CPS803:

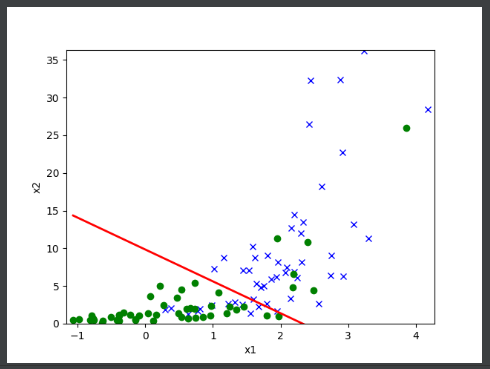
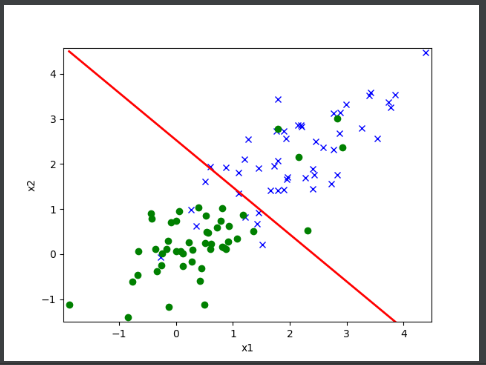
Machine Learning (Fall 2021)

Assignment 2

Name: Tusaif Azmat Student#:500660278

**Q1:**

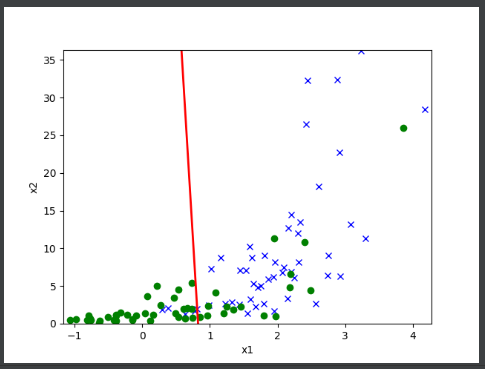
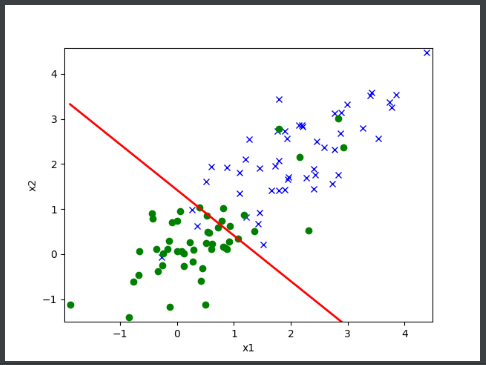
**1.1: Answer:**

Dataset.1 Dataset.2

Above are the results plots from the two datasets for Logistic Regression.

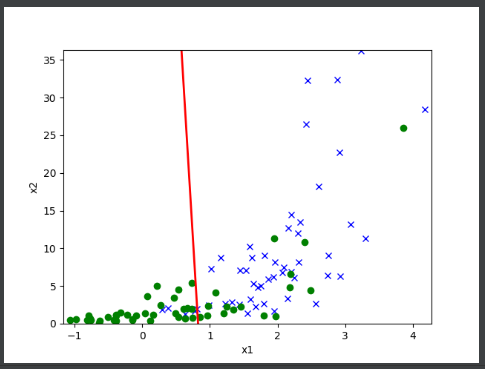
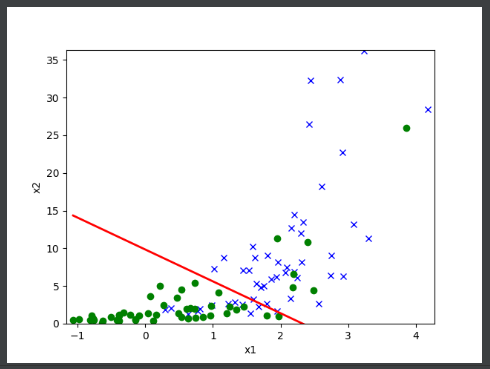
**1.2: Answer:**

Dataset.1 Dataset.2

Above are the results plots from the two datasets for GDA.

