1. ESL 3.12
2. ESL 3.14
3. ESL 4.2

4. Suppose you wish to build a regression model for some microarray data. In particular, you have 50 samples with expression measurements on each of 10,000 genes in a $50 \times 10,000$ array $X$. You also have an associated outcome variable $y$. You plan to use ridge regression to solve the problem.

(a) Write an expression for the ridge regression coefficient vector $\hat{\beta}_\lambda$. How big is the matrix that requires inversion? What order of computations are required to compute $\hat{\beta}_\lambda$?

(b) Show that $\hat{\beta}_\lambda$ is in the row-space of $X$ i.e. $\hat{\beta}_\lambda = X^T \alpha$ for some $\alpha$. [hint: start with the derivative condition on the penalized RSS criterion].

(c) If $X = UDV^T$ is the SVD of $X$, argue that you can represent $\hat{\beta}_\lambda = V\theta$. Show how you can use this representation to dramatically reduce the computations in a), and for all values of $\lambda$.

(d) How would you predict the outcome for a new sample with measurement vector $x_0$?

5. In this exercise we fit various regression models to the “spam” data. These data are available from the “Data” section of the book webpage: http://www-stat.stanford.edu/ElemStatLearn.

(a) Write a program to do ridge regression with 10-fold cross-validation and produce a plot for the spam data like that in the top right panel of Fig 3.6 of ESL. Make plots for both squared error, and misclassification error. Comment on any differences between the plots. For the rest of this question, focus on squared error.

(b) Write a program to implement an incremental forward stagewise approximation to the lasso, as discussed in class. Produce a cross-validation picture like that in figure 3.6 for the lasso. Draw a picture of the coefficient profiles like that in Figure 3.9.
(c) Make a plot of the residual sum of squares versus the L2 norm of the coefficients, with ridge and lasso on the same plot, to confirm that ridge regression produces a smaller sum of squares for a fixed value of the L2 norm. Do the same for the L1 norm, to confirm that the lasso produces a smaller sum of squares for a fixed value of the L1 norm.

(d) Using the tuning parameters estimates from cross-validation, apply both ridge regression and your approximate lasso to the spam test data, display and discuss the results.