IST 707 — Applied Machine Learning

**HW8: Deep Learning with Neural Network Algorithms**

Having focused the bulk of the class on the more traditional data mining algorithms, this HW will focus a newer class of called Neural Network algorithms.

While there are many different variations on ways that (N)eural (N)et’s work, for this HW we are going to focus on the NN variation we reviewed in the NN lecture, Multilayer Perceptrons.

Your Assignment

1. Load the drug data from Blackboard, make sure that the Drug attribute is set to target role.
2. Use the “Constant” module to predict the drugs and test the model. This is your baseline. Record the F1-score using 5-fold cross validation

With 5-fold cross validation, I got baseline F1 score of 0.285 for the constant module.

1. Use the “Tree” module to do the same and test the model. Record the F1-score.

The F1 Score for tree was 0.990.

1. Then, use the “Neural Network” module to do the same and test the model (via 5-fold cross validation) under the following conditions:

The activation method for all the following setting was tanh. I used L-BFGS-B solver. Alpha was 0.000, and maximal number of iterations was 30.

* 1. 25 neurons in hidden layers

With 100 neurons in hidden layer, I got an F1 score of 0.955.

* 1. 50 neurons in hidden layers

With 100 neurons in hidden layer, I got an F1 score of 0.966.

* 1. 100 neurons in hidden layers

With 100 neurons in hidden layer, I got an F1 score of 0.98.

* 1. 200 neurons in hidden layers

With 100 neurons in hidden layer, I got an F1 score of 0.97.

1. Record the F1 value for the NN under each condition.
2. Answer the following questions:
   1. Which configuration/model is most accurate?

It seems that tree is the most accurate, with an F1 score of 0.99.

* 1. Which model is the fastest?

Generally, the speed of each model is dependent on the tuning and the final setting that we use. In this case, decision performance was way faster than the neural network with 50, 100, and 200 neurons in hidden layers.

* 1. Describe which configuration is the best balancing accuracy vs. speed.

I was under impression that the more hidden layers we use, the better accuracy we will get. However, in this experiment, I realized that increasing the number of hidden layers will not necessarily increase the accuracy of the model (because the model will be overfitted). Moreover, Increasing the number of hidden layers will significantly decrease the model speed. In conclusion, the decision tree and the neural network with 100 neurons in hidden layers have led to the best F1 scores respectively. Thus, these models would be the best choices for prediction purposes.