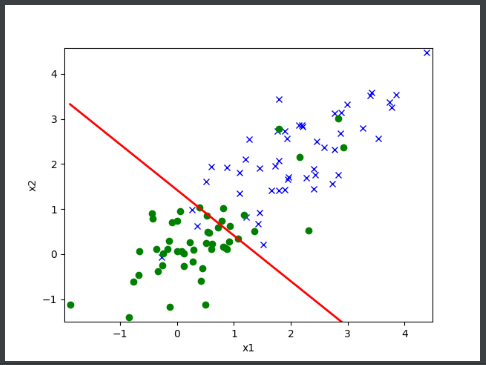
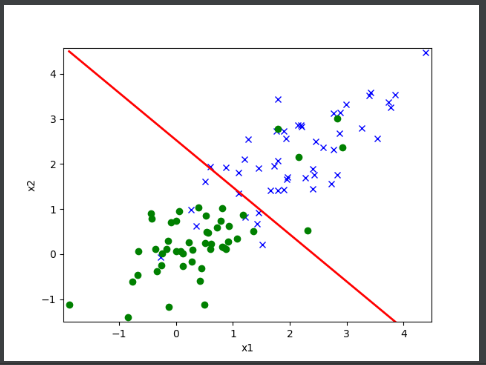
**1.3: Answer:**

****

**Logistic regression (dataset.1) GDA (datset.1)**

If we compare each plot for dataset 1 with respect with Logistic regression and GDA, it is clear that the decision boundary with logistic regression appears more reasonable than GDA on dataset 1. Hence, logistic regression has higher accuracy.

**1.4: Answer:**

****

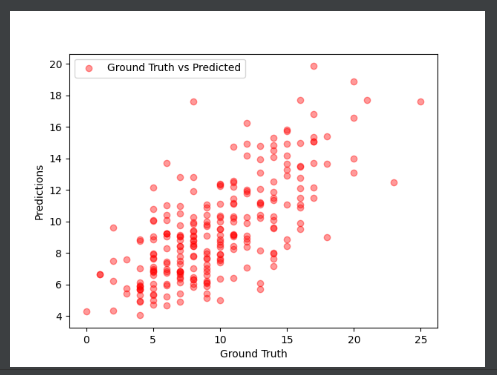
**Logistic regression (dataset.2) GDA (dataset.2)**

If we compare the validation dataset 2, both models gave rather similar results but the logistic regression model was slightly better. Overall the both learned decision boundaries are good on this data set.

As we have observed from the previous question, GDA performed worse than logistic regression on data set 1. Two classes in data set 2 don't look like Gaussian distributed.

**Q2:**

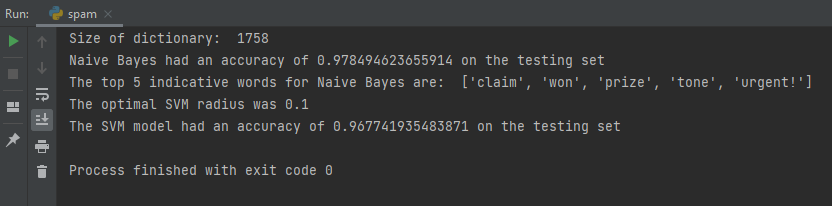
**2.3: Answer:**

****

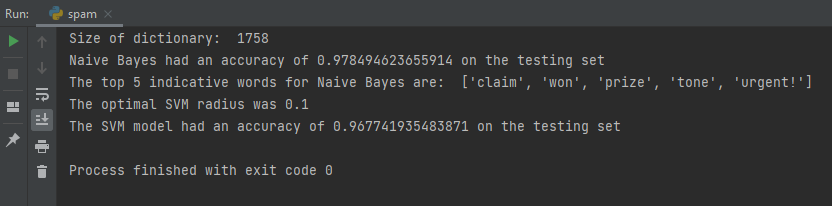
Plot: Learned Poisson GLM model on the validation set.

**Q3:**

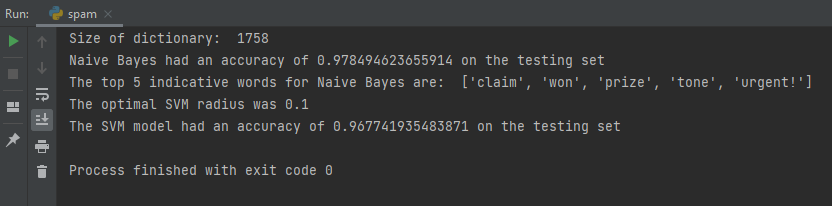
**3.1: Answer:** The output of spam.py. After preprocessing the function get words, create dictionary and transform text: we get the vocabulary size of 1758.

****

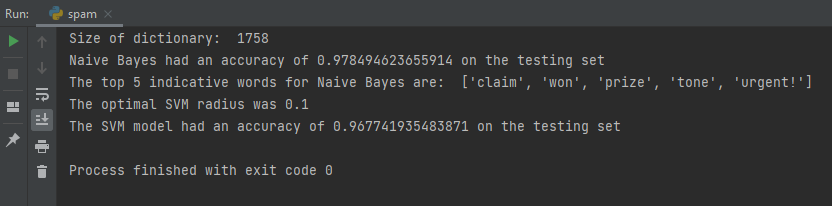
**3.2: Answer:** **Naive Bayes had an accuracy of 0.978494623655914 on the testing set.**

****

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

****

**3.4: Answer:** **The optimal SVM radius was 0.1 The SVM model had an accuracy of 0.967741935483871 on the testing set.**

****