Paul Steller
MA 637
Spring 2020
Assignment 2, Due Mar 11
In collaboration with:

---

**Question 1.** *Consider the Ridge regression problem*

$$\hat{\beta}_{ridge} = \arg\min_{\beta \in \mathbb{R}^p} ||y - X\beta||_2^2 + \alpha||\beta||_2^2.$$

*Follow the steps below to explicitly construct an example of a matrix $X$ and a vector $y$ for which $\hat{\beta}_{ridge}$ does not have a single entry equal to 0 for all $\alpha > 0$.*

*(a) Let $X = \mathbf{1}_{1\times p}$ and $y = 1$. Prove that for any $t > 0$, the optimization problem*

$$\min \sum_{i=1}^{p} \beta_i^2 \textit{ such that } \sum_{i=1}^{p} \beta_i = t$$

*admits a unique solution $\beta_i = t/p$ for all $i = 1, \ldots, p$. (Note: You may want to use the method of Lagrange multipliers).*

*(b) Now, assume $\sum_{i=1}^{p} \beta_i = t$. Use part (a) to conclude that*

$$||y - X\beta||_2^2 + \alpha||\beta||_2^2 \geq (1-t)^2 + \alpha \cdot \frac{t^2}{p},$$

*with equality if and only if $\beta_i = t/p$ for all $i$.*

*(c) Verify that the function $f(t) = (1-t)^2 + \alpha \cdot \frac{t^2}{p}$ has a global minimum at $t^* = \frac{p}{\alpha+p}$ and that $f(t^*) = \alpha/(\alpha+p)$. Hence*

$$||y - X\beta||_2^2 + \alpha||\beta||_2^2 \geq \frac{\alpha}{\alpha+p}$$

*with equality if and only if $\beta = \frac{1}{\alpha+p} \cdot \mathbf{1}_{p\times 1}$.*

*(d) Conclude that the solution of the Ridge regression problem with $X$ and $y$ given above is $\hat{\beta}_{ridge} = \frac{1}{\alpha+p} \cdot \mathbf{1}_{p\times 1}$.*

**Solution 1.**
$(a)$ First, we wish to show that $\min \sum_{i=1}^{p} \beta_i^2$ such that $\sum_{i=1}^{p} \beta_i = t$ admits a unique solution $\beta_i = t/p$ for all $i = 1, \ldots, p$. This will be shown using the method of Lagrange multipliers. Let $f(\beta_1, \beta_2, ..., \beta_p) = \sum_{i=1}^{p} \beta_i^2$, and let $g(\beta_1, \beta_2, ..., \beta_p) = \sum_{i=1}^{p} \beta_i - t$. Note that

$$\nabla f(\beta_1, \beta_2, ..., \beta_p) = \begin{bmatrix} 2\beta_1 \\ 2\beta_2 \\ \vdots \\ 2\beta_p \end{bmatrix}, \text{ and } \nabla g(\beta_1, \beta_2, ..., \beta_p) = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Then we can solve $\nabla f(\beta_1, \beta_2, ..., \beta_p) - \lambda \nabla g(\beta_1, \beta_2, ..., \beta_p) = 0$ to obtain our minimum:

$$\begin{bmatrix} 2\beta_1 \\ 2\beta_2 \\ \vdots \\ 2\beta_p \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \Rightarrow \beta_i = \frac{\lambda}{2}, \ \forall i = 1, ..., p$$

Substituting this into $\sum_{i=1}^{p} \beta_i = t$, we get

$$\sum_{i=1}^{p} \frac{\lambda}{2} = t \Rightarrow \frac{p\lambda}{2} t \Rightarrow \lambda = \frac{2t}{p}$$

And thus, $\beta_i = \frac{\lambda}{2} = \frac{2t}{2p} = \frac{t}{p}$ for all $i = 1, ..., p$ is the unique solution that minimizes $\sum_{i=1}^{p} \beta_i^2$.

$(b)$ From the first part of the problem, we know that

$$\sum_{i=1}^{p} \beta_i^2 \geq \sum_{i=1}^{p} (\frac{t}{p})^2 = p(\frac{t^2}{p^2}) = \frac{t^2}{p}.$$

And thus,

$$\alpha||\beta||_2^2 = \alpha \sum_{i=1}^{p} \beta_i^2 \geq \alpha \frac{t^2}{p}$$

Next, notice that, since $X = \mathbf{1}_{1\times p}$, we have that $X\beta = \sum_{i=1}^{p} \beta = t$. And so,

$$||y - X\beta||_2^2 = ||1 - t||_2^2 = (1-t)^2$$

Putting this all together, we get

$$||y - X\beta||_2^2 + \alpha||\beta||_2^2 \geq (1-t)^2 + \alpha \frac{t^2}{p}$$

as desired. Notice that since $||y - X\beta||_2^2 = (1-t)^2$, we can simplify the above inequality:

$$||y - X\beta||_2^2 + \alpha||\beta||_2^2 \geq (1-t)^2 + \alpha \frac{t^2}{p}$$

$$(1-t)^2 + \alpha||\beta||_2^2 \geq (1-t)^2 + \alpha \frac{t^2}{p}$$

$$\alpha||\beta||_2^2 \geq \alpha \frac{t^2}{p}$$

$$\sum_{i=1}^{p} \beta_i^2 \geq \frac{t^2}{p}$$

From part $(a)$, we have a unique solution $\beta_i = \frac{t}{p}$ $\forall i$ that minimizes the sum. Thus, we achieve equality only by minimizing the sum:

$$\frac{t^2}{p} = \sum_{i=1}^{p} (\frac{t}{p})^2 = \sum_{i=1}^{p} \beta_i^2 \geq \frac{t^2}{p}$$

In other words, we achieve equality if and only if $\beta_i = t/p$ for all $i = 1, ..., p$.

