**Project #2**

(Not to worry; this is a small project, but nonetheless gives you the chance to synthesize concepts an show your mastery.)

For this project, you may work in groups of up to three people.

DUE: Wednesday, December 5 by 2:00 PM

1. (40 pts) First, let’s play a bit with neural networks and their ability to perform multi-class classification.
2. Look at the information sheet for the “BloodTransfusion.txt” data.
3. Explain the nature of the experiment that was performed and why it is important.
4. Once you have a fundamental understanding of the data, use a neural net to perform classification. Simple code for neural network classification is attached and titled “Code\_for\_categorical\_outcomes\_NN.R”
5. Use the simple validation set approach to train/test your data. What is the performance, in terms of accuracy or error, using one hidden node?
6. Plot the network.
7. Make a plot of performance (or, conversely, error) as a function of the number of nodes in one hidden layer. What is your best performance?
8. (40 pts) Now let’s compare decision tree and random forest performance (using 10-fold cross-validation) on the dataset, “Wisconsin BC Data.”
9. Familiarize yourself with the information sheet regarding the data.
10. Run the C5.0 algorithm with 10-fold cross-validation and compute the ROC area.
11. Run the random forest algorithm on the data with 10-fold cross-validation and compute the ROC area.
12. Compare the performance in a table in terms of both accuracy and AUC.