(e) Repeat (a) for the irreducible error.

5. It is well-known that ridge regression tends to give similar coefficient values to correlated variables, whereas the lasso may give quite different coefficient values to correlated variables. We will now explore this property in a very simple setting.

   Suppose that $n = 2$, $p = 2$, $x_{11} = x_{12}$, $x_{21} = x_{22}$. Furthermore, suppose that $y_1 + y_2 = 0$ and $x_{11} + x_{21} = 0$ and $x_{12} + x_{22} = 0$, so that the estimate for the intercept in a least squares, ridge regression, or lasso model is zero: $\hat{\beta}_0 = 0$.

   (a) Write out the ridge regression optimization problem in this setting.

   (b) Argue that in this setting, the ridge coefficient estimates satisfy $\hat{\beta}_1 = \hat{\beta}_2$.

   (c) Write out the lasso optimization problem in this setting.

   (d) Argue that in this setting, the lasso coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are not unique—in other words, there are many possible solutions to the optimization problem in (c). Describe these solutions.

Answer

$n = 2, \ p = 2$

$x_{11} = x_{12} \quad , \quad x_{21} = x_{22}$

$y_1 + y_2 = 0 \ , \ x_{11} + x_{21} = 0 \ , \ x_{12} + x_{22} = 0$

$\hat{\beta}_0 = 0$

a) Ridge Regression Optimization Problem

Minimize $\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$

$\hat{\beta}_0 = 0$; $n = p = 2$

$\therefore$ Objective Function is

$$\text{Minimize} \left\{ (y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2 + (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2 + \lambda (\hat{\beta}_1^2 + \hat{\beta}_2^2) \right\}$$

b) Given that $x_{11} = x_{12} = x_1$; $x_{21} = x_{22} = x_2$

Let objective function be M

$$M = (y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)^2 + \lambda (\hat{\beta}_1^2 + \hat{\beta}_2^2)$$

Differentiating M w.r.t $\hat{\beta}_1$ and $\hat{\beta}_2$, and setting it to 0, we get

$$\frac{dM}{d\hat{\beta}_1} = -\cancel{2}(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1) x_1$$
$$- \cancel{2}(y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2) x_2$$
$$+ \cancel{2}\lambda \hat{\beta}_1 = 0$$

$$\therefore -y_1 x_1 + \hat{\beta}_1 x_1^2 + \hat{\beta}_2 x_1^2$$
$$- y_2 x_2 + \hat{\beta}_1 x_2^2 + \hat{\beta}_2 x_2^2$$
$$+ \lambda \hat{\beta}_1 = 0$$

$$\therefore \hat{\beta}_1 (+x_1^2 + x_2^2 + \lambda) = +y_1 x_1 - \hat{\beta}_2 x_1^2 + y_2 x_2 - \hat{\beta}_2 x_2^2$$

$$\therefore \hat{\beta}_1 = \frac{y_1 x_1 + y_2 x_2 - \hat{\beta}_2 x_1^2 - \hat{\beta}_2 x_2^2}{x_1^2 + x_2^2 + \lambda}$$

1

$$\frac{dM}{d\hat{\beta}_2} = 0 = -2(y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)x_1$$

$$-2(y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)x_2$$

$$+2\lambda\hat{\beta}_2$$

$$\therefore -y_1 x_1 + \hat{\beta}_1 x_1^2 + \hat{\beta}_2 x_1^2$$

$$-y_2 x_2 + \hat{\beta}_1 x_2^2 + \hat{\beta}_2 x_2^2$$

$$+\lambda\hat{\beta}_2 = 0$$

$$\therefore \hat{\beta}_2 = \frac{y_1 x_1 + y_2 x_2 - \hat{\beta}_1 x_1^2 - \hat{\beta}_1 x_2^2}{x_1^2 + x_2^2 + \lambda} \quad (2)$$

The symmetry in equations (1) and (2) suggest that $\hat{\beta}_1 = \hat{\beta}_2$.

c) Lasso Optimization Problem

$$\text{Minimize} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j=1}^{p} \hat{\beta}_j x_j) + \lambda \sum_{j=1}^{p} |\hat{\beta}_j|$$

$\therefore$ Objective Function is

$$\text{Minimize} \{ (y_1 - \hat{\beta}_1 x_{11} - \hat{\beta}_2 x_{12})^2$$
$$+ (y_2 - \hat{\beta}_1 x_{21} - \hat{\beta}_2 x_{22})^2$$
$$+ \lambda (|\hat{\beta}_1| + |\hat{\beta}_2|) \}$$

d) Given that $x_{12} = x_{11} = x_1$; $x_{21} = x_{22} = x_2$

$$M = (y_1 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_1)^2 + (y_2 - \hat{\beta}_1 x_2 - \hat{\beta}_2 x_2)^2$$

subject to $(|\hat{\beta}_1| + |\hat{\beta}_2| \leq s)$

$$\therefore M = (y_1 - x_1 (\hat{\beta}_1 + \hat{\beta}_2))^2$$
$$+ (y_2 - x_2 (\hat{\beta}_1 + \hat{\beta}_2))^2$$

$$\therefore M = y_1^2 + x_1^2 (\hat{\beta}_1 + \hat{\beta}_2)^2 - 2 x_1 y_1 (\hat{\beta}_1 + \hat{\beta}_2)$$

$$+ y_2^2 + x_2^2 (\hat{\beta}_1 + \hat{\beta}_1)^2 - 2 x_2 y_2 (\hat{\beta}_1 + \hat{\beta}_2)$$

But $y_1 + y_2 = 0 \Rightarrow y_1 = -y_2 \Rightarrow y_1^2 = y_2^2$

$x_1 + x_2 = 0 \Rightarrow x_1^2 = x_2^2$

$$\therefore M = 2y_1^2 + 2x_1^2 (\hat{\beta}_1 + \hat{\beta}_2)^2$$

$$- 4 x_1 y_1 (\hat{\beta}_1 + \hat{\beta}_2)$$

$$\therefore M = 2 \left(y_1^2 + x_1^2 (\hat{\beta}_1 + \hat{\beta}_2)^2 \right.$$

$$\left. - 2 x_1 y_1 (\hat{\beta}_1 + \hat{\beta}_2) \right)$$

$$\therefore M = 2 \left(y_1 + x_1 (\hat{\beta}_1 + \hat{\beta}_2) \right)^2$$

Subject to $|\hat{\beta}_1| + |\hat{\beta}_2| \leq s$

Since a squared term is involved in the optimization term, we can see that if both $\hat{\beta}_1$ and $\hat{\beta}_2$ are negative and positive at the same time, then the optimization problem is still satisfied since $|\hat{\beta}_1| + |\hat{\beta}_2| \leq s$.

Since more than 1 solution exists, the lasso coefficients do not have a unique solution.

The entire edge $\hat{\beta}_1 + \hat{\beta}_2 = s$ is a solution at different points in the cartesian 3d graph.

Similarly, $\hat{\beta}_1 + \hat{\beta}_2 = -s$ is also a solution.