Statistical Learning Midterm

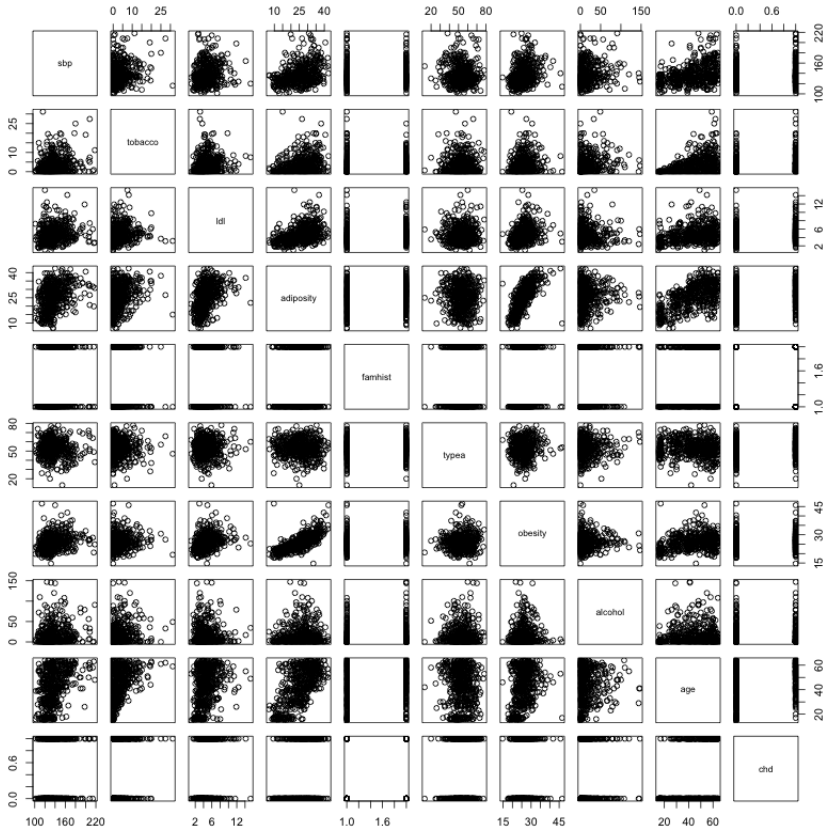
Duncan S. Wilson

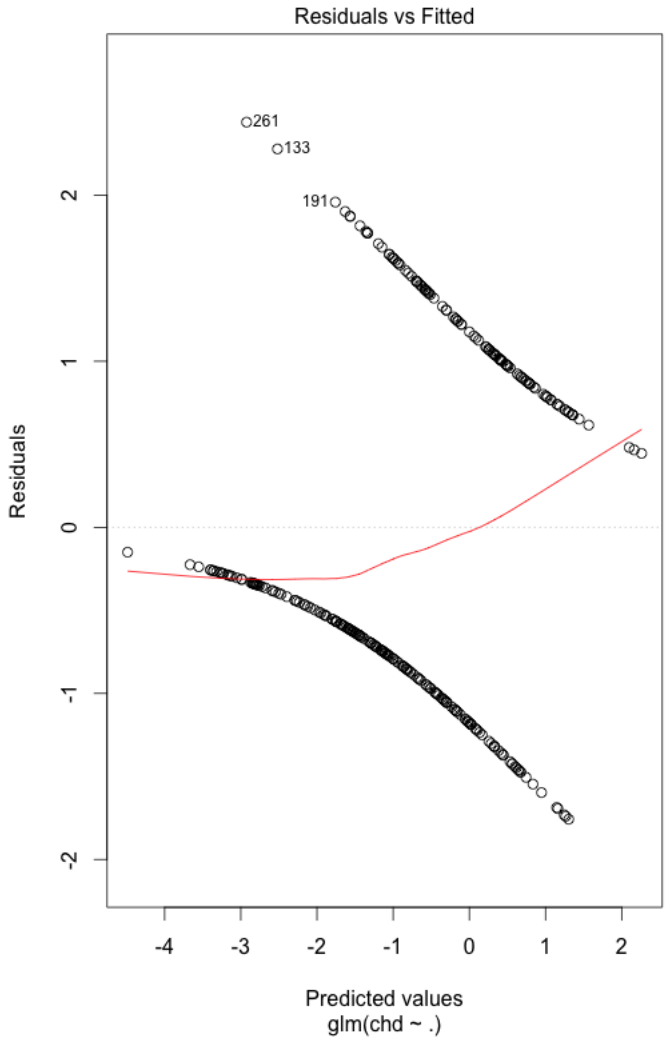
March 25, 2015

3 Logistic Regression Vs Neural Network

3.1 Logistic Regression:

Below is a scatter plot of the ten variables in the South African Heart Disease Data set:





