Linear Models for Supervised Learning Reports

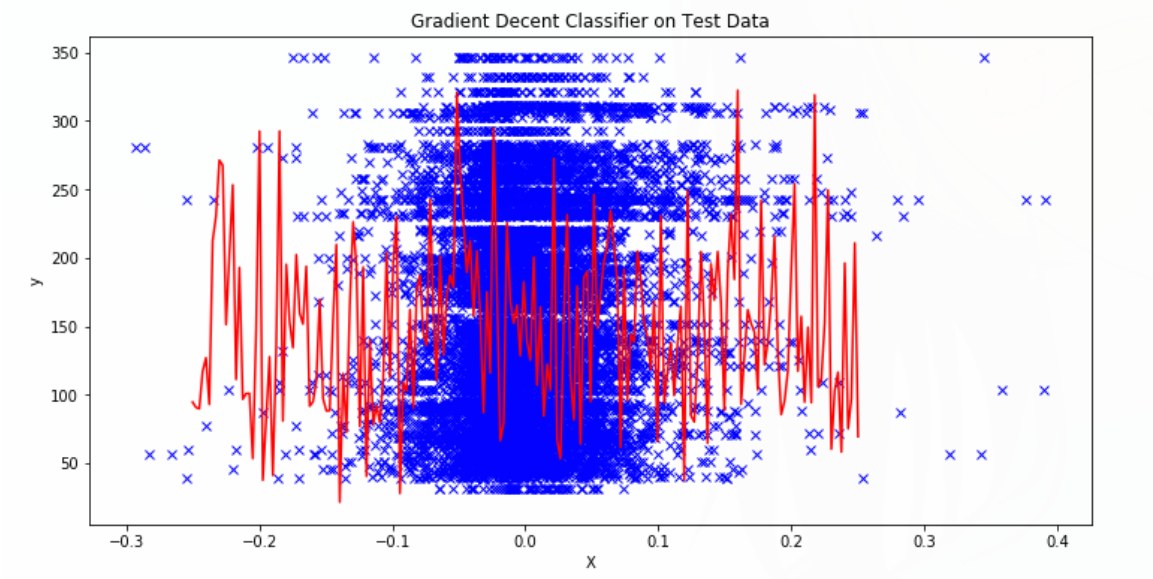
CSE 474 Group 41 Jeremy Baumann & Liam Carroll

REPORT 1

We calculated the RMSE on the training data to be 71.24 without the intercept and 72.29 with the intercept. We calculated the RMSE on the test data to be 88.91 without the intercept and 87.11 with the intercept. The intercept hurt the training data result because [???] but helped the testing data result because [???].

