**CSE 574:** Introduction to Machine Learning

Fall 2018

Submitted by: Sai Kalyan Katta (UB Person Number: 50292522)

**Overview:**

Using the datasets Human\_Observed\_Features\_data, same\_pairs, diff\_pairs I have created two datasets namely HO\_concatinated and HO\_subtracted\_1.

Using the datasets GSC\_Features\_data, same\_pairs, diff\_pairs I have created two datasets namely GSC\_concatinated and GSC\_subtracted\_1.

After taking the data, I divided it into training, validation and the testing sets where the training set contains of 80% of the data, validation set contains 10% of the data and the rest is used as the testing set. Here, the functions ‘**GenerateTrainingTarget**’, ‘**GenerateTrainingDataMatrix**’, ‘**GenerateValData**’, ‘**GenerateValTargetVector**’ divide the original data into the training sets and validation sets.

**Linear Regression:**

The function ‘**Linear\_regression**’ performs the linear regression on the data passed to the function in which the closed form and the gradiant descent solutions are found and the E\_rms values are printed.

The ‘**GenerateBigSigma**’, ‘**GetScalar**’, ‘**GetRadialBasisOut’** are used to calculate the Gaussian basis function using the input values and their means. The ‘**GenerateBigSigma**’ function calculates the part of the Gaussian radial basis functions in which the input values x and their mean µ are involved. The ‘**GetScalar**’ function calculates the whole scalar part except the exponent of the Gaussian radial basis function Ø(x). The ‘**GetRadialBasisOut’** function gives the final value of the Guassian radial basis function. The ‘**GetPhiMatrix**’ gives the matrix of the values of the basis functions.

The **‘GetWeightsClosedForm’** function returns the weights by using the formula

WML = (λI+ØT Ø)-1 ØTt

The ‘**GetValTest’** function gives thevalues of the linear regression function y(x,w) which is defined as

y(x,w) = wT Ø(x)

The ‘**GetErms**’ function gives the Root Mean Square error between the test output and the actual data.

**Logistic Regression:**

The function ‘**Logistic\_regression**’ performs the logistic regression on the data passed to the function in which the accuracies are printed at last. First, we calculate the value of xθT by taking random value for θ and then calculate the target values A by applying the sigmoid function on xθT which is defined as

Sigmoid(a) = 1/ (1 + e-a)

done by the function ‘**Sigmoid**’.

Then, the gradiant is found calling the ‘**gradiant**’ function which is used to calculate θ\_next. We calculate the accuracy of our findings of target values A by comparing it with the original Target values.

The ‘**gradiant**’ function gives the calculation of gradiant which is used in calculating the theta\_next values.

gradiant = xT(A-y)

Where A is the predicted target values

y is the actual target values

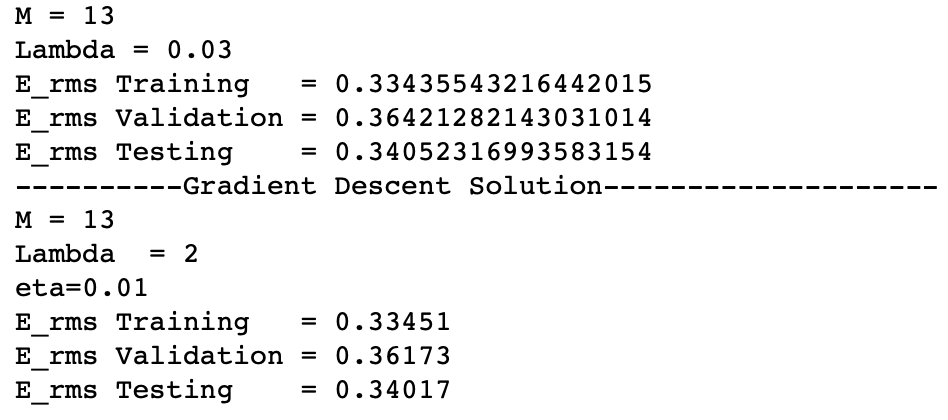
x is the data

The ‘**GetAcc**’ function gives the accuracy of the findings.

