Assignment 4
Logistic Regression

B Chetan Rao
IIB2019018

Ques 1a

Design a Predictor with two basic features which are given using Batch Gradient Descent Algorithm, Stochastic Gradient Algorithm and mini batch Gradient Descent algorithms (determining minibatch size is your choice) with and without feature scaling and compare their performances in terms of % error in prediction.

Google colab link

Batch Gradient Descent

Without feature scaling

alpha = 0.00001, epochs = 5000

Coefficients =
[-0.0037572100258432 33,
0.011335268594940533,
-0.005446243351258 86]

Accuracy is - 82.75862068965517

Batch Gradient Descent

With feature scaling

Alpha = 0.00001, epochs = 5000

Coefficients =
[0.0007110721689603617,
0.0042330637639209 73,
0.0031647622687840 68]

Accuracy is - 93.33333333333333

Stochastic Gradient Descent

Without feature scaling

alpha = 0.001, epochs = 5000

Coefficients = [-17.009419449867295, 0.150778766483117 64, 0.16322765871133627]

Accuracy is - 89.65517241379311

Stochastic Gradient Descent

With feature scaling

Alpha = 0.001, epochs = 5000

Coefficients = [0.5799812872035947, 7.806737165995214, 6.594950012053915]

Accuracy is - 83.33333333333334

Mini Batch Gradient Descent

Without feature scaling

alpha = 0.0001, epochs = 100, batchsize = 20

Coefficients = [-0.0032175980479680713, 0.0086595530335130 24, -0.0035662510332845 97]

Accuracy is - 82.75862068965517

Mini Batch Gradient Descent

With feature scaling

Alpha = 0.0001, epochs = 100, batch size = 20

Coefficients = [1.0929605679746005e-06, 0.0033654291265052 97, 0.0028506560195192313]

Accuracy is - 86.66666666666667

Ques 1b

Inject more features from the data set in the model and repeat (a)

Google colab link

Batch Gradient Descent

Without Regularization-

Alpha = 0.1, epochs = 5

Accuracy is - 80.0

With Regularization-

Alpha = 0.00001, epochs = 5000

Accuracy is - 86.66666666666667

Stochastic Gradient Descent

Without Regularization-

Alpha = 0.001, epochs = 5000

Accuracy is - 86.66666666666667

With Regularization-

Alpha = 0.001, epochs = 5000

Accuracy is - 83.33333333333334

Mini Batch Gradient Descent

Without Regularization-

Alpha = 0.0001, epochs = 100, batch size = 20

Accuracy is - 70.0

With Regularization-

Alpha = 0.001, epochs = 5000

Accuracy is - 73.33333333333333

Ques 1c

Add regularization term and repeat (b). Submit comparative analyses of your Results.

Google colab link

Batch Gradient Descent

Without Regularization-

Alpha = 0.00001, epochs = 1000

Accuracy is - 93.10344827586206

With Regularization-

Alpha = 0.0001, epochs = 5000, lambda = 1000

Accuracy is - 86.66666666666667

Stochastic Gradient Descent

Without Regularization-

Alpha = 0.0001, epochs = 5000

Accuracy is - 86.20689655172413

With Regularization-

Alpha = 0.001, epochs = 500, lambda = 1000

Accuracy is - 82.75862068965517

Mini Batch Gradient Descent

Without Regularization-

Alpha = 0.0001, epochs = 100, batch size = 32

Accuracy is - 75.86206896551724

With Regularization-

Alpha = 0.0001, epochs = 1000, batch size = 32, lambda = 1000

Accuracy is - 93.10344827586206

Ques 2

After gaining experience of solving problem No 1) Design a classifier using logistic regression on Cleveland Medical data set for heart disease diagnosis. The processed dataset with some 13 features has been given with a label that a patient has a heart disease (1) or not (0). This design should have a professional touch within your ML knowledge in terms of data preprocessing, feature scaling, selecting appropriate features etc.

Google colab link

Mini Batch Gradient descent

With feature scaling and regularized data

Alpha = 0.00001, epochs = 500, batch size = 64, lambda = 10

Accuracy is - 76.92307692307693

Confusion matrix -

[[26, 4], [17, 44]]