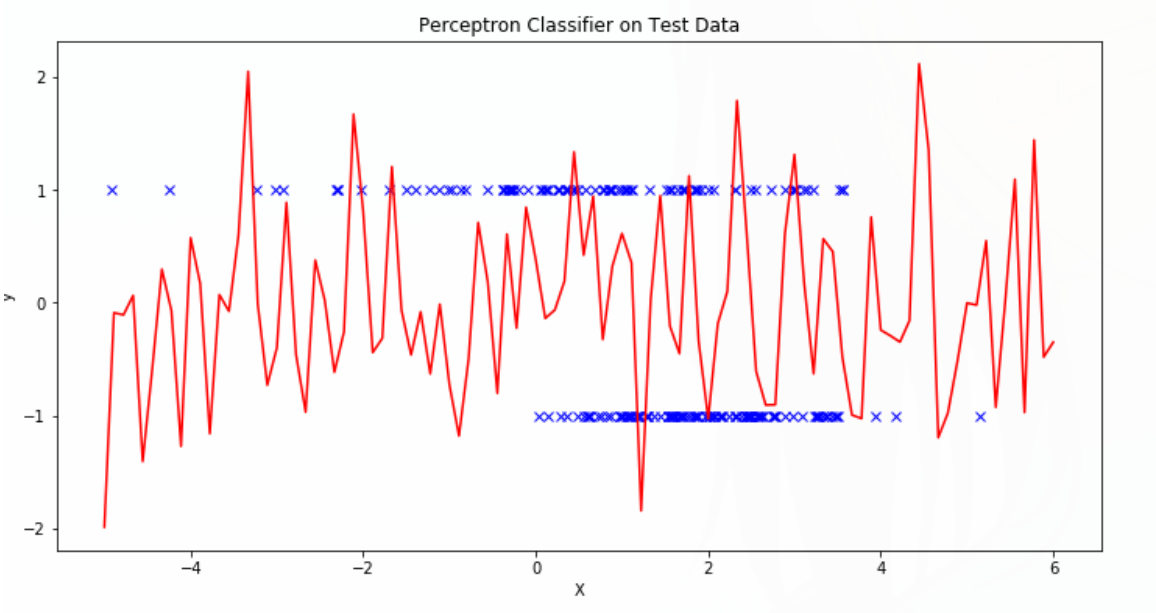
REPORT 2

We calculated the RMSE on the training data to be 72.64 and 86.48 on the test data. The RMSE for gradient decent based learning is higher than the RMSE for direct minimization. The following is a graph of the predicted values (red) and the test data (blue):



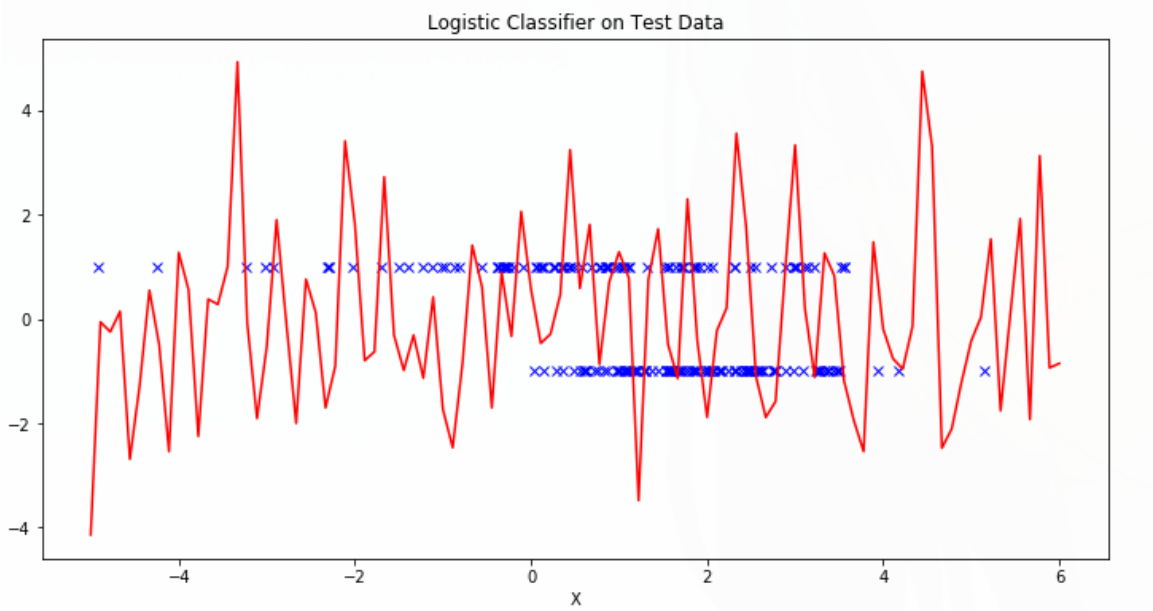
REPORT 3

The accuracy of our Perceptron model was calculated to be .71 on the training data and .81 on the testing data. The following is the classifier graphed on top of the test data:



