STAT 675 – Homework 3
Due: Oct. 20

1. Let's look at the digits data with sparse logistic regression, LDA, and Random Forest

   (a) Download the digits data (`http://yann.lecun.com/exdb/mnist/`) and read the website for relevant information about the dataset.

   (b) We wish to do two-class classification in two cases. Apply the above mentioned techniques to these data sets.

      i. Comparing 3's to 5's

      ii. Comparing 4's to 9's.

      Return the test misclassification rates (use the training/test split given on the website)

      Note: Be sure to plot the results of LDA on the plane $\mathcal{H}_1$. Try different values for the feature subsampling parameter (m) (as well as the default and bagging) for random forest.

   (c) Compare the OOB error rate for the default m random forest fit to the training and testing misclassification error rates.

   (d) Which of the methods has a better test error rate?

   (e) Find the most important pixel for each classification problem for sparse logistic regression and random forest (pick the lowest miss-classification error value for m and use permutation variable importance).