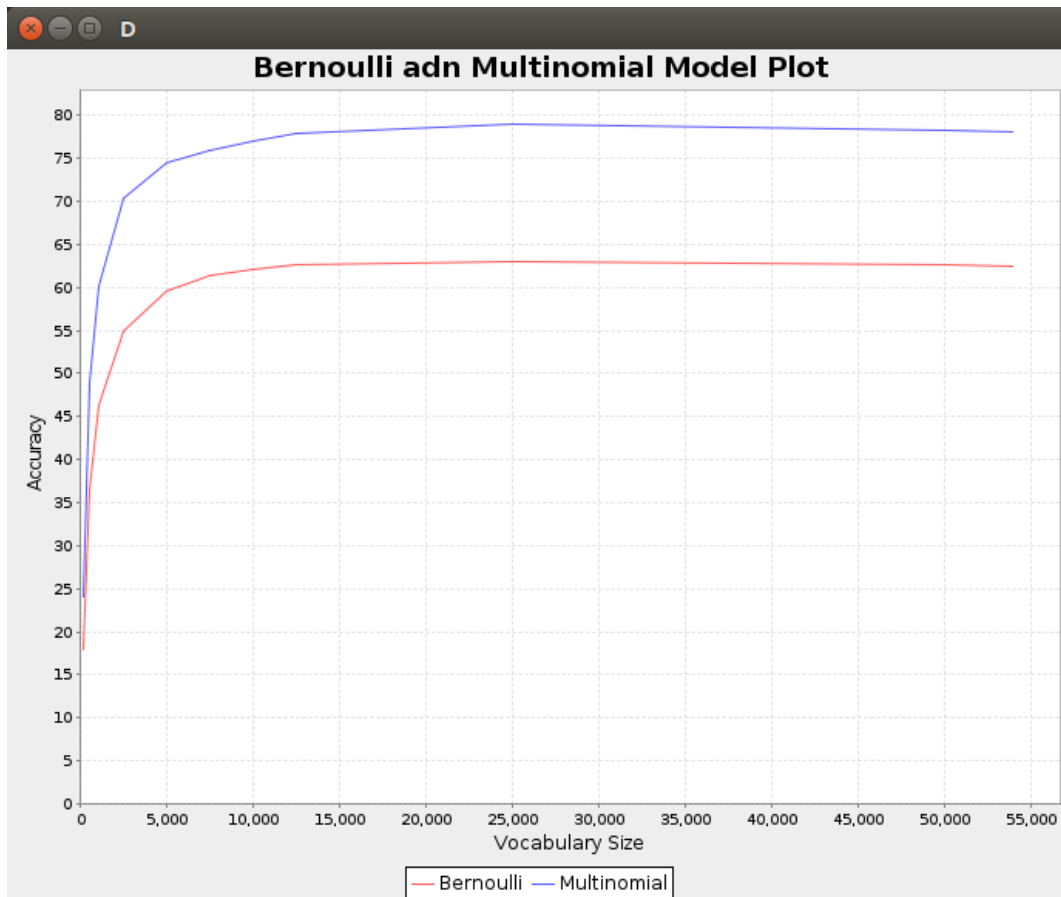


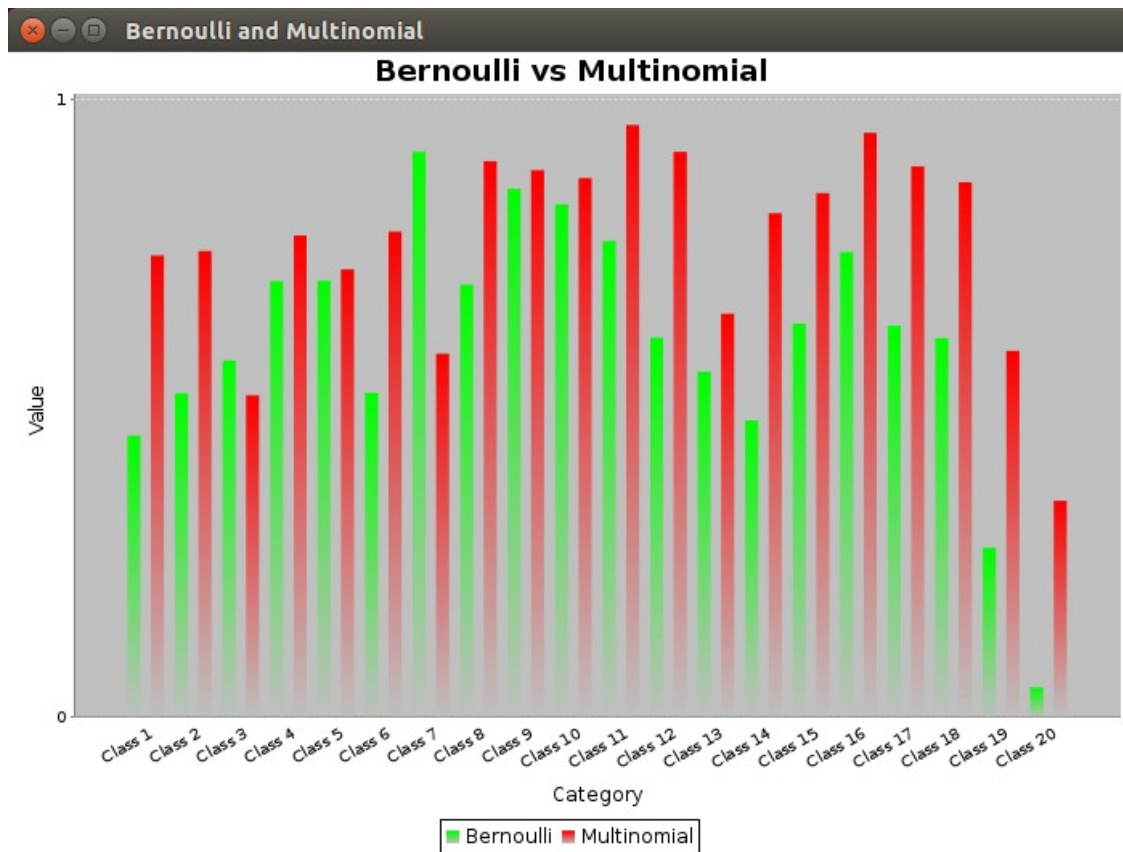
Shreyas Mahimkar

Q1)

1) Classification Accuracy Bernoulli vs Multinomial



2) Bar Plot for each Class