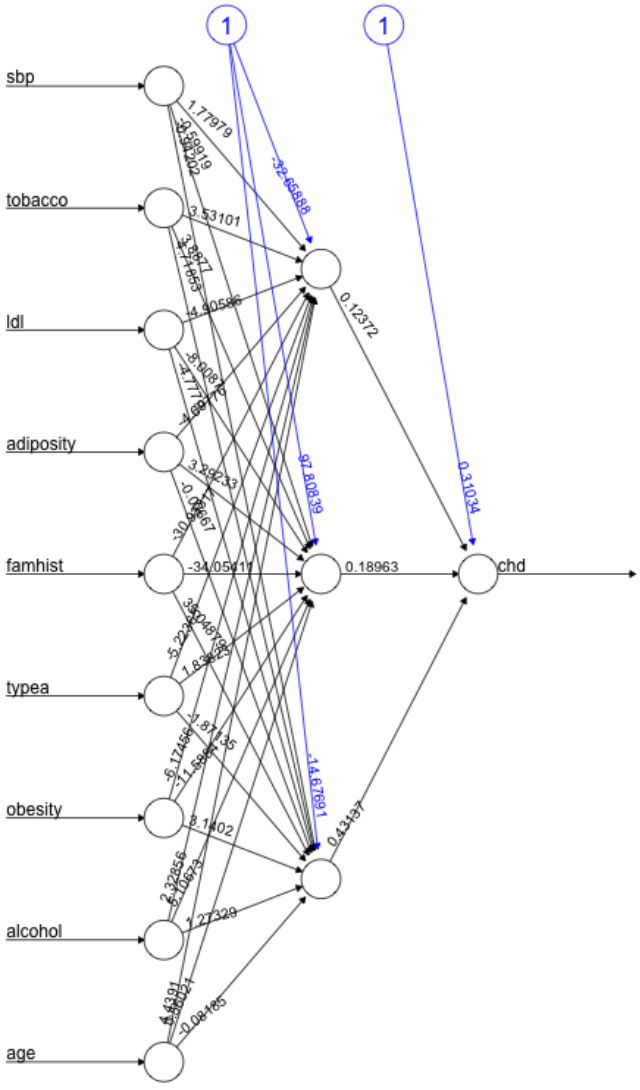
Here the glm(chd .) represents the built-in logistic model fitted to the chd variable in the South African Heart Disease data-set. As the table below shows, neither my classifier nor the built in was perfect. My code is attached.

chd variable	Frequency	My Logistic Regression	Built-In Function
1	47	26.00	39.00
0	106	90.00	68.00

Table 1: Logistic Regression on the South African Cancer Data

3.2 Neural Network with Three Hidden Neurons:

Due to time constraints as well as not having a good result for my neural network homework (pending asking you for advice Dr. Rojas) I was only able to use R's built in library neural net to compare the performance. One really neat feature is that for smaller networks it will plot the network as a graph:



Error: 32.810628 Steps: 34131

The results for this neural net was awful! I played with all sorts of different parameters and couldn't improve my output better than what is shown below:

