Math 680 - Assignment #2

Kevin McGregor

October 15th, 2015

Question 1

We're trying to minimize $\|Y - X\beta\|^2$ with respect to β. We know that the least squares criterion is a convex function, therefore the minimum we find will be a global minimum. The function can be rewritten as:

$$\begin{aligned} \|Y - X\beta\|^2 &= (Y - X\beta)^\top (Y - X\beta) \\ &= Y^\top Y - (X\beta)^\top Y - Y^\top X\beta + (X\beta)^\top X\beta \\ &= Y^\top Y - 2(X\beta)^\top Y + \beta^\top X^\top X\beta. \end{aligned}$$

Then, differentiating with respect to β gives:

$$\nabla_\beta \|Y - X\beta\|^2 = -2X^\top Y + 2X^\top X\beta.$$

So, if we plug in the centered least squares estimate for β into this equation, we know it will be a global minimum for the least squares criterion.

WILL COME BACK TO THIS!

Question 2

(a)

$$\begin{aligned} \|\tilde{Y} - \tilde{X}\beta\|^2 + \lambda\|\beta\|^2 &= (\tilde{Y} - \tilde{X}\beta)^\top (\tilde{Y} - \tilde{X}\beta) + \lambda\beta^\top\beta \\ &= \tilde{Y}^\top \tilde{Y} - 2\beta^\top \tilde{X}^\top \tilde{Y} + \beta^\top \tilde{X}^\top \tilde{X}\beta + \lambda\beta^\top\beta \\ \nabla f(\beta) &= -2\tilde{X}^\top \tilde{Y} + 2\tilde{X}^\top \tilde{X}\beta + 2\lambda\beta \end{aligned}$$

(b)

The function f is quadratic in β, and the matrix in the quadratic term is $\tilde{X}^\top \tilde{X} + \lambda I$, which is positive definite. Therefore f is strictly convex.

(c)

Since f is strictly convex, there is only one global minimizer of f.

(d)

See R code in "a2_q2d.R" and "centeredRidge_func.R"

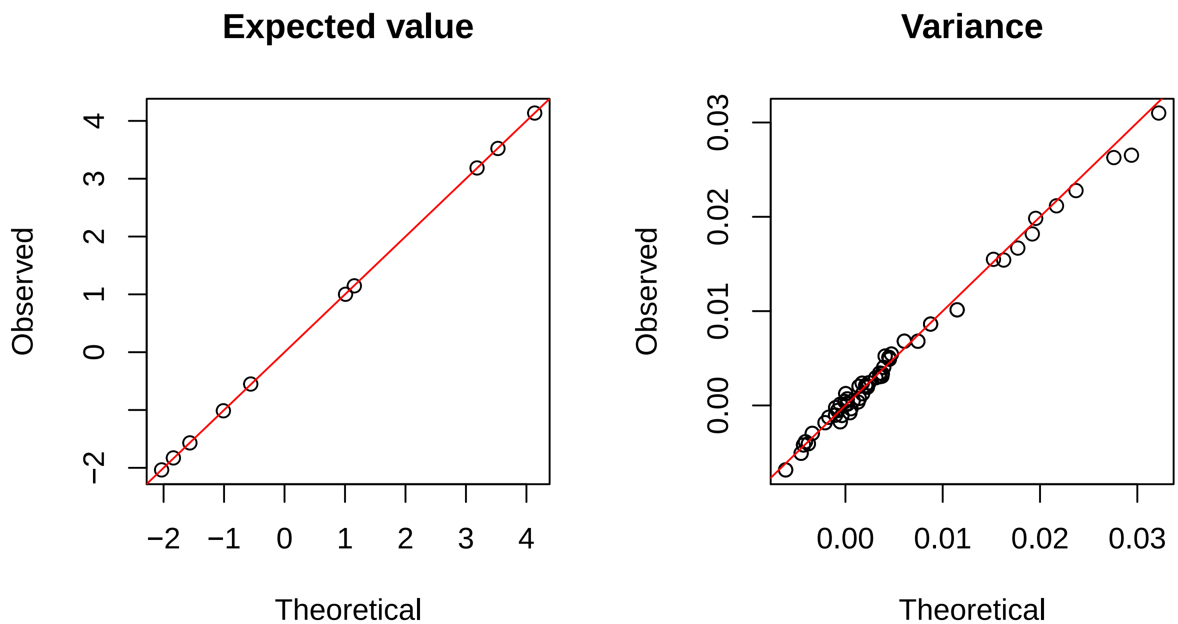

Figure 1: Observed vs. Theoretical Expectation and Variance for the centered ridge estimator