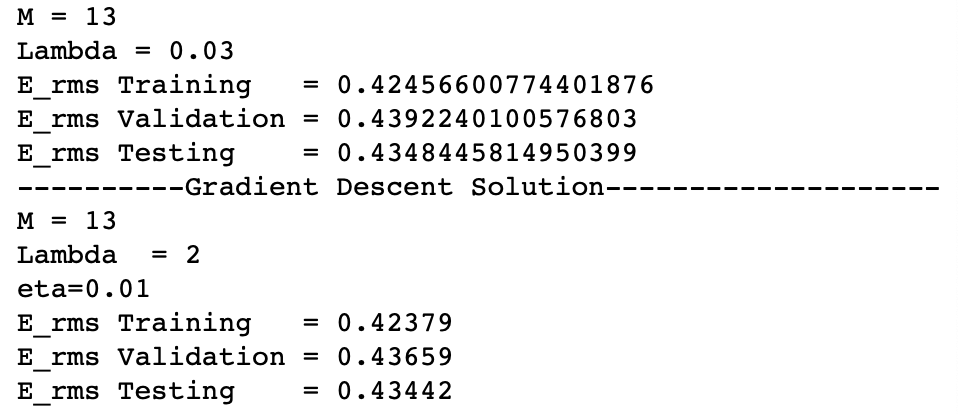
**Results:**

**Linear Regression:**

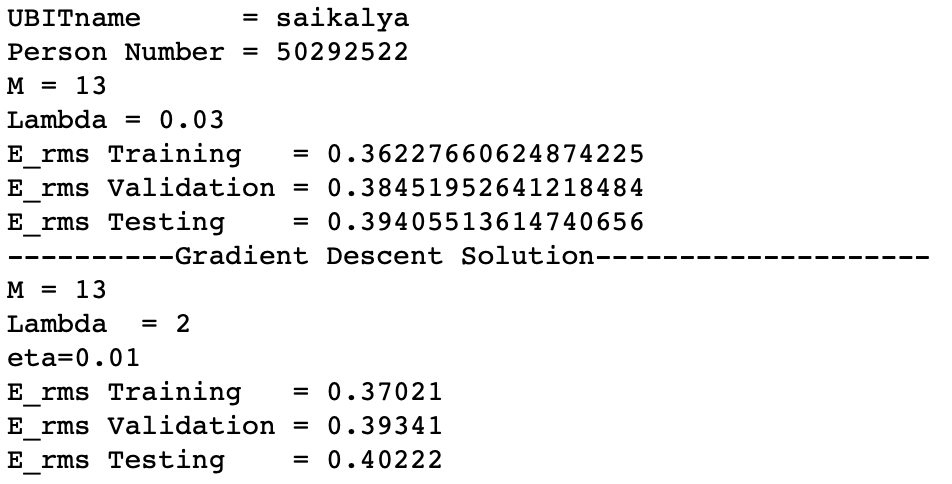
Results for Human\_Observed\_features\_concatinated:



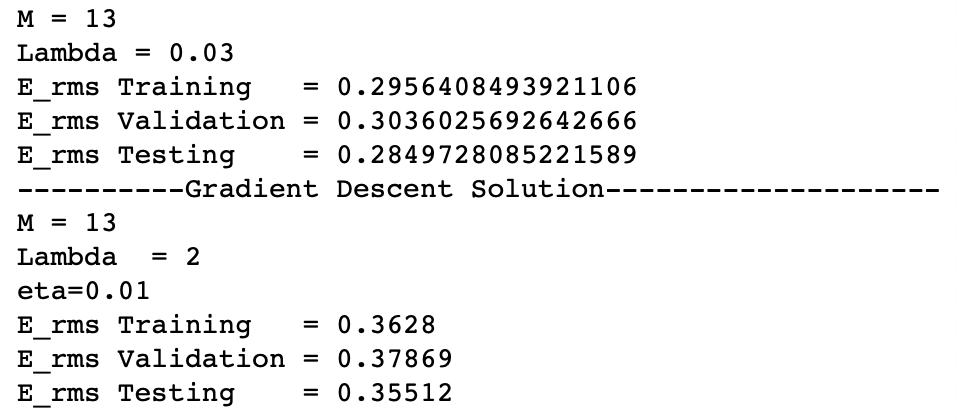
Results for Human\_Observed\_features\_subtracted:



Results for GSC\_features\_concatinated:

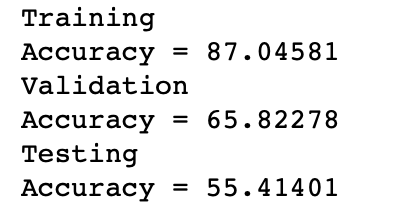


Results for GSC\_features\_subtracted:

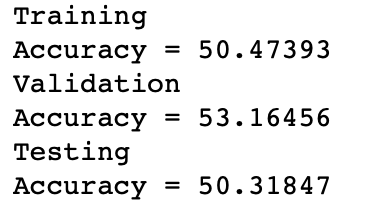


**Logistic Regression:**

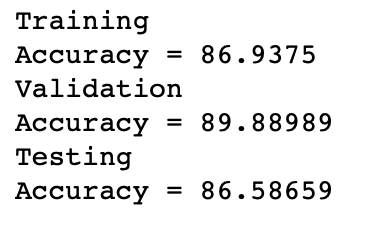
Training, Validation and Testing accuracies for Human\_Observed\_features\_concatinated:



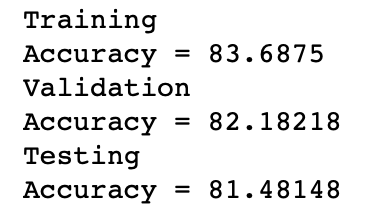
Training, Validation and Testing accuracies for Human\_Observed\_features\_subtracted:



Training, Validation and Testing accuracies for GSC\_features\_concatinated:



Training, Validation and Testing accuracies for GSC\_features\_subtracted:



**Conclusion:**

From the results obtained, it is seen that the linear regression produces the E\_rms that are less than 0.5 for all the datasets, which is a good sign of performance. Whereas, the logistic regression produces the accuracies ranging from 50 to 90 for all the datasets. The logistic regression performed well for the GSC\_Features datasets but under-performed for the Human\_Observed\_Features datasets. It can be said that the linear regression is consistently performing well with all the datasets.