation set Accuracy:97.39%
Testing set Accuracy:97.33000000000001%
C = 90.0
Training set Accuracy:99.542%
Validation set Accuracy:97.36%
Testing set Accuracy:97.34%
C = 100.0
Training set Accuracy:99.612%
Validation set Accuracy:97.41%
Testing set Accuracy:97.39999999999999%
----------SVM END----------

----------Multilabel Logistic Regression----------
Training set Accuracy:93.096%
Validation set Accuracy:92.45%
Testing set Accuracy:92.52%