BITS F464: MACHINE LEARNING

Done by:

Rohan Maheshwari 2017B4A70965H
Keshav Kabra 2018AAPS0527H
Godhala Meganaa 2017B3A70973H

2A: Logistic Regression

Model Description

Our model first standardises the feature values and appends a column of 1s to take care of the constant w0 in the calculations. We then define a logistic_regression class which has all the important functions.

There are 2 fit methods corresponding to gradient descent and stochastic gradient descent. These functions fit the model and return a list of costs, accuracies, iteration list and the final cost achieved by the model(Useful for plotting purposes later).

There is a predict function which takes in a set of testing values and returns their labels as predicted by our model. Finally there is a metrics function which calculates the accuracy, precision, recall and fmeasure achieved by the model on the passed predicted and true labels.

Finally we used the matplotlib library to plot 2 graphs for each model namely Costs vs Iterations and Accuracy vs Iterations. We create 10 models and display the mean of the metrics achieved by the 10 model.

The most important feature in the dataset

```
GD
Final Weights =  [[-0.14260727]
 [-1.02531131]
 [-0.59626425]
 [-0.13479226]
 [-0.023052  ]]
SGD
Final Weights =  [[-0.00381883]
 [-0.01849885]
 [-0.01129026]
 [ 0.00402135]
 [-0.00063002]]
```

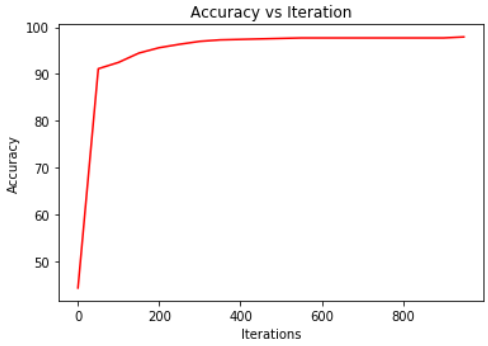
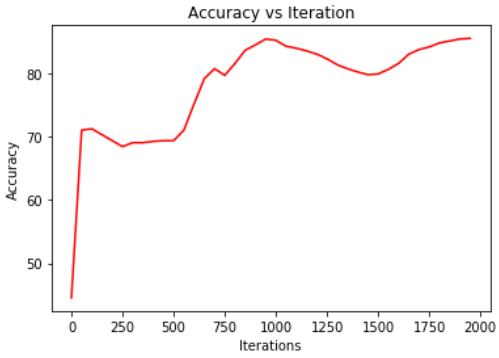
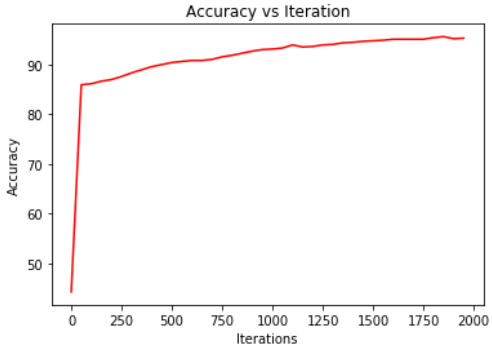
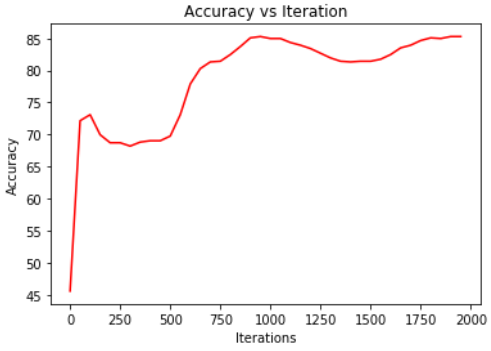
We can clearly see that in both models the maximum magnitude is of w1 hence attr1 is the most important feature in the dataset.

Metrics

Model / Measure	GD		SGD	
	Train	**Test**	**Train**	**Test**
Loss	0.08166377337945444	0.08166377337945444	0.04002991060023056	0.04002991060023056
Accuracy	97.778936392075 08	98.029197080291 98	98.884254431699 69	98.613138686131 4
Precision	0.95478956 63144832	0.95963268 00254715	0.98446271 05567991	0.97600449 74749214
Recall	0.99405066 51921865	0.99371834 26331349	0.98835011 2839937	0.98793421 56796337
Fmeasure	0.97401990 10949663	0.97634144 12309107	0.98638746 214942	0.98185126 81365602

(Learning Rate = 0.01(GD) and 0.1(SGD), Iterations = 10000)

Accuracy Plots

Alpha	GD	SGD
0.1		
0.01		
0.001