STAT 745 – Fall 2014
Assignment 4

Group 1: Francene Cicia, Doug Raffle, Melissa Smith

September 22, 2014

1 Group Lasso Derivation

Let X be an orthogonal (i.e. $X'X = I$) and let $\hat{\beta}^{(ls)}$ be the least squares estimator, and in this case $\hat{\beta}^{(ls)} = X'y$. We will show that the lasso has a closed form of:

$$\hat{\beta}^{(lasso)} = sign\left(\hat{\beta}_j^{(ls)}\right)\left(|\hat{\beta}_j^{(ls)}| - \gamma\right)^+$$

Where $sign$ returns the sign of its input, $(t)^+ = max\,(0, t)$ and γ is determined so that the condition of $\Sigma|\hat{\beta}_{(j)}| = t$.

Recall: $||y - X\beta||_2^2 + \lambda||\beta||_1^1$ solves to $\hat{\beta}^{(lasso)}$.

Proof:

$$\begin{aligned}
\arg\min_\beta ||y - X\beta||_2^2 + \lambda||\beta||_1^1 &= \arg\min_\beta (y - X\beta)^T (y - X\beta) + \lambda|\beta| \\
&= y^T y - \beta^T X^T y - y^T X\beta + \beta^T X^T X\beta + \lambda|\beta| \\
&= y^T y - 2y^T X\beta + \beta^T X^T X\beta + \lambda|\beta|
\end{aligned}$$

To find minimum, take the derivative with respect to β and set equal to 0.

$$\begin{aligned}
0 &= \frac{\partial}{\partial\beta} y^T y - 2y^T X\beta_j + \beta - j^T X^T X\beta + \lambda|\beta_j| \\
0 &= -2X^T y + 2(X^T X)\hat{\beta}_j^{(lasso)} + \lambda sign\left(\hat{\beta}_j^{(lasso)}\right) \\
0 &= -X^T y + \hat{\beta}_j^{(lasso)} I + \frac{\lambda}{2} sign\left(\hat{\beta}_j^{(lasso)}\right) \\
0 &= -\hat{\beta}_j^{(ls)} + \hat{\beta}_j^{(lasso)} I + \frac{\lambda}{2} sign\left(\hat{\beta}_j^{(lasso)}\right) \\
\hat{\beta}_j^{(lasso)} &= \hat{\beta}_j^{(ls)} - \frac{\lambda}{2} sign\left(\hat{\beta}_j^{(lasso)}\right)
\end{aligned}$$

Notice that $\hat{\beta}_j^{(lasso)}$ and $\hat{\beta}_j^{(ls)}$ will have the same sign. Thus,

$$\hat{\beta}_j^{(lasso)} = \hat{\beta}_j^{(ls)} - \frac{\lambda}{2} sign\left(\hat{\beta}_j^{(ls)}\right)$$

This means,

$$f(x) = \begin{cases} \hat{\beta}_j^{(ls)} - \frac{\lambda}{2} & : \hat{\beta}_j^{(ls)} \le 0 \\ \hat{\beta}_j^{(ls)} + \frac{\lambda}{2} & : \hat{\beta}_j^{(ls)} > 0 \end{cases}$$

It follows that,

$$\hat{\beta}_j^{(lasso)} = sign\left(\hat{\beta}_j^{(ls)}\right)\left(|\hat{\beta}_j^{(ls)}| - \gamma\right)^+ . \blacksquare \tag{1}$$

$\frac{\lambda}{2}$ is some threshold constatnt that determines if a given $\hat{\beta}_j^{(lasso)}$ will be zeroed out, we will call it γ in (1). If $|\hat{\beta}_j^{(ls)}|$ is less than γ, it means that inside the parentheses results in a negative number and the function (t) will return a value of 0, thus indicating that particular $|\hat{\beta}_j^{(ls)}|$ should not be included in the final model.