PROJECT REPORT COMP-526 FALL 2021
NISHEE AGRAWAL

PROBLEM STATEMENT

Student admission into a university is based on several criteria. Typically the following information is analyzed in the decision process: GRE score, GPA and Class Rank. You need to obtain a model that predicts whether a student gets admitted into the university. The model uses logistic regression and is trained using historical data. The historical data is available from different sources 1. The data shows when a student was accepted (admit = 1) or not accepted (admit = 0). For model training, you will use the gradient descent method. You have to write your own code in C or FORTRAN.

In this project, you will use the pseudocode provided in this document. The procedures presented are:
MEAN X: It calculates the mean average of a vector X's components.
STD X: It calculates the standard deviation of a vector X's components.
NORMALIZE X: It normalizes the columns of a matrix X.
ADD UNITCOL2 X: Given a matrix X with N columns, it adds a new first column of 1's. The result is a new matrix X with (N+1) columns.
SIGMOID: Evaluates the sigmoid function, used for Logistic Regression.
GET COST: Calculates the value of the cost function and the gradient used in the optimization problem for the calculation of the model parameters.
GRAD DESCENT: Implements the gradient descent method to find the Logistic Regression Model parameters (Θ).
In addition, you need to create the following procedures:
PRECISION(Y ,Yˆ ,M): It calculates the precision of the model using the predicted output Yˆ and the actual Y . There are M actual output values. The precision is calculated as
Precision = TP/(TP + FP), where TP is the number of " admit = 1 " outcomes correctly predicted by the model and FP is the number
of " admit = 1 " outcomes incorrectly predicted by the model. It is assumed that the prediction is 1 when Yˆ ≥ 0.5, or 0 otherwise.
RECALL(Y ,Yˆ ,M):It calculates the recall of the model using the predicted output Yˆ and the actual Y . There are M actual output values. The recall is calculated as
Recall = TP/(TP + FN), where TP is the number of " admit = 1" outcomes correctly predicted by the model and F N is the number of " admit = 0 " outcomes incorrectly predicted by the model. It is assumed that the prediction is 0 when Y <ˆ 0.5 , or 1 otherwise.

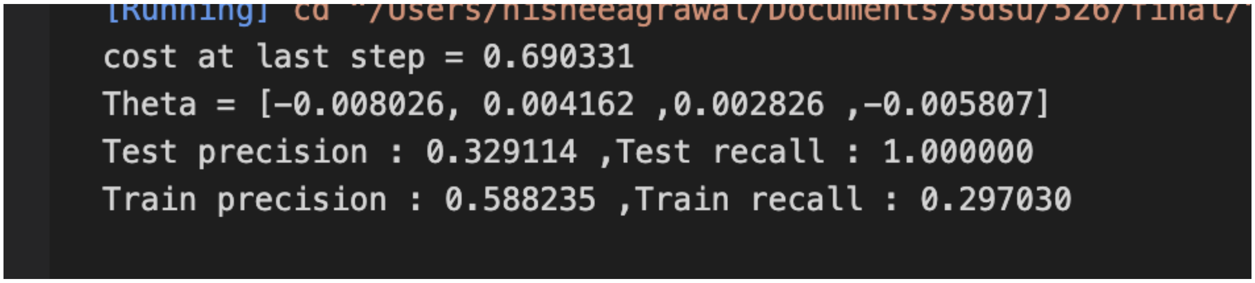
PREDICT(Θ,X) it predicts the output using the model. The product of matrix X and vector Θ
Z = X × Θ gives a vector Z. The predictions are obtained from evaluating each entry Zi as follows
Yˆi = SIGMOID(Zi)

METHODS AND PROCEDURES

I have used C programming for above mention problem statement, For creating a Predictive model I have used all the pseudo code provided and all the methods to calculate Mean, Standard Deviation, Normalize Function, to add a column function , Sigmoid function , Get COST function and Gradient DESCENT also created two more functions PRECISION which calculates the precision of the predicted model using output Y and the actual Y value and calculates the precision and then I have created Recall method to calculate the the recall of the model using the predicted output Yˆ and the actual Y .

RESULTS

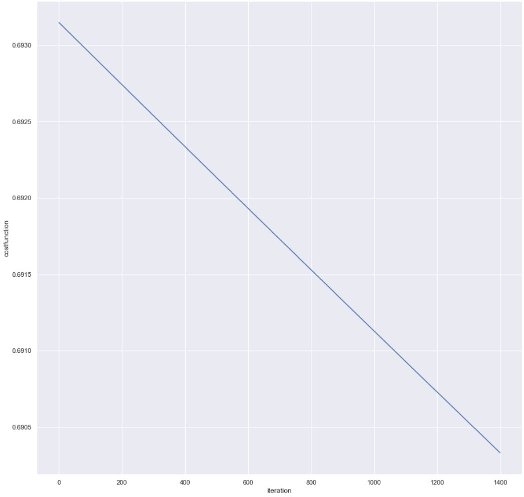
The results obtained are as follows



```
[Running] cd "/Users/nisheeagrawal/Documents/sdsu/526/final/
cost at last step = 0.690331
Theta = [-0.008026, 0.004162 ,0.002826 ,-0.005807]
Test precision : 0.329114 ,Test recall : 1.000000
Train precision : 0.588235 ,Train recall : 0.297030
```

PLOT

Plot obtained iteration vs cost function is shown below


costfunction
iteration