Statistics 315a
Homework 1, due Wednesday January 25, 2012.

"ESL" refers to the course textbook, and ESL 2.4 refers to exercise 2.4 in ESL. Since the 4 homework assignments count 80% of your final grade, you must do them on your own. Problem 1 is computing intensive, and is partly there to get you up to speed in R. You can form teams of up to 3 students to collaborate on problem 1, but must still write up your results on your own. If so, clearly indicate in your writeup who is on the team.

1. *Error curves*

   (a) Write a function to simulate data as described on page 17 in ESL for one of the classes. Your function should take as inputs a 10×2 matrix of centroids, the sample size, and the noise variance. Generate a training sample of size 100 for each class, as well as a test sample of 5,000 per class. (Best to generate the centroids matrices per class once and store them). Try and write elegant code, that makes use of the matrix/vector facilities in `R`.

   (b) Evaluate the misclassification performance of K-nearest neighbor classification on the training and test set (`library(class)` in `R`), for $k = \{1, 3, 5, 9, 15, 25, 45, 83, 151\}$. Evaluate also the performance of the linear regression procedure. Produce a plot as in Figure 2.4.

   (c) Using the training data, use 10-fold cross-validation to estimate the errors in the cases above. Include these errors in your plot (average fold errors and estimated standard error of this average).

   (d) Summarize what you see.

2. ESL 2.4

3. ESL 2.7

4. Consider a linear regression model with p parameters, fit by least squares to a set of training data $(x_1, y_1), \ldots, (x_N, y_N)$ drawn at random from a population. Let $\hat{\beta}$ be the least squares estimate. Suppose we have some test data $(\tilde{x}_1, \tilde{y}_1), \ldots, (\tilde{x}_M, \tilde{y}_M)$ drawn at random from the same population as the training data. If $R_{tr}(\beta) = \frac{1}{N}\sum_1^N (y_i - \beta^T x_i)^2$ and $R_{te}(\beta) = \frac{1}{M}\sum_1^M (\tilde{y}_i - \beta^T \tilde{x}_i)^2$, prove that

$$E[R_{tr}(\hat{\beta})] \leq E[R_{te}(\hat{\beta})],$$

where the expectations are over all that is random in each expression.

5. ESL 3.2

6. Suppose $p \gg N$, you have a data matrix $\mathbf{X}$ and a quantitative response vector $\mathbf{y}$, and you plan to fit a linear regression model.

   (a) Explain why the ordinary least squares solution is not unique. What can you say about the residuals of any of the solutions.

   (b) Is the ridge regression solution unique? why?

   (c) Suppose you compute a series of ridge solutions, letting λ get successively smaller. What can you say about the limiting ridge solution in this case, as $\lambda \downarrow 0$.

   (d) Using the SVD of $\mathbf{X}$, write a closed form expression for this limiting solution.