

Computational Statistics Midterm

Liam Dillingham

April 7, 2019

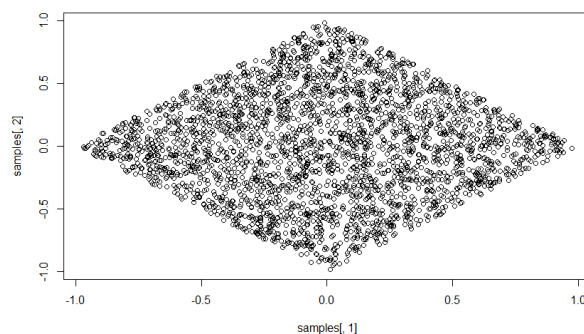
1 Problem 1

Derive any one scheme for drawing samples from the diamond shaped area:

The R code for this is shown below:

```
arfunction = function() {  
  while(TRUE) {  
    x = runif(1,min = -1,max = 1)  
    y = runif(1,min = -1, max = 1)  
    if (abs(x) + abs(y) <= 1){  
      rv = c(x, y)  
      return(rv)  
    }  
  }  
}  
  
samples = matrix(nrow = 3000, ncol = 2)  
for(i in 1:3000) {  
  x = arfunction()  
  samples[i,1] = x[1]  
  samples[i,2] = x[2]  
}  
plot(samples[,1], samples[,2])
```

And the corresponding plot:



2 Problem 2

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

2.1 Write out the ridge regression optimization problem in this setting

I'm trying to visualize the problem at hand here. We have a system that looks like this:

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

So we want to find the values of the $\hat{\beta}_i$'s in the equation $x_{i1}\hat{\beta}_1 + x_{i2}\hat{\beta}_2 = \hat{y}_i$ such that the computed value, $(\hat{y}_i - y_i)^2$ is minimized for all i . In our $p = 2$ case, we want to minimize $(\hat{y}_1 - y_1)^2 + (\hat{y}_2 - y_2)^2$. However, since we are performing ridge regression, we alter this equation such that we are minimizing the equation:

$$(y_1 - (m_i x_{1i} + b_i))^2 + (y_2 - (m_i x_{2i} + b_i))^2 + \lambda m_i^2$$

For each $i \in p = 2$. By doing some algebra, we rearrange the equation to:

$$(y_1^2 + y_2^2) - 2m_i(x_{1i}y_1 + x_{2i}y_2) - 2b_i(y_1 + y_2) + m_i^2(x_{1i}^2 + x_{2i}^2) + 2m_i b_i(x_{1i} + x_{2i}) + 2b_i^2 + \lambda m_i^2$$

Since $\bar{x}_i = \frac{x_1 + x_2 + \dots + x_n}{n}$, and in this case, $n = p = 2$, then $\bar{x}_i = \frac{x_1 + x_2}{2}$, or $x_1 + x_2 = 2\bar{x}_i$. So we can replace a few of these terms:

$$2\bar{y}^2 - 4m_i\bar{x}_i\bar{y} - 4b_i\bar{y} + 2m_i^2\bar{x}_i^2 + 4m_i b_i\bar{x}_i + 2b_i^2 + \lambda m_i^2$$

To find the line of best fit, we want to find the minimum point on the error surface produced by m and b . We can find those values by calculating the partial derivatives for m and b , and setting them equal to 0.

$$\frac{\partial SSE}{\partial m_i} = -4\bar{x}_i\bar{y} + 4m_i\bar{x}_i^2 + 4b_i\bar{x}_i + 2\lambda m_i = 0$$

$$\frac{\partial SSE}{\partial b_i} = -4\bar{y} + 4m_i\bar{x}_i + 4b_i = 0$$

We want to coerce these equations into a $y = mx + b$ form, so we do some more algebraic manipulation:

$$\frac{\partial SSE}{\partial m_i} = m_i\left(\frac{4\bar{x}_i^2}{4\bar{x}_i} + 2\lambda\right) + b_i = \frac{\bar{x}_i\bar{y}}{\bar{x}_i} = 0$$

$$\frac{\partial SSE}{\partial b_i} = m_i\bar{x}_i + b_i = \bar{y} = 0$$

And then we have a system of equations, and we can solve for m and b . we multiply the partial derivative with respect to m by -1, and add it to the other equation:

$$m_i(\overline{x_i} - \frac{4\overline{x_i^2} + 2\lambda}{4\overline{x_i}}) = \overline{y} - \frac{\overline{x_i y}}{\overline{x_i}}$$

Then, to isolate m_i , we divide both sides by its coefficient and simplify:

$$m_i = \frac{4\overline{y} \overline{x_i} - \overline{x_i y}}{4(\overline{x_i})^2 - 4\overline{x_i^2} + 2\lambda}$$

And to solve for b_i we substitute back into the equation and solve: $m_i \overline{x_i} + b_i = \overline{y} \Rightarrow b_i = \overline{y} - m_i \overline{x_i}$.

$$b_i = \overline{y} - \frac{4\overline{y} \overline{x_i} - \overline{x_i y}}{4(\overline{x_i})^2 - 4\overline{x_i^2} + 2\lambda} \overline{x_i} = \overline{y} - \frac{4\overline{y} (\overline{x_i})^2 - \overline{x_i y} \overline{x_i}}{4(\overline{x_i})^2 - 4\overline{x_i^2} + 2\lambda}$$

Now, it is worth noting a few key facts about this problem. Since $y_1 + y_2 = 0$, $x_{12} + x_{22} = 0$, and $x_{11} + x_{21} = 0$, then $x_{1i} + x_{2i} = 0$, for all i . Thus $\overline{x_i} = 0$. So we can eliminate a few terms.

$$m_i = \frac{4\overline{y} \cdot 0 - \overline{x_i y}}{4(0)^2 - 4\overline{x_i^2} + 2\lambda} = \frac{-\overline{x_i y}}{-4\overline{x_i^2} + 2\lambda}$$

$$b_i = 0 - \frac{0 - \overline{x_i y} \cdot 0}{4(0)^2 - 4\overline{x_i^2} + 2\lambda} = 0$$

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

note that $x_{11} = x_{12}$, $x_{21} = x_{22}$. Since $b_i = 0$ for all i , then we don't need to calculate b . However, let's calculate m_1 .

