STA9792 - Homework 4

Christopher Lang

November 29, 2017

Please see the Excel Spreadsheet called **HW04 Submit - Christopher Lang.xlsx**, in the worksheet tab **PREDICTION**, highlighted in yellow, my answer for this homework

Please see the source code file **STA9792 - Homework 4 code.R** for the R code used to generate those classification prediction

In regards to the question:

How could we get better results for both the in-sample and out-sample (output predictions)?

For improved in-sample, we can continue to tune the Neural Network model further, using different learning algorithms, more hidden layers, etc. It very well could be that the greater complexity introduced by more hidden layers can figure out the nuances needed for scoreboard number recognition

For better out-sample prediction accuracy, one of the methods you could use is a monte-carlo k-fold cross validation technique (or other optimization algorithms, such as metaheuristics e.g. genetic algorithm, particle swarm, evolutionary strategy, etc.)

The goal of using such techniques is to figure out the model's *stability*. Currently, the script only captures accuracy, but it is tough to really predict the model's out-of-sample performance without some measure of variation on new data(this is similar to looking at a population's mean without the accompanying variance or standard deviation). This way, we can get a sense about how well the model (as in the neural network's hyperparameters) would do when new data is introduced

Upon seeing such variation, it would be up to the developer, and the use case of the model, to determine the optimal model