Lecture 18
Theory of the Lasso

30 November 2015

Taylor B. Arnold
Yale Statistics
STAT 312/612

Class Notes

- XXXXX XXXXX XXXXX

Recall that sasso regression replaces the ℓ_2 penalty with an ℓ_1 penalty, and looks deceptively similar to the ridge regression:

$$\widehat{\beta}_\lambda = \arg\min_b \left\{ ||y - Xb||_2^2 + \lambda ||b||_1 \right\}$$

Where the ℓ_1-norm is defined as the sum of the absolute values of the vector's components:

$$||\beta||_1 = \sum_i |\beta_i|$$

There are three types of errors that we commonly are concerned with in lasso regression.

There are three types of errors that we commonly are concerned with in lasso regression. The prediction loss:

$$||X(\beta - \widehat{\beta})||_2^2$$

There are three types of errors that we commonly are concerned with in lasso regression. The prediction loss:

$$||X(\beta - \widehat{\beta})||_2^2$$

Parameter estimation:

$$||\beta - \widehat{\beta}||_2^2$$

There are three types of errors that we commonly are concerned with in lasso regression. The prediction loss:

$$||X(\beta - \widehat{\beta})||_2^2$$

Parameter estimation:

$$||\beta - \widehat{\beta}||_2^2$$

And model selection:

$$\text{supp}(\beta) = \text{supp}(\widehat{\beta})$$

There are three types of errors that we commonly are concerned with in lasso regression. The prediction loss:

$$||X(\beta - \widehat{\beta})||_2^2$$

Parameter estimation:

$$||\beta - \widehat{\beta}||_2^2$$

And model selection:

$$\text{supp}(\beta) = \text{supp}(\widehat{\beta})$$

Today we are going to focus on the first two types of estimation errors.

The type of results we want to establish may look different than you are used to seeing in more introductory courses.

The type of results we want to establish may look different than you are used to seeing in more introductory courses. What we will want to be able to construct is a set $\mathcal{A}$ such that:

$$\mathbb{P}\mathcal{A} = 1 - \epsilon$$

For some small $\epsilon > 0$, and

$$||X(\beta - \widehat{\beta})||_2^2 \leq \delta$$

Conditioned on being in event $\mathcal{A}$.

We want to establish some result about the lasso solution; where to begin?

Let's start with the one relationship we know to be true between $\widehat{\beta}$ and β:

$$||y - X\widehat{\beta}||_2^2 + \lambda||\widehat{\beta}||_1 \leq ||y - X\beta||_2^2 + \lambda||\beta||_1$$

I'll start by expanding the loss terms and canceling on both sides:

$$||y - X\widehat{\beta}||_2^2 + \lambda||\widehat{\beta}||_1 \leq ||y - X\beta||_2^2 + \lambda||\beta||_1$$
$$y^t y + \widehat{\beta}^t X^t X\widehat{\beta} - 2y^t X\widehat{\beta} + \lambda||\widehat{\beta}||_1 \leq y^t y + b^t X^t Xb - 2y^t Xb + \lambda||b||_1$$
$$\widehat{\beta}^t X^t X\widehat{\beta} - 2y^t X\widehat{\beta} + \lambda||\widehat{\beta}||_1 \leq b^t X^t Xb - 2y^t Xb + \lambda||b||_1$$

The next step is grouping the terms on each side of the equation

$$\widehat{\beta}^t X^t X \widehat{\beta} - 2y^t X \widehat{\beta} + \lambda ||\widehat{\beta}||_1 \leq b^t X^t X b - 2y^t X b + \lambda ||b||_1$$
$$||X(\widehat{\beta} - b)||_2^2 \lambda ||\widehat{\beta}||_1 \leq 2y^t X(\beta - \widehat{\beta}) + \lambda ||b||_1$$