

## **DEPARTMENT OF STATISTICS**

### **STATS 760** A Survey of Modern Applied Statistics

Assignment 2 2016 part 2

Due Friday April 8. Hand in or mail to Alan by 5pm.

This is part 2 of the assignment.

Q2. If you took STATS 330 in 2015, you might remember the Boston Housing data. This was a data set consisting of observations on 506 suburbs in Boston. The aim is to predict the median house price for a suburb using 13 explanatory variables. The data, and a document describing the data in more detail is on the course web page.

Your task is to construct a predictor of the median house price using (a) a regression tree, (b) a neural network, and (c) a random forest. Use the caret package to explore different settings for the tuning parameters. For each method, calculate both the training error and the test error. Which method is doing the best job?