GOV 51 Section

Week 5: Penalized Regression

Pranav Moudgalya

Harvard College

Housekeeping

- Problem set 1 is complete now!

Housekeeping

- Problem set 1 is complete now!
- We are grading, will return grades well before the midterm

Housekeeping

- Problem set 1 is complete now!
- We are grading, will return grades well before the midterm
- Midterm (conceptual + coding) coming up; I will release a list of study topics to help you prepare

Housekeeping

- Problem set 1 is complete now!
- We are grading, will return grades well before the midterm
- Midterm (conceptual + coding) coming up; I will release a list of study topics to help you prepare

Agenda

- Lasso

Agenda

- Lasso
- Lasso Proof

Agenda

- Lasso
- Lasso Proof
- Implementation of Lasso

Agenda

- Lasso
- Lasso Proof
- Implementation of Lasso

Issues with Regression

- Regression allows us to estimate the relationship between two variables

Issues with Regression

- Regression allows us to estimate the relationship between two variables
- **Problem:** How do we choose variables?

Issues with Regression

- Regression allows us to estimate the relationship between two variables
- **Problem:** How do we choose variables?

1. Intuition - perhaps, we have prior knowledge about the relationship and think certain things affect outcomes

Issues with Regression

- Regression allows us to estimate the relationship between two variables
- **Problem:** How do we choose variables?

1. Intuition - perhaps, we have prior knowledge about the relationship and think certain things affect outcomes
2. Literature - perhaps, the scholarship tells us that there are things that are important

Issues with Regression

- Regression allows us to estimate the relationship between two variables
- **Problem:** How do we choose variables?

1. Intuition - perhaps, we have prior knowledge about the relationship and think certain things affect outcomes
2. Literature - perhaps, the scholarship tells us that there are things that are important

- **Problem 2:** What if we have lots and lots of variables?

Issues with Regression

- Regression allows us to estimate the relationship between two variables
- **Problem:** How do we choose variables?

1. Intuition - perhaps, we have prior knowledge about the relationship and think certain things affect outcomes
2. Literature - perhaps, the scholarship tells us that there are things that are important

- **Problem 2:** What if we have lots and lots of variables?
  - Intuition and the literature cannot really help us if we have 4500 variables

Lasso

- **Lasso** - Least absolute shrinkage and selection operator

Lasso

- **Lasso** - Least absolute shrinkage and selection operator
- Similar to ordinary least squares regression by using a loss function to calculate $\widehat{\beta}$

Lasso

- **Lasso** - Least absolute shrinkage and selection operator
- Similar to ordinary least squares regression by using a loss function to calculate $\widehat{\beta}$

$$\widehat{\beta} = \underset{\widetilde{\beta}}{\text{argmin}} \frac{1}{2} \sum_{i=1}^{N} (Y_i - \widetilde{\beta} X)^2 + \boxed{\lambda |\widetilde{\beta}|} \tag{1}$$

- Short version: Add a penalty term to our loss function to only keep relevant covariates in our model
  - Generally referred to shrinkage and regularization

Lasso

- **Lasso** - Least absolute shrinkage and selection operator
- Similar to ordinary least squares regression by using a loss function to calculate $\widehat{\beta}$

$$\widehat{\beta} = \underset{\widetilde{\beta}}{\text{argmin}} \frac{1}{2} \sum_{i=1}^{N} (Y_i - \widetilde{\beta} X)^2 + \boxed{\lambda |\widetilde{\beta}|} \tag{1}$$

- Short version: Add a penalty term to our loss function to only keep relevant covariates in our model
  - Generally referred to shrinkage and regularization
- Advantages of lasso is the ease of which to interpret its penalty: keep or drop

Lasso

- **Lasso** - Least absolute shrinkage and selection operator
- Similar to ordinary least squares regression by using a loss function to calculate $\widehat{\beta}$

$$\widehat{\beta} = \underset{\widetilde{\beta}}{\text{argmin}} \frac{1}{2} \sum_{i=1}^{N} (Y_i - \widetilde{\beta} X)^2 + \boxed{\lambda |\widetilde{\beta}|} \qquad (1)$$

- Short version: Add a penalty term to our loss function to only keep relevant covariates in our model
  - Generally referred to shrinkage and regularization
- Advantages of lasso is the ease of which to interpret its penalty: keep or drop

$$\hat{\beta}_{\text{lasso}} = \left(\beta_{\text{OLS}} - \text{sgn}(\beta_{\text{OLS}}) \cdot \frac{\lambda}{B} \right) \mathbf{1} \left(|\beta_{\text{OLS}}| > \frac{\lambda}{B} \right)$$

Lasso Intuition

- Intuition from lecture: The size of the boxed term, $\lambda|\widetilde{\beta}|$, determines the penalty size, or how stringent lasso is
  - If a beta coefficient is insufficiently large, it goes to zero (e.g. dropped)
  - If lambda is too large, we exclude everything
  - What happens if lambda is zero?

