STAT 675 – Homework 4
Due: Nov. 13

1. Let's look at the digits data with boosting. Download the digits data[1] and read the website for relevant information about the dataset. We are going to compare classifying the 4's and 9's, which tends to be difficult. Create a training and a test data set.

   FOR each of the following base classifiers: logistic regression with 1 covariate, logistic regression with 10 covariates, trees with 1 split (stump), and trees with 10 splits; DO:

   (a) For AdaBoost, make a plot of the training error and test error as a function of the number of boosting iterations. Do you see evidence of overfitting?

   (b) Do the same for LogitBoost (You can look at `http://stat.ethz.ch/~dettling/boosting.html` for an implementation for stump classifiers)

   Which procedure combination works best?

2. Try the above, but with random forest instead, trying different combinations of `mtry` and number of bootstrap sample.

[1] `http://yann.lecun.com/exdb/mnist/`