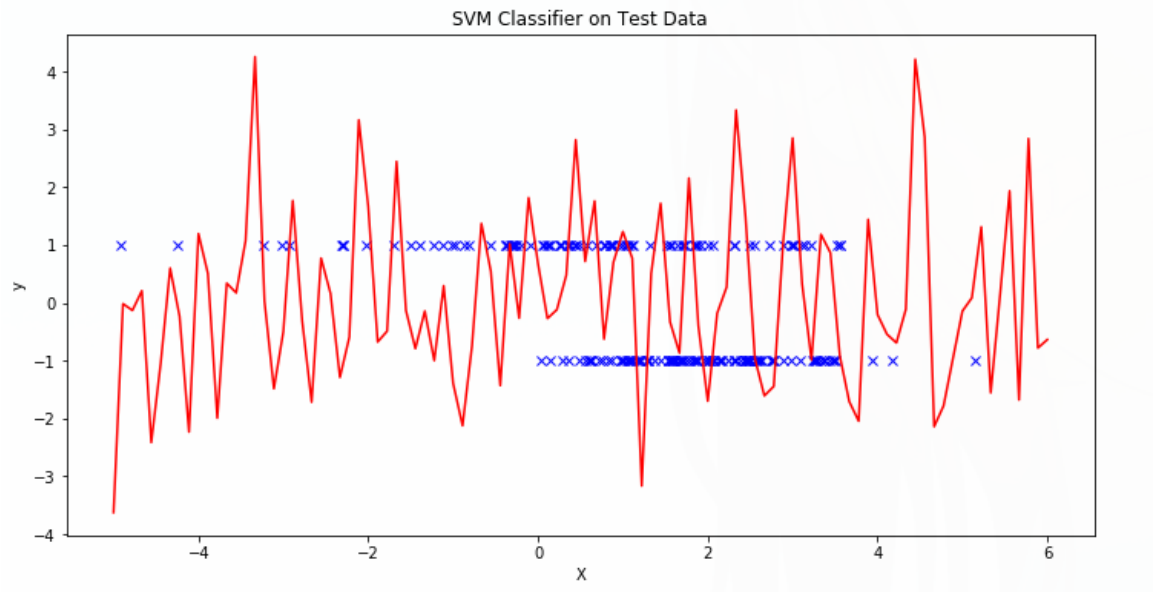
REPORT 4

The accuracy of our Logistic Regression model on the training data was 0.80 while the accuracy on the testing data was 0.83. This was better than our perceptron model by .09 on the training data and .02 on the test data. The following is the classifier graphed on top of the test data:



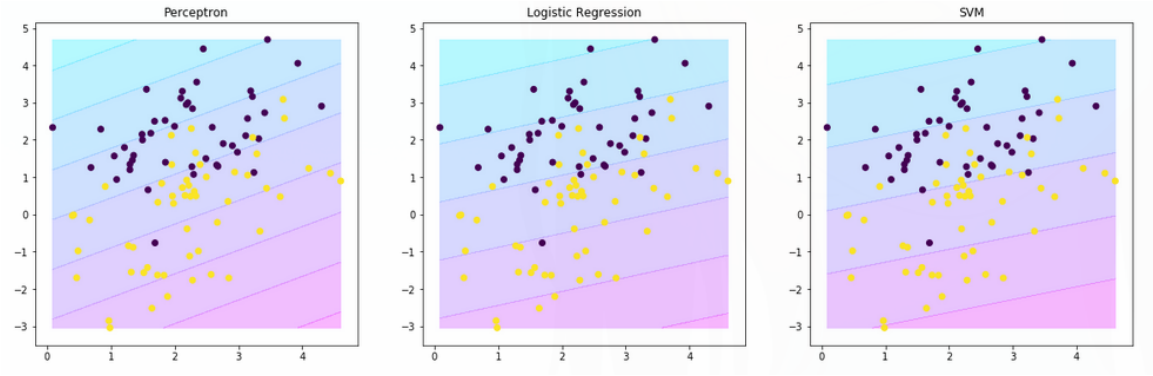
REPORT 5

The accuracy of our SVM model after 200 iterations was calculated to be .83 on both the test and training data, which was better than our Perceptron model and better than our Logistic Regression model on average. The following is the classifier graphed on top of the test data:



REPORT 6

The SVM was our most accurate model. The following is our three models graphed with decision boundaries:



After analyzing these graphs, it is evident that the SVM has the widest boundaries and the least steep slope. The perceptron has the most narrow boundaries and the steepest slope. Looking back at our reports we can see that the SVM was the most accurate model, possibly correlating the slope and wideness of the boundaries to accuracy.