LOGISTIC REGRESSION

ASHNA SWAIKA	MUDIT CHATURVEDI	ROHAN SACHAN
(2018A7PS0027H)	(2018A7PS0248H)	(2018B3A70992H)

1)MODEL DESCRIPTION AND IMPLEMENTATION

We have implemented Logistic Regression for binary classification. The implementation has been done in Python from scratch:

1. The dataset is read using the read_csv() function of the pandas library into a pandas dataframe.
2. Next the dataset is randomized for each of the 10 iterations individually and split 70:30 into train set and test set. These sets are further divided into X, Y, X_Test, Y_Test and converted to numpy arrays.
3. Each iteration makes a call to LogisticRegressionGD(lr, iterations, X, y) and LogisticRegressionSGD(lr,iterations ,X , y) which return the weights, bias, loss and accuracy obtained for training data split corresponding to that iteration.
4. The weights, bias obtained above are used to predict the results on the training dataset. Further, calls to evaluate(y_pred , y), calc_precision(tr_pos, fal_pos), calc_recall(tr_pos, fal_neg), calc_f1(precision, recall), calc_accuracy(tr_pos, tr_neg, fal_pos, fal_neg) are made, which return the precision, recall, f1, accuracy values for the training dataset
5. Procedure similar to above is followed for the testing dataset, for both Gradient Descent and Stochastic Gradient Descent.
6. The model is retrained from scratch for 10 different 70:30 data splits and the average loss, accuracy, recall, precision and fscore are reported.

2) MOST IMPORTANT FEATURE IN THE DATASET

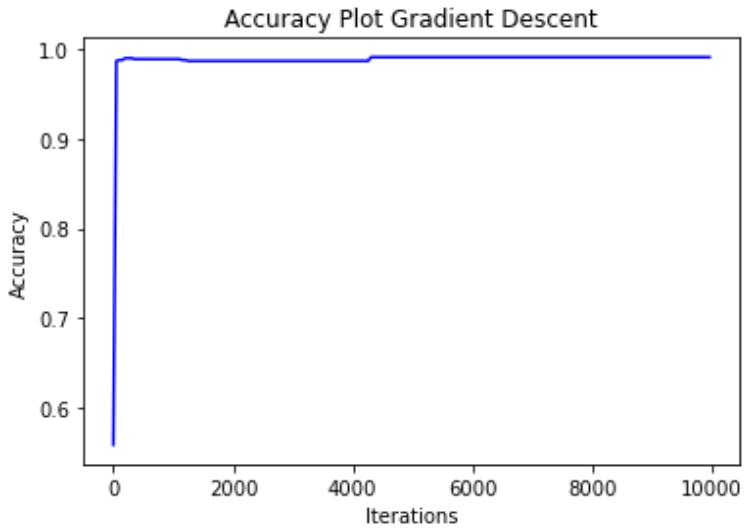
Attribute 1 was obtained to be the most important feature in the dataset, based on weight matrices' observations for 10 different data splits over Gradient Descent as well as Stochastic Gradient Descent.

3) FINAL AVERAGE TRAIN AND TEST METRICES

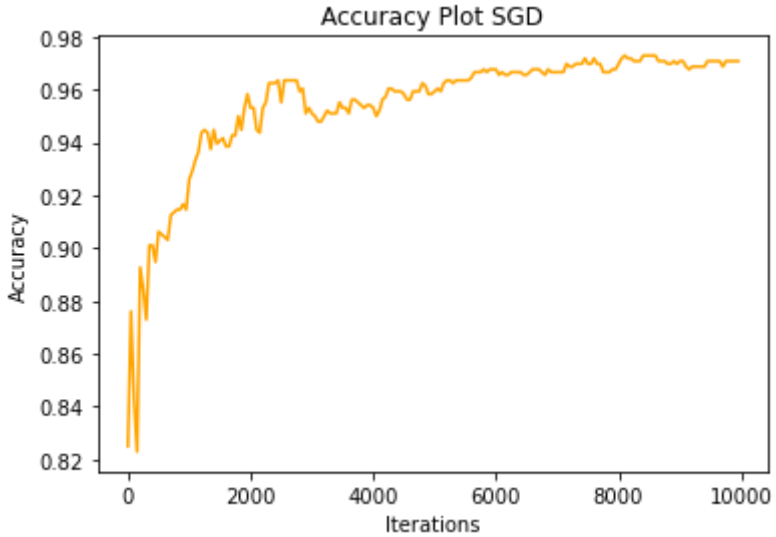
VALUE	Gradient Descent	Stochastic Gradient Descent
Train Loss	17.99320	0.042741
Train Accuracy	0.990729	0.973229
Train Recall	0.989252	0.958163
Train Precision	0.989941	0.981349
Train FScore	0.989594	0.9696101
Test Accuracy	0.991262	0.9718447
Test Recall	0.992272	0.9548644
Test Precision	0.988155	0.9807858
Test FScore	0.990179	0.9676095

4) ACCURACY PLOTS FOR Gradient Descent and Stochastic Gradient Descent with 3 different learning rates.

