

ETC3250 2018 - Lab 10 solution

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.

- (a) iii is TRUE
- (b) iii is TRUE
- (c) ii is TRUE

Question 2

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

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

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:

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

where $\lambda' \geq 0$.

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

$$\left\{ \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \sum_{j=1}^p [(1 - \alpha)\beta_j^2 + \alpha|\beta_j|] \right\} \quad (3)$$

$$= \left\{ \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda(1 - \alpha) \|\beta\|_2^2 + \lambda\alpha \|\beta\|_1 \right\} \quad (4)$$

$$= \left\{ \left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{bmatrix} \mathbf{X} \\ \sqrt{\lambda(1 - \alpha)} \mathbf{I}_p \end{bmatrix} \beta \right\|_2^2 + \lambda\alpha \|\beta\|_1 \right\} \quad (5)$$

where $\mathbf{0}$ is a zero-vector of dimension p , and \mathbf{I}_p is a $p \times p$ identity matrix.

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.

The variables have different variances. If not scaled, variables with higher variance will be penalized more since they will have higher RSS.

Question 3

Run the following commands:

```
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)
```

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

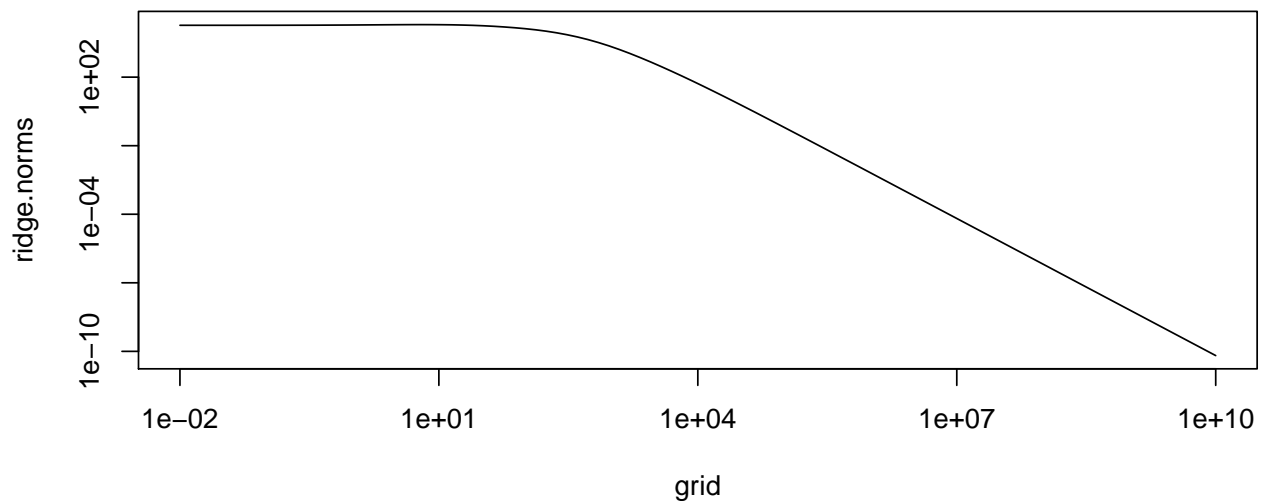
This compute ridge regression and lasso estimates, over the whole sequence of *lambda* values specified by *grid*. The flag *alpha* = 0 notifies glmnet to perform ridge regression, and *alpha*=1 notifies it to perform lasso regression

Question 4

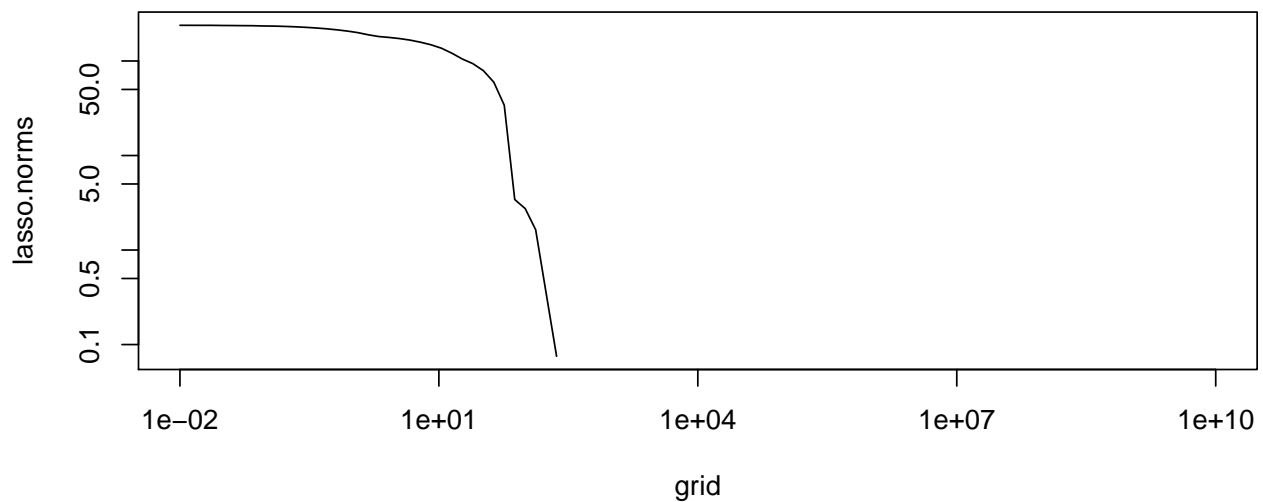
- For each model, verify that as *lambda* decreases, the value of the penalty term only increases. In other words, the squared L_2 and the L_1 norm of the coefficients only gets bigger as *lambda* decreases for ridge and the lasso, respectively. The plot should be on a log-log scale.

```
nlambda <- ncol(ridge.model$beta)
ridge.norms <- lasso.norms <- numeric(nlambda)

for(ilambda in seq(nlambda)){
  ridge.norms[ilambda] <- sum(ridge.model$beta[, ilambda]^2)
  lasso.norms[ilambda] <- sum(abs(lasso.model$beta[, ilambda]))
}
plot(grid, ridge.norms, type = 'l', log="xy")
```



```
plot(grid, lasso.norms, type = 'l', log="xy")
```



Question 5

- For both ridge and the lasso, explain what happens to the coefficients for very small and very large values of λ .

For a very small value of λ , the estimates are very close to the least squares estimate. For a very large value of λ , the estimates approach 0 in all components (except the intercept, which is not penalized by default).

Question 6

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

```
cv.ridge <- cv.glmnet(x, y, nfolds = 5, alpha = 0, lambda = grid)
cv.lasso <- cv.glmnet(x, y, nfolds = 5, alpha = 1, lambda = grid)
```

```
idbest_ridge <- match(cv.ridge$lambda.min, grid)
idbest_lasso <- match(cv.lasso$lambda.min, grid)
```

Question 7

- a. For both ridge and lasso, compute the estimates using (1) the best value of λ 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?

```
best_ridge <- predict(ridge.model, type = "coef")[, idbest_ridge]
best_lasso <- predict(lasso.model, type = "coef")[, idbest_lasso]

print(best_ridge)
# (Intercept)      AtBat      Hits      HmRun      Runs
# 152.52893912 -1.67284639  5.86106113  0.92327754 -0.51634440
#      RBI      Walks      Years      CAtBat      CHits
# -0.01416825  5.36007853 -9.93306467 -0.06593201  0.20771259
#      CHmRun      CRuns      CRBI      CWalks      LeagueN
#  0.69782050  0.75594735  0.38243059 -0.63089682  61.69452203
# DivisionW      PutOuts      Assists      Errors      NewLeagueN
# -122.22631375  0.28016906  0.29979472 -3.73139449 -27.96471550
print(best_lasso)
# (Intercept)      AtBat      Hits      HmRun      Runs
# 146.83733357 -1.84795456  6.36025683  0.53321464 -0.20796869
#      RBI      Walks      Years      CAtBat      CHits
#  0.00000000  5.27692705 -9.37605656 -0.02686293  0.00000000
#      CHmRun      CRuns      CRBI      CWalks      LeagueN
#  0.44385268  0.88762007  0.46369992 -0.66338590  38.73155632
# DivisionW      PutOuts      Assists      Errors      NewLeagueN
# -117.98879627  0.28057238  0.24942119 -2.59606200 -3.40537234
```

We can clearly see that the lasso estimate is sparser than the ridge estimate with 6 coefficients equal to zero.

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?

Look at the lasso coefficients that are non-zero with large positive values.