**Logistic Classifier**

**Dataset:**

The dataset I chose represents a set of possible advertisements on Internet pages. The features encode the geometry of the image (if available) as well as phrases occurring in the URL, the image's URL and alt text, the anchor text, and words occurring near the anchor text.

**The task** is to predict whether an image is an advertisement ("ad") or not ("non-ad").

**Number of features**: 1558, in which 3 are continuous and others are binary.

**Feature Engineering:**

* I removed the continuous variables and worked on the remaining 1555 variables.
* 0.5% of the values were missing, Hence, substituted the missing values with the binary digit 0

**Total number of instances**: 3279

The accuracy I received by running this dataset on Winnow algorithm was 97% for training data and 96% for test data.

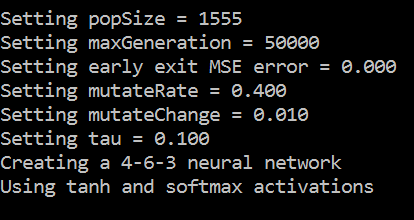
When I ran this data using logistic regression algorithm, I got the accuracy as 93% on training data and 94% for test data. Since logistic regression works best when dependent variables are categorical and my data has all the dependent variables as binary therefore this was one of the reasons for the decrease of model accuracy.

It is imperative to note that the program ran for 25minutes.

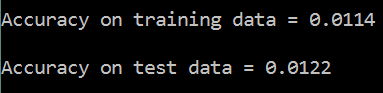
**Genetic Algorithm**

The data goes through Selection, Crossover, Mutation and Placement in Genetic Algorithm.

I ran my data on genetic algorithm with the following parameters in the program.



I got the accuracy on training data as % and on test data as %.



I tried changing the variables *maxGeneration*, *mutateRate*, *mutateChange* and *tau* but there was no increase in the accuracy. I kept the maxGeneration this high because it represents the number of timed crossover is happening and I believe it would have helped in better accuracy. Anyways I tries with a lower value as well but there was no affect in the accuracy. The program ran for 10minutes for reading all the features of all the data.

After running Winnow, Logistic and Genetic I have realized that it’s not necessary that all the algorithms will be best suitable for all types of data. For implementing a successful machine learning algorithm, we need to know the data and then apply the suitable algorithm to it.

The code and output screenshots can be found in the appropriate folders in the assignment submission.