A. Learning rate = 0.001

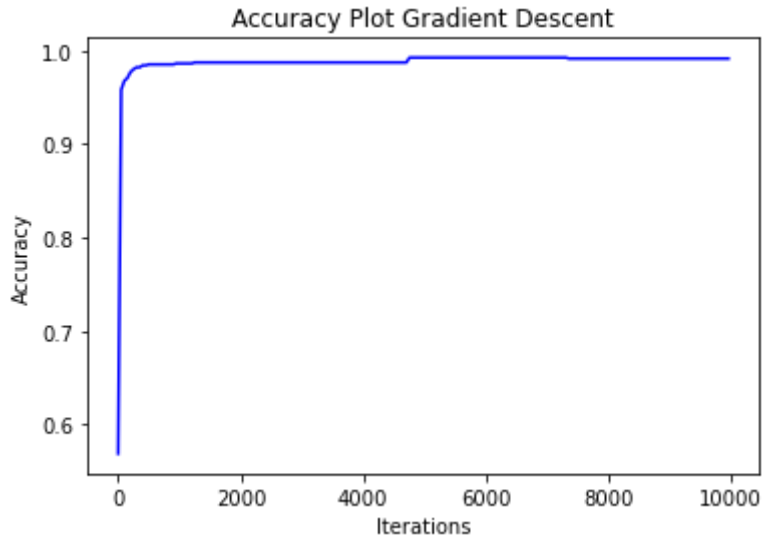


Gradient Descent Accuracy = 0.990625

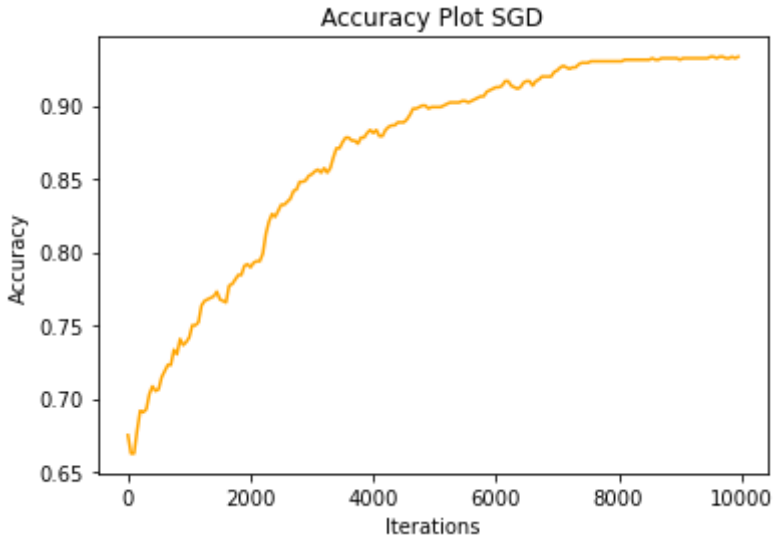


Stochastic Gradient Descent Accuracy: 0.9708333333333333

B. <u>Learning rate = 0.0001</u>

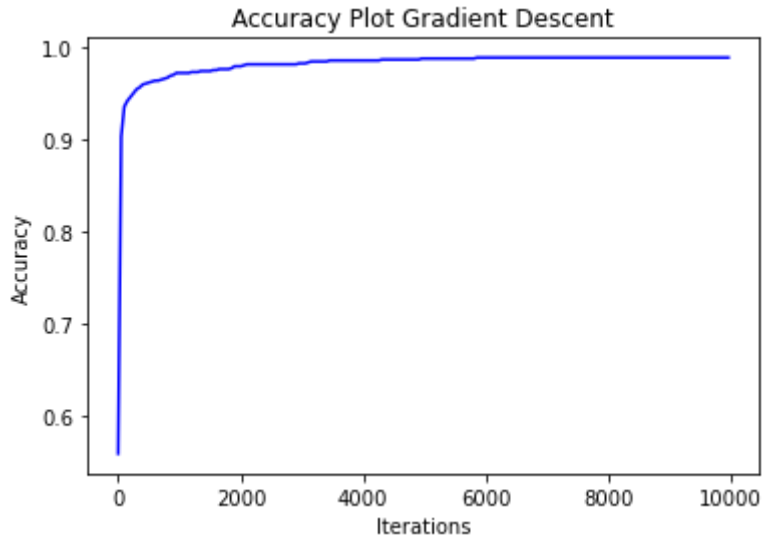


Gradient Descent Accuracy = 0.9916667

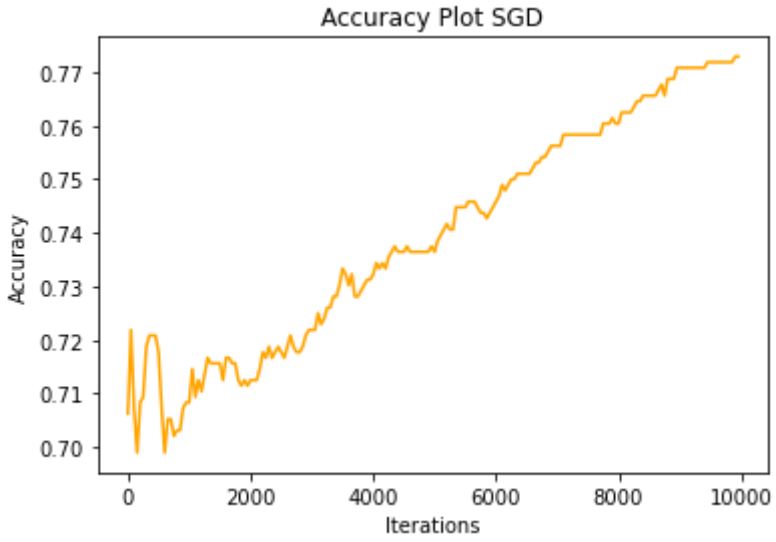


Stochastic Gradient Descent Accuracy: 0.933333

C. Learning rate = 0.00001



Gradient Descent Accuracy = 0.98854



Stochastic Gradient Descent Accuracy: 0.772916