# Stat 710 Homework 6

Due Tuesday, March 26, 11:59pm

## Assignment

In this homework you will look at two methods for penalized regression that also perform variable selection. Recall that in the standard regression problem, we minimize $\sum_{i=1}^{n}(y_i - x_i^T\beta)^2$.

The standard regression problem can be generalized to losses beyond the standard squared error loss, in which case we minimize $\sum_{i=1}^{n} P(y_i - x_i^T\beta)$, where $P$ is a penalty function.

A second generalization involves putting a penalty on the coefficient vector in addition to the residuals, in which case we minimize

$$\sum_{i=1}^{n} P_1(y_i - x_i^T\beta) + \sum_{j=1}^{p} P_2(\beta_j)$$

where $P_1$ and $P_2$ are penalty functions.

Here we'll look at four penalty functions:

- Quadratic/$\ell_2$: $Q(z) = z^2$
- Lasso penalty/$\ell_1$: $L(z) = |z|$
- Huber penalty:

$$H(z, M) = \begin{cases} z^2 & |z| \leq M \\ 2M|z| - M^2 & |z| \geq M \end{cases} \tag{1}$$

- Berhu penalty:

$$B(z, M) = \begin{cases} |z| & |z| \leq M \\ \frac{z^2+M^2}{2M} & |z| > M \end{cases} \tag{2}$$

Next week we'll fit regressions with these penalties; this week we will just look at some of their properties.

1. Suppose that $\mathbf{z} \in \mathbb{R}^n$, and $z_i$ is the $i$th element of $z$. Write down the gradients for the following functions:
   - $\sum_{i=1}^{n} Q(z_i)$
   - $\sum_{i=1}^{n} L(z_i)$
   - $\sum_{i=1}^{n} H(z_i)$

- $\sum_{i=1}^{n} B(z_i)$

2. Are the four functions defined in question 1 convex? Why or why not?
3. For each of the four functions defined in question 1, write a function that computes the penalty and a function that computes the gradient of the penalty.

   The penalty function should take a vector and return a single number for the functions based on $Q$ and $B$, and for the functions based on $H$ and $B$ it should take a vector and a scalar and return a single number.

   The gradient function should take a vector and return a vector for the functions based on $Q$ and $B$, and for the functions based on $H$ and $B$ it should take a vector and a scalar and return a vector.
4. On a fine grid between $-2$ and 2, plot the four functions as a function of a scalar argument $z$.
5. On a fine grid between $-2$ and 2, plot the derivatives of the four functions as a function of a scalar argument $z$.

## Submission parameters

Submit two files:

- A pdf giving your answers to the questions.
- An R script with the functions and other code you used.