

# Handout 08: Text Analysis and Coordinate Descent

## LAB QUESTIONS

1. Start by doing the text analysis questions in the Rmarkdown file.
2. Assume that a magical creature has told you all but the first coefficient of  $\beta^{LASSO}$  for some value of  $\lambda$ . How would you go about solving for the first coefficient if you know all of the others? Start by defining the partial residual:

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

This is what is “left over” in  $y$  that we need to predict using the first variable. Finding the best value for  $\beta_1$  is surprisingly similar to the uncorrelated columns case you derived last class. We want to minimize the following quantity over  $b_1$ :

$$f(b_1) = \|y - Xb\|_2^2 + 2\lambda|b|_1 \quad (8.2)$$

$$= \|r - X_1 b_1\|_2^2 + 2\lambda \sum_j |b_j| \quad (8.3)$$

$$= r^t r + X_1^t X_1 b_1^2 - 2r^t X_1 b_1 + 2\lambda \sum_j |b_j| \quad (8.4)$$

$$\propto X_1^t X_1 b_1^2 - 2r^t X_1 b_1 + 2\lambda \cdot |b_1| \quad (8.5)$$

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  $\beta_1$  by minimizing the function  $f(b_1)$ . You should be able to carry through the method from last time almost exactly.

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

$$r = y - \sum_{j \neq k} \beta_j \cdot X_j \quad (8.6)$$

Your question also gives (replacing  $X_1$  with  $X_k$  and  $b_1$  with  $b_k$ ) a formula for  $\beta_k$  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 descent on the  $\ell_1$ -penalized regression problem.

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