References:

<http://nlp.stanford.edu/IR-book/pdf/13bayes.pdf>

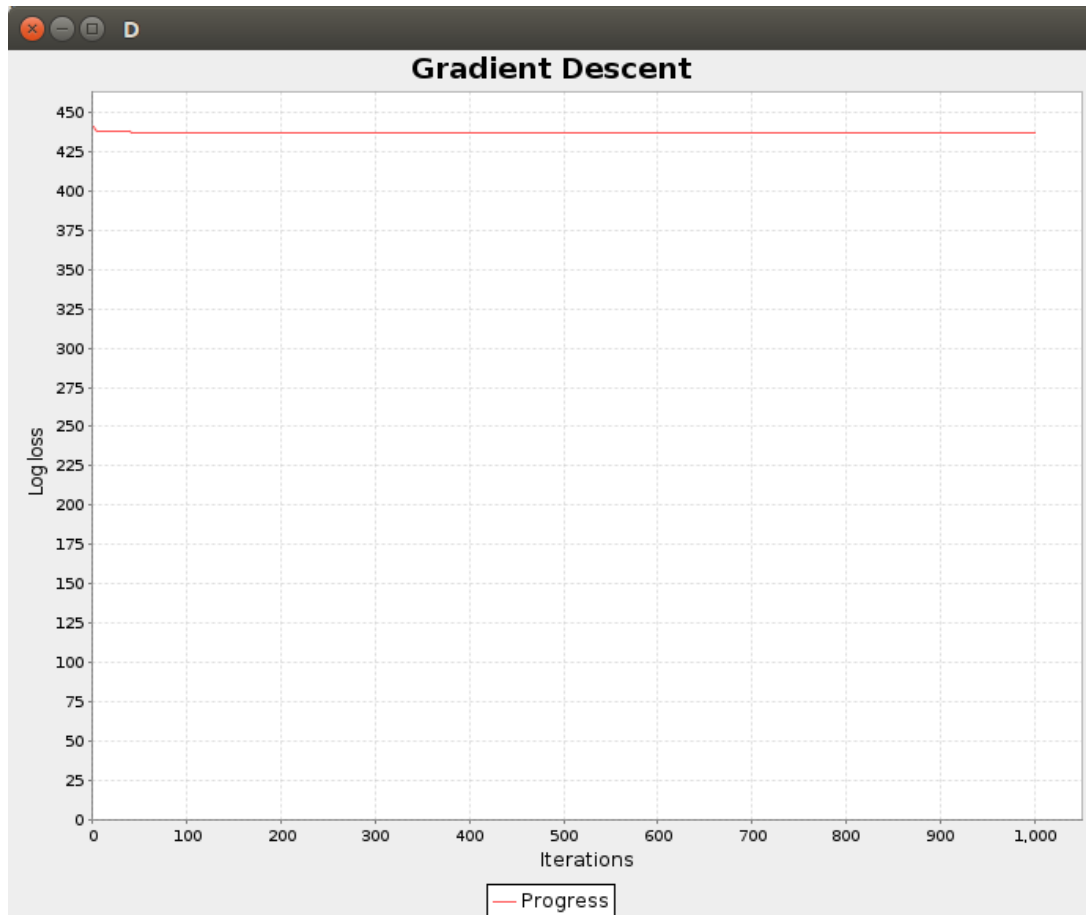
Q 2) Logistic Regression and plot to select proper tolerance and maximum iterations