(e)

The solution to ridge regression is found by setting $\nabla f(\beta) = 0$, which gives us $\hat{\beta}^{(\lambda)}_{-1} = (\tilde{X}^\top \tilde{X} + \lambda I)^{-1} \tilde{X}^\top \tilde{Y}$. So we have the expected value:

$$\begin{aligned} E[\hat{\beta}^{(\lambda)}_{-1}] &= E[(\tilde{X}^\top \tilde{X} + \lambda I)^{-1} \tilde{X}^\top \tilde{Y}] \\ &= (\tilde{X}^\top \tilde{X} + \lambda I)^{-1} \tilde{X}^\top E[\tilde{Y}] \\ &= (\tilde{X}^\top \tilde{X} + \lambda I)^{-1} \tilde{X}^\top \tilde{X} \beta_{*-1}, \end{aligned}$$

and the variance:

$$\begin{aligned} var[\hat{\beta}^{(\lambda)}_{-1}] &= var[(\tilde{X}^\top \tilde{X} + \lambda I)^{-1} \tilde{X}^\top \tilde{Y}] \\ &= [(\tilde{X}^\top \tilde{X} + \lambda I)^{-1} \tilde{X}^\top] var[\tilde{Y}] [(\tilde{X}^\top \tilde{X} + \lambda I)^{-1} \tilde{X}^\top]^\top \\ &= \sigma_*^2 [(\tilde{X}^\top \tilde{X} + \lambda I)^{-1} \tilde{X}^\top] [(\tilde{X}^\top \tilde{X} + \lambda I)^{-1} \tilde{X}^\top]^\top, \end{aligned}$$

since $var[\tilde{Y}] = \sigma_*^2 I$.

In the simulation study, we independently generate Y 1000 times. The R code for this can be seen in the file “a2_q2d.R”. Figure 1 plots the theoretical vs. observed expectation and variance. Since most points fall close to the diagonal, the observed expectation and variance were close to the theoretical values.

Question 3

See R code in “ridge_crossval.R”.

Question 4

Question 5

(a)

The objective function can be rewritten as:

$$\begin{aligned} g(\beta, \sigma^2) &= \frac{n}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\|\tilde{Y} - \tilde{X}\beta\|^2 + \frac{\lambda}{2}\|\beta\|^2 \\ &= \frac{n}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}(\tilde{Y}^\top\tilde{Y} - 2\beta^\top\tilde{X}^\top\tilde{Y} + \beta^\top\tilde{X}^\top\tilde{X}\beta) + \frac{\lambda}{2}\beta^\top\beta, \end{aligned}$$

then the partials are calculated as:

$$\nabla_\beta g(\beta, \sigma^2) = \frac{1}{2\sigma^2}\left(-2\tilde{X}^\top\tilde{Y} + 2\tilde{X}^\top\tilde{X}\beta\right) + \lambda\beta$$

$$\nabla_{\sigma^2} g(\beta, \sigma^2) = \frac{n}{2\sigma^2} - \frac{(\tilde{Y}^\top\tilde{Y} - 2\beta^\top\tilde{X}^\top\tilde{Y} + \beta^\top\tilde{X}^\top\tilde{X}\beta)}{(\sigma^2)^2}.$$

(b)

The second derivative with respect to σ^2 is given by:

$$\nabla^2_{\sigma^2} g(\beta, \sigma^2) = -\frac{n}{2(\sigma^2)^2} + \frac{2(\tilde{Y}^\top\tilde{Y} - 2\beta^\top\tilde{X}^\top\tilde{Y} + \beta^\top\tilde{X}^\top\tilde{X}\beta)}{(\sigma^2)^3},$$

Since this could be negative, we know that g cannot be convex.

Question 6