1. Consider a linear regression problem where $p \gg N$, and assume the row rank of $X$ is $N$. Let the SVD of $X = UDV^T = RV^T$, where $R$ is $N \times N$ nonsingular, and $V$ is $p \times N$ with orthonormal columns.

(a) Show that there are infinitely many least-squares solutions all with zero residuals.

(b) In HW2 Problem 4(c) you showed (or were meant to!) that the ridge-regression estimate for $\beta$ can be written

$$\hat{\beta}_\lambda = V(R^TR + \lambda I)^{-1}R^Ty$$

Show that when $\lambda = 0$, the solution $\hat{\beta}_0 = VD^{-1}U^Ty$ has residuals all zero, and is unique in that it has the smallest Euclidean norm among all zero-residual solutions.

2. Piling. Exercise 4.2 in ESL shows that the two-class LDA solution can be obtained by a linear regression of a binary response vector $y$ consisting of $-1$s and $+1$s. The prediction $\hat{\beta}^T x$ for any $x$ is (up to a scale and shift) the LDA score $\delta(x)$. Suppose now that $p \gg N$.

(a) Consider the linear regression model $f(x) = \alpha + \beta^T x$ fit to a binary response $Y \in \{-1, +1\}$. Using the results of the previous problem, show that there are infinitely many directions defined by $\hat{\beta}$ in $R^p$ onto which the data project to exactly two points, one for each class. These are known as data piling directions.

(b) Show that the distance between the projected points is $2/||\hat{\beta}||$, and hence these directions define separating hyperplanes with that margin.

(c) Argue that there is a single maximal data piling direction for which this distance is largest, and is defined by $\hat{\beta}_0 = VD^{-1}U^Ty = X^{-}y$, where $X = UDV^T$ is the SVD of $X$.

(d) How does this distance compare to the optimal separating distance one would obtain using a linear support-vector machine.

(e) Generate 40 points in 50 dimensions, using standard gaussian coordinates. Assign 20 to class $+1$, and 20 to $-1$. Compute the data piling direction and the optimal separating hyperplane direction (you can use, for example, the svm procedure in R package e1071
for the latter). Compare the projected data in both cases (on the same scale), and summarize what you see.

3. Explore the use of non-linearities on the South African heart-disease data using logistic regression. The data are available from the book website. For each of the quantitative predictors you need to decide if a linear term suffices, or whether a non-linear term is more suitable. Use natural cubic splines with 4 interior knots (knot placement strategy is up to you). Use likelihood-ratio tests (deviance tests) to make your assessment.

4. ESL 7.4

5. ESL 7.5

6. ESL 9.2

7. Write a program to compare the linear SVM, and SVM with radial kernel, in the setup and style of Table 12.2. Include the lasso (via the lars package) as a 3rd competitor. (You don’t need to include the other methods in Table 12.2 in your study). In R use the library e1071 to compute the SVM. Interpret the results.