-----diabetes-----

Mean : 63.23529411764707

Standard Deviation : 8.20254927974584

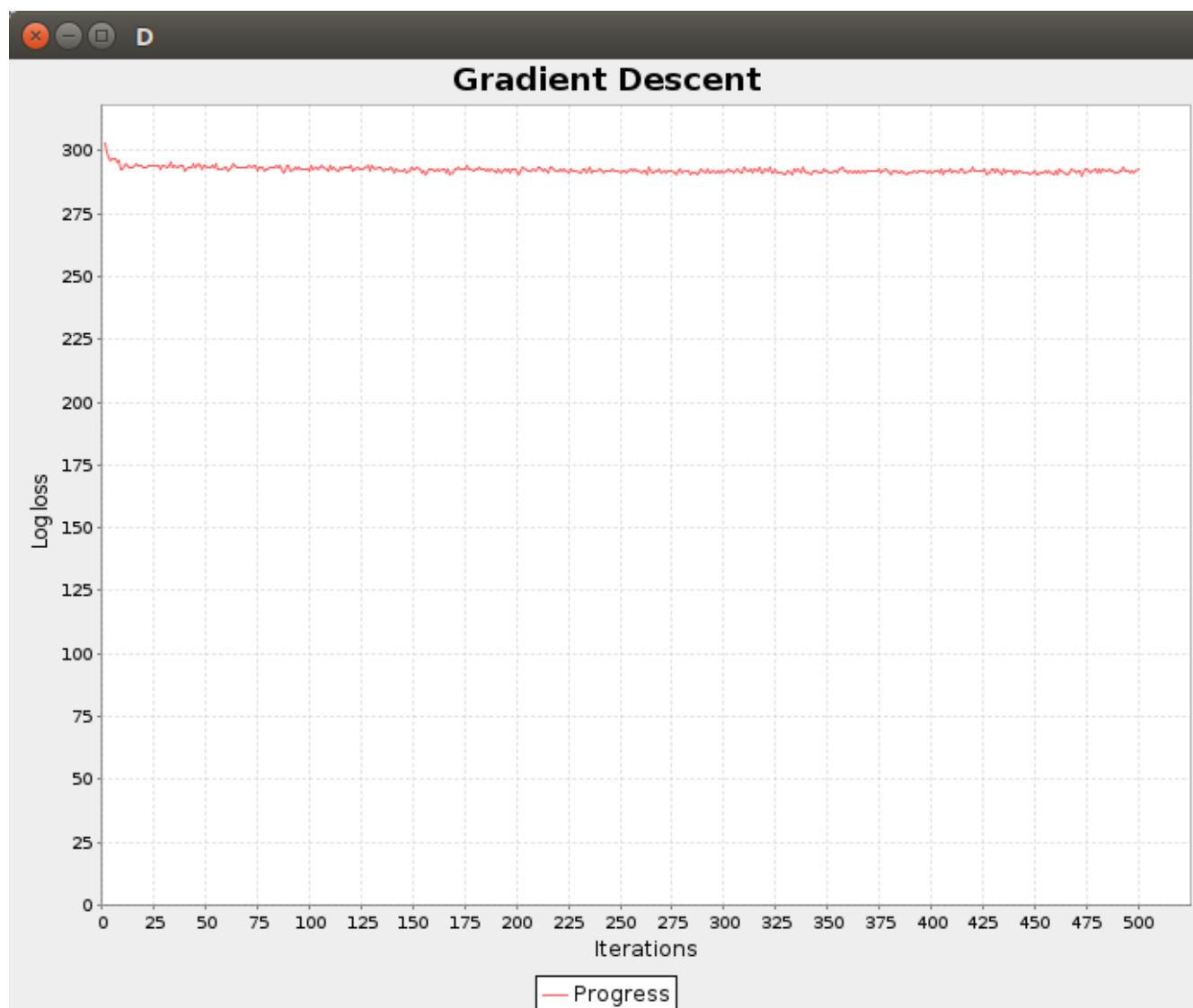
Plot:



