LASSO

Sean Hellingman

https://github.com/shellingman

April 26, 2023

Presentation Overview

1. Introduction
2. Theory
3. Example
4. Conclusions

LASSO

- Least Absolute Shrinkage and Selection Operator (LASSO)
- Regularization method for model selection
- The LASSO solution can yield a reduction in variance at the expense of a small increase in bias

Formulation

The LASSO coefficients, $\hat{\beta}_\lambda^L$ minimize the quantity

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = \text{RSS} + \lambda\sum_{j=1}^{p}|\beta_j|. \tag{1}$$

The ℓ_1 norm of a coefficient vector β is given by $||\beta||_1 = \sum|\beta_j|$.

Some Properties

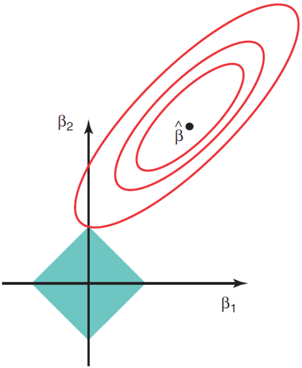
- LASSO shrinks the coefficient estimates towards zero
- With a sufficiently large λ some of the coefficient estimates shrink to be exactly zero
- LASSO performs variable selection
- Choice of λ is important and is often done through cross-validation

Another Formulation

$$\min_{\beta}\left\{\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2\right\} \text{ subject to } \sum_{j=1}^{p}|\beta_j|\leq s. \qquad (2)$$

For every value of λ, there is some s such that (2) will give the same LASSO coefficient estimates.

Variable Selection Property



- Two parameters ($p = 2$)
- $\hat{\beta}$: OLS solution
- Blue rectangle: $|\beta_1| + |\beta_2| \leq s$
- Red ellipses: regions of constant RSS

Comments on LASSO

- When $\lambda = 0$, OLS estimates
- Reduction in variance at the expense of a small increase in bias
- Can be a useful tool for model selection

Problem

Want to model *horsepower* (hp) dependent on *Miles/gallon* (mpg), *weight* (wt), *rear axle ratio* (drat), and *1/4 mile time* (qsec) using data from the mtcars dataset.

Data Exploration

Model	hp	mpg	wt	drat	qsec
Mazda RX4	110	21	2.620	3.900	16.460
Mazda RX4 Wag	110	21	2.875	3.900	17.020
Datsun 710	93	22.800	2.320	3.850	18.610
Hornet 4 Drive	110	21.400	3.215	3.080	19.440
Hornet Sportabout	175	18.700	3.440	3.150	17.020

Scaling

Code used to scale the variables:

```
####Min-Max Scaling [0-1] Function####
norm_minmax <- function(x){
  (x- min(x)) /(max(x)-min(x))
}

####Define Model####
#Response Variable
y <- Cars$hp #horse power

#Predictors
x <- data.matrix(Cars[,c('mpg', 'wt', 'drat', 'qsec')])
#Scaled#
y <- norm_minmax(y)
x <- norm_minmax(x)
```

Example Models

Variable	OLS	$\lambda = 0.001$	$\lambda = 5$
mpg	-0.329	-0.325	0.000
wt	2.980	2.886	-
drat	0.551	0.335	-
qsec	-2.375	-2.365	-
Intercept	1.529	1.541	0.335

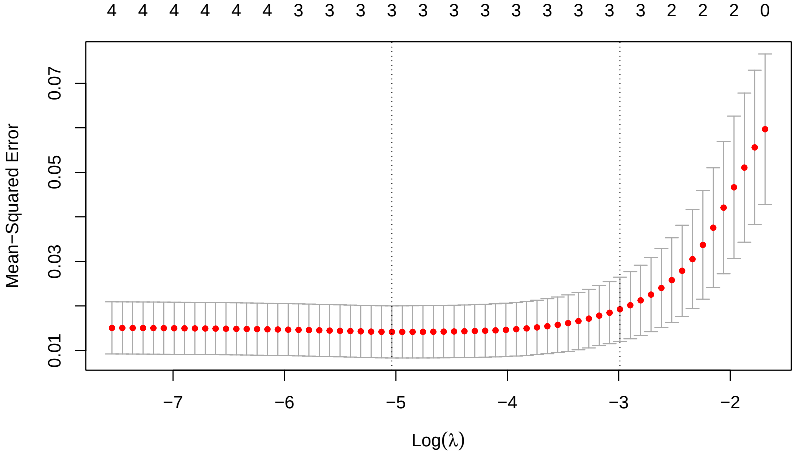
```
##OLS Model
lm <- lm(y~x)

##Small lambda
#alpha = 1 for LASSO
m1 <- glmnet(x,y,alpha=1, lambda = 0.001)

##Large lambda
m2 <- glmnet(x,y,alpha=1, lambda = 5)
```

Cross Validation

- Find optimal lambda value that minimizes test mean squared error (MSE)
- Perform 10-fold cross-validation to find optimal lambda value
- Functionality in the *glmnet* R package

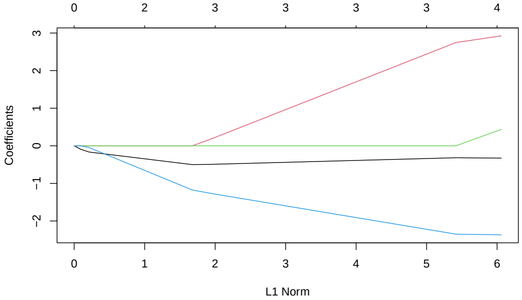
```
cv1 <- cv.glmnet(x, y, nfolds = 10, alpha = 1)

best_lambda <- cv1$lambda.min
best_lambda
[1] 0.006498586
```

Plot of the Cross-validation Curve



Model Comparisons

Variable	OLS	$\lambda = 0.001$	$\lambda = 5$	$\lambda = 0.006498586$
mpg	-0.329	-0.325	0.000	-0.330
wt	2.980	2.886	-	2.574
drat	0.551	0.335	-	-
qsec	-2.375	-2.365	-	-2.278
Intercept	1.529	1.541	0.335	1.537



LASSO Conclusions

- Penalizes β values by *shrinking* them to zero
- Useful for variable selection
- Choice of λ is important and is often done through cross-validation
- Related Topics:
  - Ridge regression
  - Elastic net regularization
  - Methods for dimension reduction

References & Resources

1. Hastie, T., Qian, J., Tay, K. (2021). An Introduction to glmnet. *CRAN R Repositary.*
2. James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

Link to code used: github.com/shellingman/LASSO-Presentation