Lab Solutions 08

1. Start by doing the text analysis questions in the Rmarkdown file.

See the solutions in the HTML file for Lab08.

2. Assume that a magical creature has told you all but the first coefficient of β^{LASSO} for some value of λ. How would you go about solving for the first coefficient if you know all of the others? Start by defining the partial residual:

$$r = y - \sum_{j=2}^{p} \beta_j \cdot X_j \tag{8.1}$$

This is what is "left over" in y that we need to predict using the first variable. Finding the best value for β_1 is surprisingly similar to the uncorrelated columns case you derived last class. We want to minimize the following quantity over b_1:

$$\begin{align} f(b_1) &= ||y - Xb||_2^2 + 2\lambda ||b||_1 \tag{8.2}\\ &= ||r - X_1 b_1||_2^2 + 2\lambda \sum_j |b_j| \tag{8.3}\\ &= r^t r + X_1^t X_1 b_1^2 - 2r^t X_1 b_1 + 2\lambda \sum_j |b_j| \tag{8.4}\\ &\propto X_1^t X_1 b_1^2 - 2r^t X_1 b_1 + 2\lambda \cdot |b_1| \tag{8.5} \end{align}$$

Note that b_1 is a scalar value, so that's why I have b_1^2 in the equation and not $b_1^t b_1$. I've already done some of the work; now for your part: derive a formula for β_1 by minimizing the function $f(b_1)$. You should be able to carry through the method from last time almost exactly.

If $b_1 > 0$ then:

$$\begin{align} f(b_1) &= X_1^t X_1 b_1^2 - 2r^t X_1 b_1 + 2\lambda \cdot b_1 \tag{8.6}\\ f'(b_1) &= 2X_1^t X_1 b_1 - 2r^t X_1 + 2\lambda \tag{8.7} \end{align}$$

Setting equal to zero:

$$\begin{align} X_1^t X_1 b_1 &= r^t X_1 - \lambda \tag{8.8}\\ b_1 &= \frac{r^t X_1 - \lambda}{X_1^t X_1} \tag{8.9} \end{align}$$

Which gives a consistent value of $b_1 > 0$ if $r^t X_1 > \lambda$. If $b_1 < 0$ the function becomes:

$$f(b_1) = X_1^t X_1 b_1^2 - 2r^t X_1 b_1 - 2\lambda \cdot b_1 \tag{8.10}$$
$$f'(b_1) = 2X_1^t X_1 b_1 - 2r^t X_1 - 2\lambda \tag{8.11}$$

Setting equal to zero:

$$X_1^t X_1 b_1 = r^t X_1 + \lambda \tag{8.12}$$
$$b_1 = \frac{r^t X_1 + \lambda}{X_1^t X_1} \tag{8.13}$$

Which gives a consistent value of $b_1 < 0$ if $r^t X_1 < -1\lambda$. Putting this together gives:

$$\beta_1^{LASSO} = \begin{cases} \frac{r^t X_1 + \lambda}{X_1^t X_1} & r^t X_1 \leq -\lambda \\ 0 & |r^t X_1| \leq \lambda \\ \frac{r^t X_1 - \lambda}{X_1^t X_1} & r^t X_1 \geq \lambda \end{cases} \tag{8.14}$$

Which looks a lot like the correlated case, with the partial residuals replacing the value of y from before.

3. There's nothing special above about the first coefficient. If we define the partial residual as:

$$r = y - \sum_{j \neq k} \beta_j \cdot X_j \tag{8.15}$$

Your question also gives (replacing X_1 with X_k and b_1 with b_k) a formula for β_k in terms of the other coefficients. Coordinate descent is an optimization technique that minimizes a function f over several variables by optimizing the first variable assuming the others are fixed, then optimizing the second variable assuming the others are fixed, and so on through convergence. There is nothing specific that you need to do here, but understand how you now have a formula for running coordinate decent on the ℓ_1-penalized regression problem.

As said, there is really nothing you need to do here other than understand the technique.

4. Finally, input these formulas into the final section of the Rmarkdown file.

See the solutions in the HTML file for Lab08.