Analysis report
IIT2018171

Qn1A .

Our task was to build a Logistic Regression based classification model that estimates an applicant probability of getting admission to an institution based on the scores from those two examinations whose data have been attached in the folder names 'data.csv'.

1. I have Design a Predictor with two basic features which are given using Batch Gradient Descent Algorithm, Stochastic Gradient Algorithm and mini batch Gradient Descent Algorithms and calculate the error generated .

    a) Batch gradient Descent
    Without feature scaling %error is : 20.0 % .
    With feature scaling % error is : 20.0 %

    b) Mini Batch Gradient Descent
    Without feature Scaling % error is 26.666666 % .
    With feature scaling % error is : 20.0 % .

    c) Stochastic Gradient Descent
    without feature scaling % error is 20.0 %
    With feature scaling % error is : 20.0 % .

So we have seen our prediction error is maximum in case of Mini Batch Gradient Descent Without feature Scaling . Rest Methods gave almost the same error percentage of 20 % .

Qn1B . By injecting more features from the data see we get the result as follows :

a) Without feature scaling :
(i) Batch Gradient - % error = 20 %
% accuracy =80%
(ii)Stochastic Gradient - % error = 40.0 %
% accuracy = 60 %
(iii) Mini Gradient - % error 20 %
% accuracy is 80 %

b) With feature scaling :
(i) Batch Gradient - % error = 20 %
% accuracy =80%
(ii)Stochastic Gradient - % error = 20.0 %
% accuracy = 80 %
(iii) Mini Gradient - % error 20 %
% accuracy is 80 %

So we have seen w/o feature scaling the percentage error in Stochastic Gradient is maximum having 40 % . IN rest cases either by applying feature scaling or w/o applying we have accuracy of almost 80 % .

QN1C. By adding regularization term in the above approach

(a) Without feature scaling :

(i) Batch Gradient - % error = 13.33333 %
% accuracy =86.666%

(ii)Stochastic Gradient - % error = 13.3333 %
% accuracy = 86.666666 %

(iii) Mini Gradient - % error 16.6666666 %
% accuracy is 83.333333 %

c) With feature scaling :

(i) Batch Gradient - % error = 20 %
% accuracy =80%

(ii)Stochastic Gradient - % error = 20.0 %
% accuracy = 80 %

(iii) Mini Gradient - % error 20 %
% accuracy is 80 %

So we have seen by adding regularization we are able to reduce error by almost 7 %- 4 % in case of without adding feature scaling .

The minimum error achieved in case of regularization without feature scaling in Stochastic Gradient prediction .

Qn 2 .I have Design a classifier using logistic regression on Cleveland Medical data set(given as data2 in the folder) for heart disease diagnosis. We find :

The % error w/o feature scaling in case of Batch Gradient is
Error: 19.78021978021978 %
Accuracy: 80.21978021978022 %

I have also Used Confusion matrix to evaluate the performance of classifier and find :

Accuracy is : 80.21978021978022
Recall is : 69.76744186046511
True Negative rate is : 89.58333333333334
Precision is : 85.71428571428571