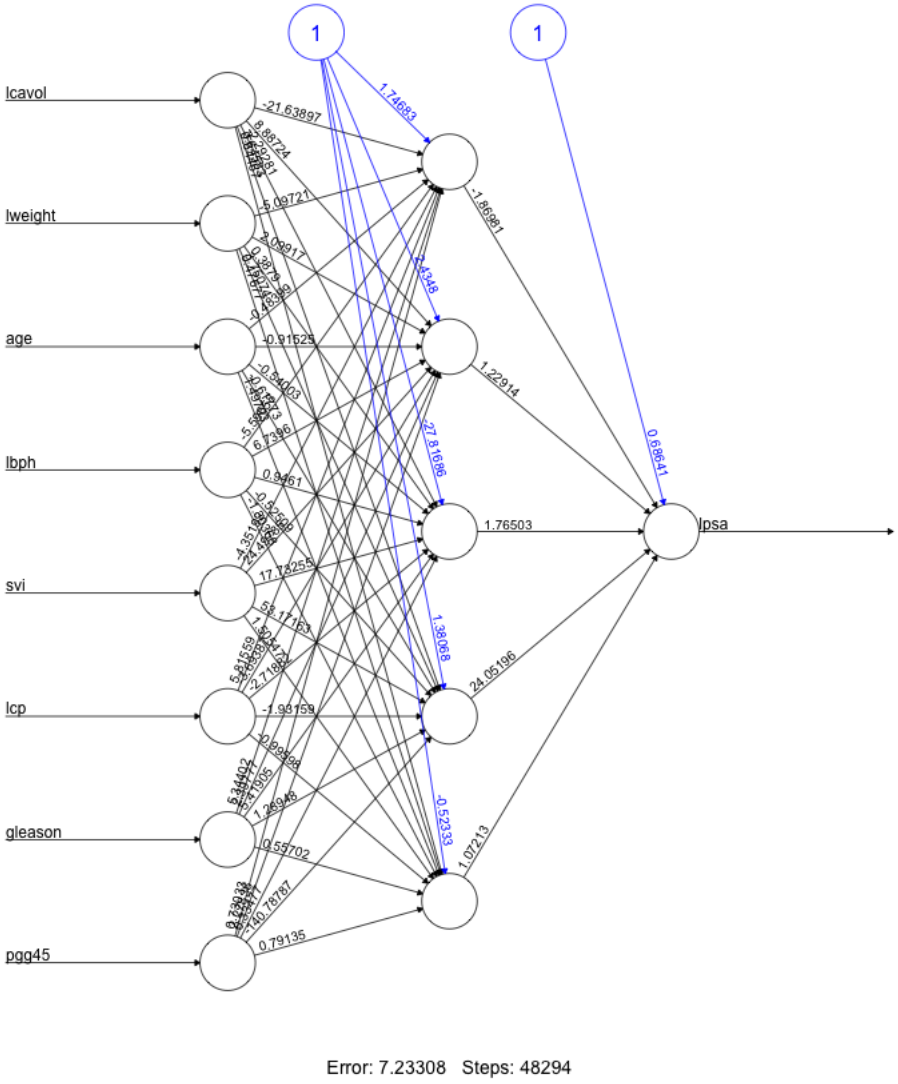
	train.target	Freq	builtInCorrect
1	0	196	41.00
2	1	113	51.00

Table 2: Logistic Regression on the South African Cancer Data

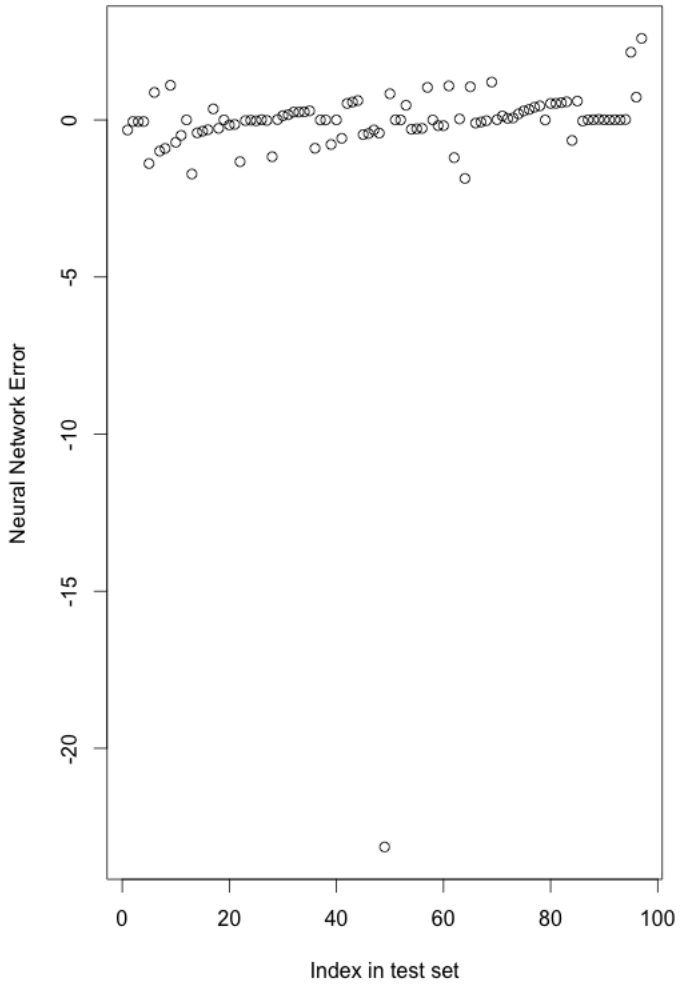
4 Neural Network Vs Linear Regression

4.1 Neural Network:

The neural network worked extremely well for the regression over the Prostate Cancer data-set. Below is the graph of the network:



Below is a graph of the error between the test and the predicted value. As you can see, the neural network did quite well on this data-set!



4.2 Linear Regression:

Linear regression failed horribly at the job of predicting the values of the Prostate Cancer data-set.

Below is a graph of the error between the test and the predicted value: