ETC3250 2018 - Lab 10

Advanced regression

Souhaib Ben Taieb

3 May 2018

Question 1

Read Section 6.2 of ISLR and do the exercise 2 in Section 6.8.

Question 2

Let $\mathbf{y} = (y_1, \dots, y_n)^{'} \in \mathbb{R}^n$, $\mathbf{X}^{'} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n] \in \mathbb{R}^{p \times n}$ with $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})' \in \mathbb{R}^p$, and consider the following optimization problem:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \left\{ \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \sum_{j=1}^{p} \left[(1-\alpha)\beta_j^2 + \alpha|\beta_j|\right] \right\} \tag{1}$$

where $\lambda \geq 0$, $\alpha \in [0, 1]$ and $\|\cdot\|_2$ is the L_2 norm.

Show how one can turn this into a lasso optimization problem. In other words, consider the following optimization problem:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \left\{ \left\|\mathbf{y}^{'} - \mathbf{X}^{'}\beta\right\|_2^2 + \lambda^{'} \sum_{j=1}^{p} |\beta_j| \right\}, \tag{2}$$

where $\lambda^{'} \geq 0$.

What are $\mathbf{y}^{'}$, $\mathbf{X}^{'}$ and $\lambda^{'}$ so that the two optimization problems are equivalent?

We will use ridge regression and the lasso to estimate the salary of various baseball players based on several predictor measurements. This data set is taken from the *ISLR* package. Download the file *hitters.Rdata* at https://github.com/bsouhaib/BA2018/blob/master/data/hitters.rdata. We will use the implementation of these algorithms available in the *glmnet* package.

Question 2

The *glmnet* function, by default, internally scales the predictor variables so that they will have standard deviation 1, before solving the ridge regression or lasso problems. Explain why such scaling is important in our application.

Question 3

Run the following commands:

```r
library(glmnet)

load("../../data/hitters.Rdata")
grid <-  10^seq(10, -2, length=100)
ridge.model <-  glmnet(x, y, lambda = grid, alpha = 0)
lasso.model <-  glmnet(x, y, lambda = grid, alpha = 1)
```

a. Using the help page of the *glmnet* function, briefly describe what the previous two lines are doing. In particular, what is *lambda* and *alpha*?

Question 4

a. For each model, verify that as *lambda* decreases, the value of the penalty term only increases. In other words, the squared L_2 and the L_1 norm of the coefficients only gets bigger as *lambda* decreases for ridge and the lasso, respectively. The plot should be on a log-log scale.

Question 5

a. For both ridge and the lasso, explain what happens to the coefficients for very small and very large values of *lambda*.

Question 6

a. For both ridge and lasso, produce a plot of the 5-fold cross-validation error curve as a function of *lambda*, with standard errors drawn, for both the ridge and lasso models. Determine the value of *lambda* that minimize the cross-validation error. You can use *cv.glmnet*.

Question 7

a. For both ridge and lasso, compute the estimates using (1) the best value of *lambda* you obtained in question 5, and (2) the model you fitted in question 2. You can use the *predict* function with *type* = *"coef"* to compute the estimates for a given model. How do the ridge estimates compare to those from the lasso?

Question 8

Suppose that you were coaching a young baseball player who wanted to strike it rich in the major leagues. What handful of attributes would you tell this player to focus on?