Data Project 8

Sree Teja Kongara,

1. List the output of your Neural Network analysis and describe the results in no less than 5 sentences.

**Ans:** The model accuracy shown on running the Neural Network analysis on the select attributes of the dataset, returns an accuracy of 79.48%. This is after running the model initially on a training set of 67% split of the dataset (having considered only 1000 tuples out of the total 3220, for ease of processing) and then running the model on a test data set of the remaining 34% data. This needs to be further tuned as the initial weights have been randomized.

1. Tune your Neural Network such that your prediction results approach the quality of your Random Forest run. Include your code and explain in no less than 5 sentences how you adjusted the hidden layers, and what your adjustments have changed about how the algorithm works.

**Ans:** On adjusting the hidden layers multiple times (from hidden columns (12,11)) we have arrived at the conclusion that changing hidden layers to (16,12) would produce an accuracy (47.25%) which is closer to the random forest model’s accuracy obtained earlier (44.80%). This means that the weights assigned to the algorithm have been changed, leading to more accurate values that correlate to those obtained while running the random forest with the same variables.

1. Then look at your runs with Multiple and Logistic Regression. Write a Gradient Descent routine to see if you can improve on the quality of one of them. Include your code and your visualization(s) and explain in no less than 5 sentences how using Gradient Descent has improved your regression formula. If it hasn’t, explain why not.

**Ans:** From the graph produced via the gradient descent model, we can see that the regression formula has only slightly changed. We believe this is because of the accuracy of our prediction, which correlates to the values obtained from both the random forest model and also the neural network model.