-----breastcancer-----

Mean : 88.42105263157893

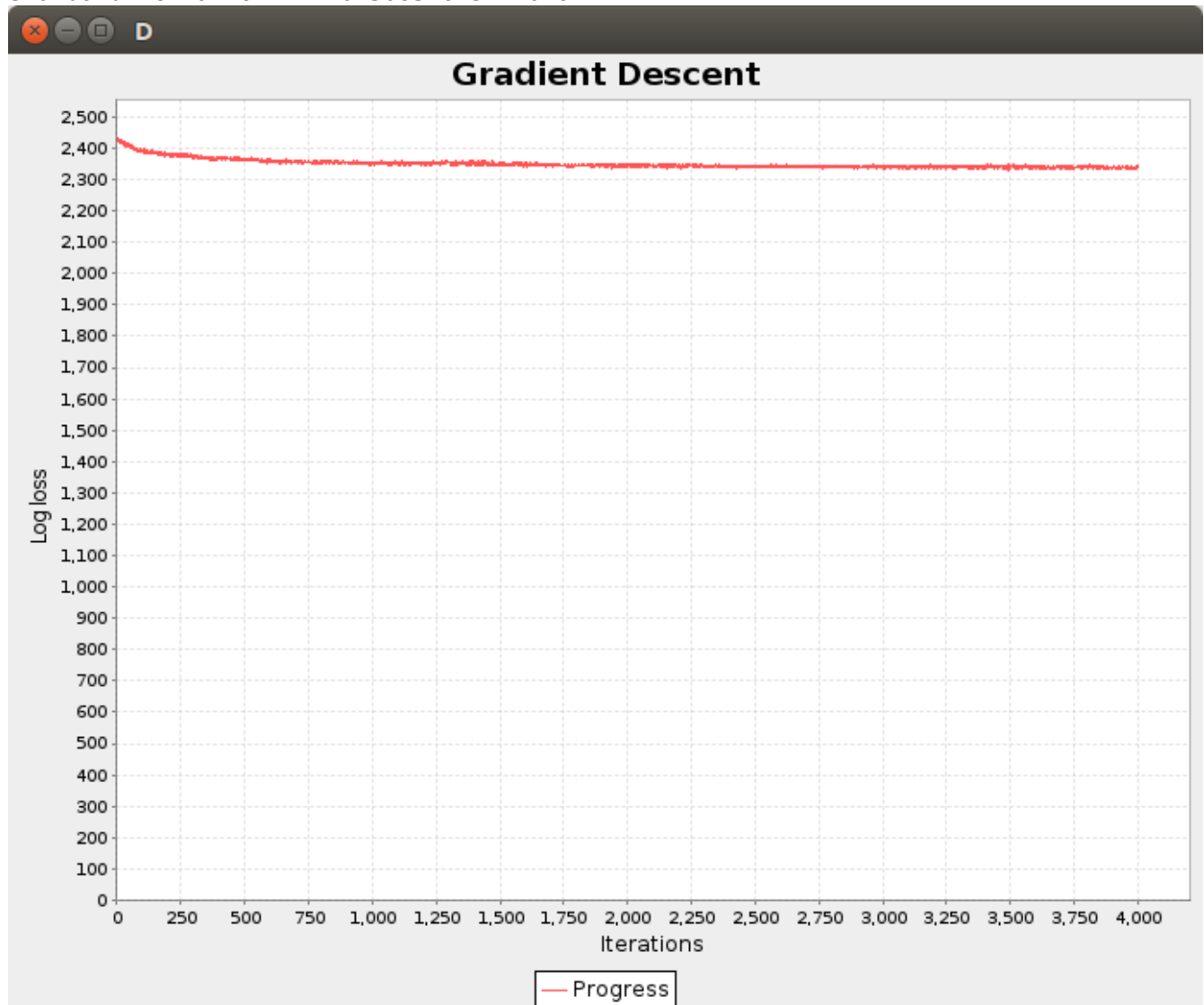
Standard Deviation : 8.602655227337213



-----spambase-----

Mean : 89.19739696312365

Standard Deviation : 2.0750854673114016



3. As you can see above the for spambase, after around 1000 iterations the value of the log loss becomes stable, So we can stop the maximum iterations to 1000 to spambase with learning rate = 0.0001.

Similar trial an error I used to determine maximum iterations and tolerance where the curve becomes straight.