$-\overline{x_1 y} = \frac{1}{2}(x_{11}y_1 + x_{21}y_2)$, and $\overline{x_1^2} = \frac{1}{2}(x_{11}^2 + x_{21}^2)$. So,

$$m_1 = \frac{-\frac{1}{2}(x_{11}y_1 + x_{21}y_2)}{-2(x_{11}^2 + x_{21}^2) + 2\lambda} \text{ and likewise,}$$

$$m_2 = \frac{-\frac{1}{2}(x_{12}y_1 + x_{22}y_2)}{-2(x_{12}^2 + x_{22}^2) + 2\lambda}$$

using the identities $x_{11} = x_{12}$, $x_{21} = x_{22}$ in the equation for m_2 ,

$$m_2 = \frac{-\frac{1}{2}(x_{11}y_1 + x_{21}y_2)}{-2(x_{11}^2 + x_{21}^2) + 2\lambda} = m_1$$

Using the notation $\hat{\beta}_1 = m_1$ and $\hat{\beta}_2 = m_2$, we have satisfied $\hat{\beta}_1 = \hat{\beta}_2$.

2.3 Write out the lasso optimization problem in this setting.

Note that for lasso regression, instead of penalizing the SSE by λm_i^2 , we penalize it by $\lambda |m_i|$. We can actually skip a few of the initial steps of the previous problem, since the differences begin at derivation. Note that instead of differentiating m_i^2 , we differentiate $|m_i|$, where $\frac{\partial SSE}{\partial m_i} = \frac{m_i}{|m_i|}$. Starting with our original equation adjusted for lasso regression:

$$2\bar{y}^2 - 4m_i\bar{x}_i\bar{y} - 4b_i\bar{y} + 2m_i^2\bar{x}_i^2 + 4m_ib_i\bar{x}_i + 2b_i^2 + \lambda|m_i|$$

We then take the partial derivatives:

$$\frac{\partial SSE}{\partial m_i} = -4\bar{x}_i\bar{y} + 4m_i\bar{x}_i^2 + 4b_i\bar{x}_i + \lambda \frac{m_i}{|m_i|} = 0$$

$$\frac{\partial SSE}{\partial b_i} = -4\bar{y} + 4m_i\bar{x}_i + 4b_i = 0$$

However, due to the absolute value of m_i , we will be breaking this equation down into two cases, $m_i = m_i$ when $m_i \geq 0$, and $m_i = -m_i$ when $m_i < 0$. And again, coercing the equations into our $mx + b$ form:

Case 1: $m_i < 0$

$$\frac{\partial SSE}{\partial m_i} = -4\bar{x}_i\bar{y} + 4m_i\bar{x}_i^2 + 4b_i\bar{x}_i + \lambda \frac{m_i}{-m_i} = 0$$

$$\frac{\partial SSE}{\partial m_i} = m_i \frac{\bar{x}_i^2}{\bar{x}_i} + b_i = \frac{\bar{x}_i\bar{y}}{\bar{x}_i} + \frac{1}{4}\lambda \quad \text{And for } b_i,$$

$$\frac{\partial SSE}{\partial b_i} = m_i\bar{x}_i + b_i = \bar{y} = 0$$

We multiply the equation for m_i by -1, add the two equations together, and simplify

$$m_i = \frac{\bar{x}_i\bar{y} - \bar{x}_i\bar{y} - \frac{\lambda}{4}}{(\bar{x}_i)^2 - \bar{x}_i^2}$$

Let's first simplify this equation further with the previously known fact that $\bar{x}_i = 0$.

$$m_i = \frac{-\bar{x}_i\bar{y} - \frac{\lambda}{4}}{-\bar{x}_i^2} \quad \text{then } b_i \text{ equals:}$$

$$b_i = \frac{-\bar{x}_i\bar{y} - \frac{\lambda}{4}}{-\bar{x}_i^2} \quad 0 + b_i = 0, \text{ Since we already know } \bar{x}_i \text{ and } \bar{y}_i = 0$$

now for case 2.

Case 2: $m_i \geq 0$

$$\frac{\partial SSE}{\partial m_i} = -4\bar{x}_i\bar{y} + 4m_i\bar{x}_i^2 + 4b_i\bar{x}_i + \lambda \frac{m_i}{m_i} = 0$$

$$\frac{\partial SSE}{\partial m_i} = m_i \frac{\bar{x}_i^2}{\bar{x}_i} + b_i = \frac{\bar{x}_i\bar{y}}{\bar{x}_i} - \frac{1}{4}\lambda \quad \text{And for } b_i,$$

$$\frac{\partial SSE}{\partial b_i} = m_i\bar{x}_i + b_i = \bar{y} = 0$$

The situation is very similar. We can already infer that $b_i = 0$, due to the multiplication from the fact that \bar{x}_i and $\bar{y}_i = 0$. Thus,

$$m_i = \frac{-\bar{x}_i\bar{y} + \frac{\lambda}{4}}{-\bar{x}_i^2}$$

- 2.4 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 (3). Describe these solutions.

The values for the two $\hat{\beta}_1$ and $\hat{\beta}_2$