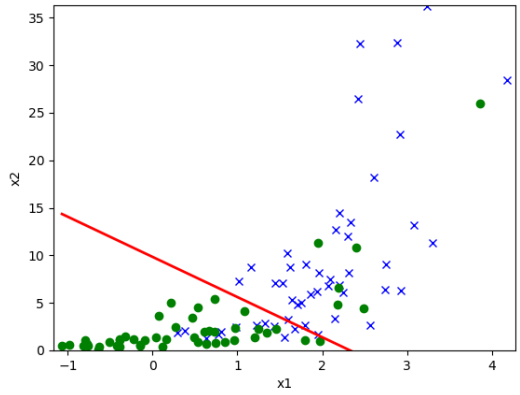
Jeremy Ng - 500882192
CPS 803 - Assignment 2

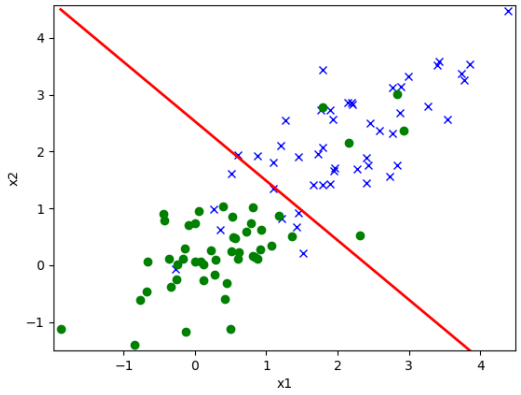
1. Logistic regression and GDA

1.1. Logistic Regression plot

Dataset 1

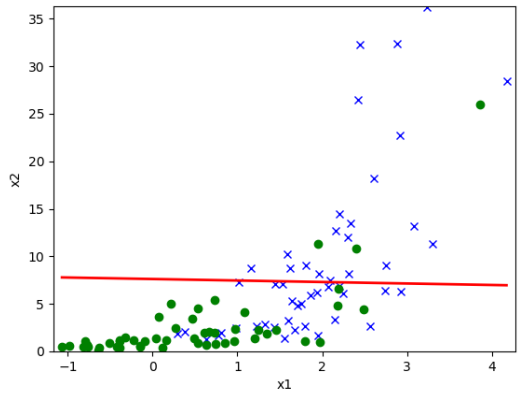


Dataset 2

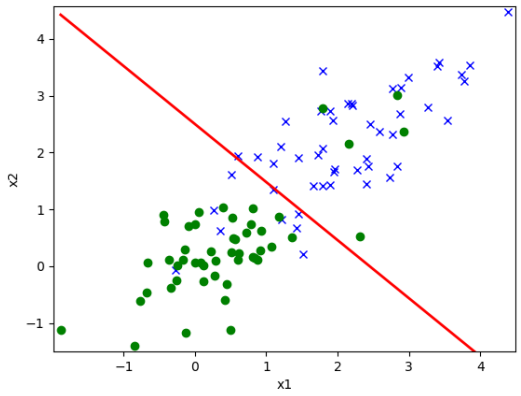


1.2. Gaussian Discriminant Analysis plot

Dataset 1



Dataset 2



Jeremy Ng - 500882192
CPS 803 - Assignment 2

1.3. For dataset 1, Logistic Regression seems to divide the data more accurately than GDA.

1.4. For dataset 2, both Logistic Regression and GDA perform well. The divider for both models seem to be identical.

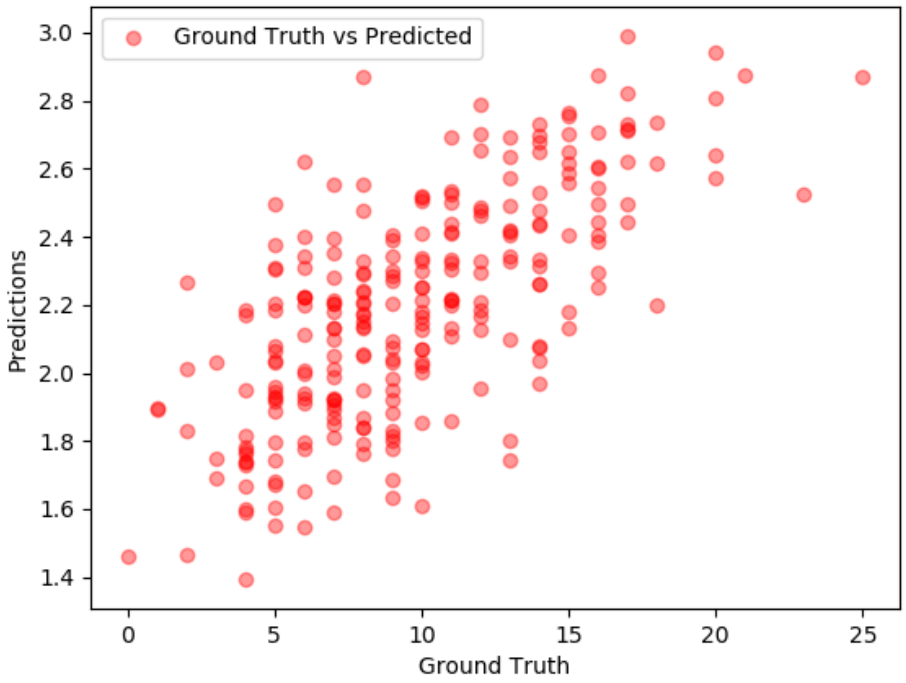
1.5. CP8318 Only Question

2. Poisson Regression

2.1. CP8318 Only Question

2.2. CP8318 Only Question

2.3. We can see from the plot that the data is following a linear trend. The gradient of our slope is close to 1.



Jeremy Ng - 500882192
CPS 803 - Assignment 2

3. Spam Classification

3.1. Size of dictionary: 1721

3.2. Naive Bayes had an accuracy of 0.978494623655914 on the testing set

3.3. The top 5 indicative words for Naive Bayes are: ['claim', 'won', 'prize', 'tone', 'urgent!']

3.4. The optimal SVM radius was 0.1
The SVM model had an accuracy of 0.9695340501792115 on the testing set