$(c)$ First, we wish to verify that $f(t) = (1-t)^2 + \alpha\frac{t^2}{p}$ has a global minimum at $t^* = \frac{p}{\alpha+p}$. To do this, we take the derivative and set it equal to zero:

$$f'(t^*) = -2(1-t^*) + 2\alpha\frac{t^*}{p} = 0$$

$$2t^* - 2 + \frac{2\alpha}{p}t^* = 0$$

$$t^*(\frac{2p+2\alpha}{p}) = 2$$

$$t^* = \frac{p}{\alpha+p}$$

To verify that it is a minimum, we examine the second derivative:

$$f''(t^*) = 2 + \frac{2\alpha}{p}$$

Since $p$ and $\alpha$ are both positive, $f''(t^*) > 0$, and so we have a minimum. Since $f(t)$ is quadratic, we have a unique minimum.

$$f(t^*) = (1-\frac{p}{\alpha+p})^2 + \alpha\frac{(\frac{p}{\alpha+p})^2}{p} = (\frac{p+\alpha-p}{p+\alpha})^2 + \frac{\alpha p}{(\alpha+p)^2} = \frac{\alpha^2}{(p+\alpha)^2} + \frac{\alpha p}{(p+\alpha)^2} = \frac{\alpha(\alpha+p)}{(\alpha+p)^2} = \frac{\alpha}{\alpha+p}$$

Next, applying this to our result from part $(b)$, we see that

$$||y - X\beta||_2^2 + \alpha||\beta||_2^2 \geq (1-t)^2 + \alpha\frac{t^2}{p} \geq \frac{\alpha}{\alpha+p}$$

Note, the first inequality obtains equality if and only if $\beta_i = t/p$ for all $i = 1, ..., p$. The second inequality obtains equality if and only if $t = \frac{p}{p+\alpha}$. Putting this together, we see that

$$||y - X\beta||_2^2 + \alpha||\beta||_2^2 = \frac{\alpha}{\alpha+p}$$

if and only if $\beta_i = \frac{\frac{p}{p+\alpha}}{p} = \frac{1}{p+\alpha}$ for all $i = 1, ..., p$.

$(d)$ If $\beta = \frac{1}{\alpha+p} \cdot \mathbf{1}_{p\times 1}$, parts $(a), (b)$, and $(c)$ have shown that $||y - X\beta||_2^2 + \alpha||\beta||_2^2$ is minimized. Note, this only applies for $X = \mathbf{1}_{1\times p}$ and $y = 1$. Thus, this is the solution to the Ridge regression problem for this specific case.

**Question 2.** *Suppose $X \in \mathbb{R}^{n\times p}$ has orthonormal columns and $y \in \mathbb{R}^{n\times 1}$. Show that the solution of the lasso problem*

$$\hat{\beta}^{lasso} = \arg\min_{\beta\in\mathbb{R}^p} \frac{1}{2}||y - X\beta||_2^2 + \alpha||\beta||_1 \ (\alpha > 0)$$

*is obtained by soft-thresholding the lease squares solution, i.e.,*

$$\hat{\beta}_i^{lasso} = sgn(\hat{\beta}_i^{LS})(|\hat{\beta}_i^{LS}| - \alpha)_+ \ (i = 1, \ldots, p),$$

*where $\hat{\beta}^{LS}$ denotes the lease squares solution of the problem,*

$$sgn(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ 0 & x = 0 \end{cases},$$

*and where $(x)_+ := \max(x, 0)$.*

**Solution 2.**
First, note that since $X \in \mathbb{R}^{n\times p}$ has orthonormal columns, $X^TX = I,\ ||X_i||_2^2 = 1$, and so

$$\beta^{LS} = (X^TX)^{-1}X^Ty = I^{-1}X^Ty = X^Ty$$

From our notes, we know that $x^* = \eta^S_{\alpha/||X_i||_2^2}(\frac{X_i^T(y-X_{-i}x_{-i})}{X_i^TX_i})$ is the solution to the lasso problem, which can be simplified as

$$x^* = \eta^S_{\alpha/1}(\frac{X_i^Ty - X_i^TX_{-i}x_{-i}}{1}) = \eta^S_\alpha(X_i^Ty - X_i^TX_{-i}x_{-i})$$

But $X_i^TX_{-i}$ is the $1 \times (p-1)$ all zero matrix, and so $X_i^TX_{-i}x_{-i} = 0$. So we have

$$x^* = \eta^S_\alpha(X_i^Ty) = \eta^S_\alpha(\beta_i^{LS}) = sgn(\beta_i^{LS})(|\beta_i^{LS}| - \alpha)_+$$

as desired.