CMPS 142 Machine Learning
Spring 2018, Homework #3 Part 2

Aaron Steele, atsteele@ucsc.edu
Tommy Tran, ttran56@ucsc.edu

2.1 Logistic Regression

1. Length of the weight vector: 1364

2. Norm of the learned weights = 35.9868

3. Accuracy = 0.9818

4. P, R, and F1 score of the positive class=0.8902 0.9829 0.9343

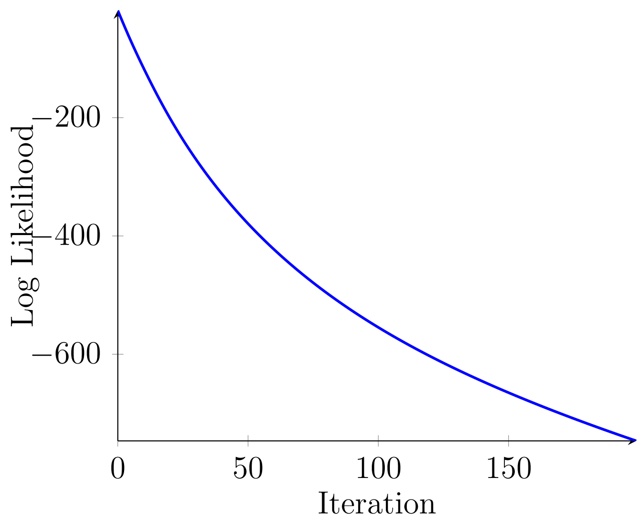
5. P, R, and F1 score of the negative class=0.9973 0.9816 0.9894

6. ```
   Confusion Matrix
   576 10
   71 3802
   ```

7. Accuracy = 0.9560

8. P, R, and F1 score of the positive class=0.8111 0.9068 0.8563
   P, R, and F1 score of the negative class=0.9839 0.9643 0.9740

9. ```
   Confusion Matrix
   146 15
   34 920
   ```

10. 

11. There are a lot more negative instances than positive ones. We could reduce the number of negative instances to account for the imbalance, or we could try using a different algorithm of some kind.

2.2 Logistic Regression with a Bias Term

1. Length of the weight vector: 1365

2. `Confusion Matrix`
   `144 17`
   `2 952`

3. The accuracy on the test set is 99% with the bias, and 95% without, so it does seem helpful. Although the train accuracy is 99%, so it seems as though there might be some overfitting going on.

2.3 L2-Regularized Logistic Regression

1. Length of the weight vector = 1364

2. `Confusion Matrix`
   `147 14`
   `33 921`

3. In 2.3 the training set accuracy goes down from 98% to 97%, but the test set accuracy goes up slightly. This implies that the amount of overfitting has decreased with the L2-regularization, which is the entire point.

4. Norm of the learned weights = 26.1916
   This value is quite a bit less than the one from 2.1. This says that the overall values of the weights have decreased, which is what we're going for with regularization, penalizing high weights.