Using A Multi-Perception Classifier to Predict the Likelihood of Obtaining Heart Disease

Joshua Stallings

December 2020

1 Introduction

In order to see if there is any correlation between various cardiovascular and respiratory conditions and the likelihood of an individual being diagnosed with heart disease, a neural network was trained on a database and fed new information to predict a result. A data set including various factors, such as heart rate, blood sugar, age, and sex were standardized and used to train a classifier. Then, the classifier was scored on data it had not seen before and the accuracy was record.

2 Data Acquisition

A data set made up of data from several experiments on people who had and did not have heart disease was acquired from
`https://www.kaggle.com/nareshbhat/health-care-data-set-on-heart-attack-possibility`

While each individual experiment had up to 76 different attributes, some of the sets were incomplete or missing data. So, data that overlapped from each experiment was taken, and then condensed to limit the amount of data that had to handled at once. The age, sex, type of chest pain, blood pressure, amount of serum cholesterol, blood sugar, maximum heart rate, and exercise-induced angina were selected to input into the neural network.

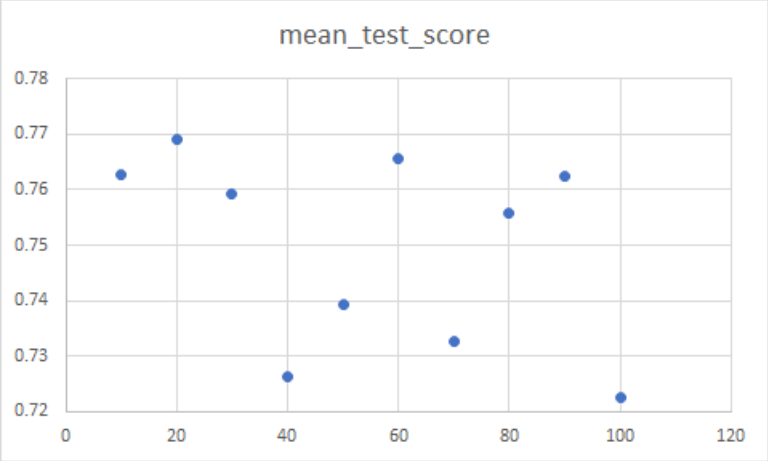
The data is already numerical, so the only modification the data needs is to be standardized / normalized.

An example data point fed into the neural network would be:
(63, 1(male), 3, 145, 233, 1, 150, 0(no agina)

3 Optimizing the Neural Network

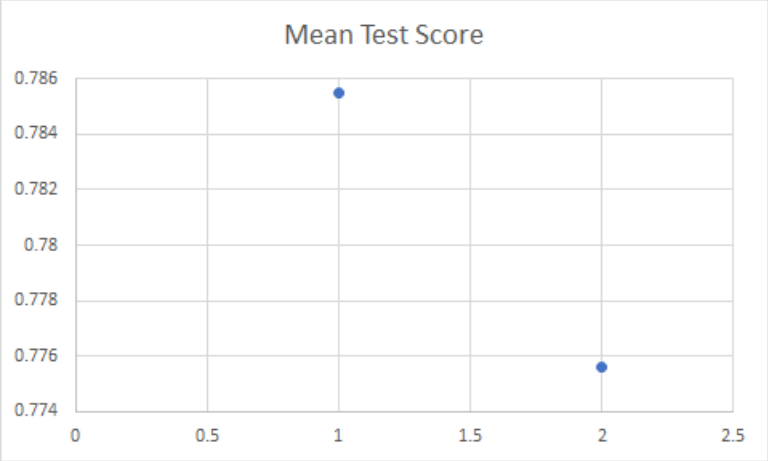
In order to get the best results, various settings had to be tested on the neural network.

The first thing that was changed was the number of hidden layers, and whether or not a two dimensional one would work better than a one dimensional one.



Since the data is scattered with seeming little relationship, it is best to go with whatever layer size gave the highest accuracy. So, a hidden layer size of 10 will be used.

Next is the type of solver used by the classifier. There are three types, 'Ibfgs', 'sgd' , and 'Adam'.



Overall, there was not much difference between the different solvers. 0 = Ibfgs, 1 = Adam, 2 = Sgd. Since Adam was the highest, but not by much, it

was used in the classifier.

4 NN Results

Once the above settings were determined, they were inputted to the neural network. Then, the network was cross-validated. Half of the data from the data set was used to train it, then the other half was put in as new data for the network to output a result for.

The two accuracy results reported were 0.80921053 and 0.7615894, or 0.81 and 0.76 when rounded. The rounded average was 0.79.

5 Conclusion

Overall, the neural network performed well. The accurate results most likely mean that there is some correlation between the inputs given, the likelihood of an individual having heart disease. The sample size could be increased, and the number of inputs could be decreased to narrow down what the tell-tale signs of heart disease may be, or whether or not it is all these symptoms combined. Furthermore, the output only decided if someone was more likely or less likely to have heart disease. Research could be done with a more specific output, such as predicting the severity of the heart disease and so on.