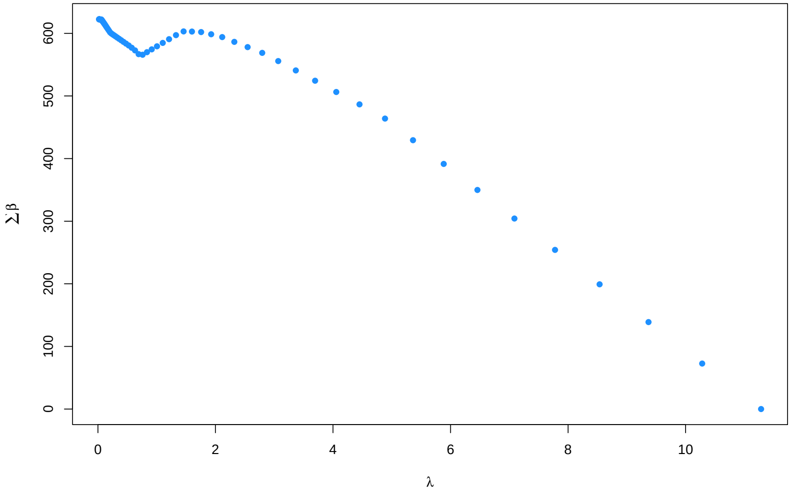
Toy Example

Variable names are respectively field goals, field goal percentage, 3 pointers made, total rebounds, minutes played, field goals attempted, 3 pointers attempted, 3 point percentage, 2 pointers made, 2 pointers attempted, 2 pointer percentage, free throws made, free throws attempted, free throw percentage, offensive rebounds, defensive rebounds, assists, steals, blocks, turnovers, personal fouls, points, conference dummy

Or. . .

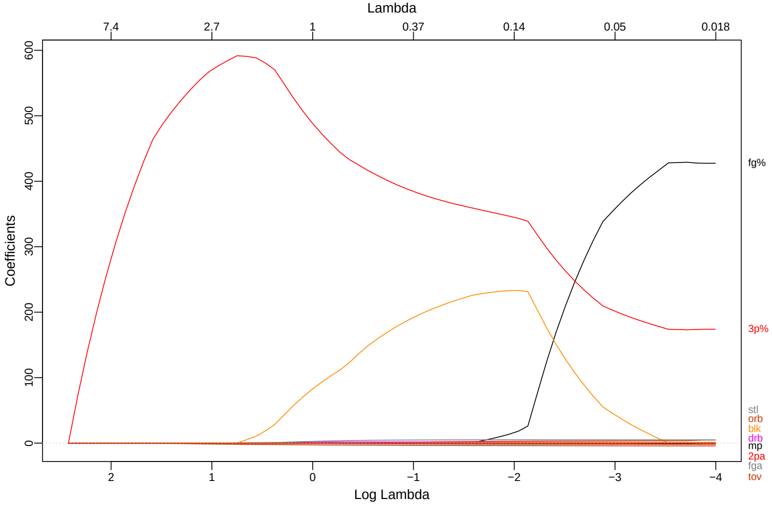
What factors contribute to winning basketball?

Lassoing some Lambdas



Lasso Coefficients

```
## Warning: package 'plotmo' was built under R version 4.3.
```



Issues with Lasso

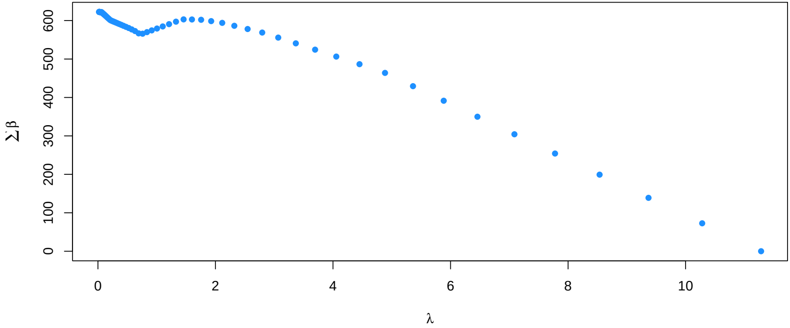
- Researchers can choose λ - theoretically, infinite models and infinite results, so opportunity to cherry pick
- OLS gives us **B**est **L**inear **U**nbiased **E**stimator with relatively limited assumptions (e.g. linearity, conditional mean = zero, independence of error)
  - In short, if all our OLS assumptions are met, regularization and shrinkage methods will bias our estimates
- Lasso does NOT have a closed-form solution because of matrix limitations
  - Ridge regression, another shrinkage/regularization method, does

Implementation of Lasso

```r
library(glmnet)

nba_data <- as.matrix(nba[, -c(1,25,26)])
winning <- as.matrix(nba[, c(25)])

lasso <- glmnet(x = nba_data,
                y = winning)

# sum absolute values of the betas
sum_beta <- colSums(abs(lasso$beta))
```

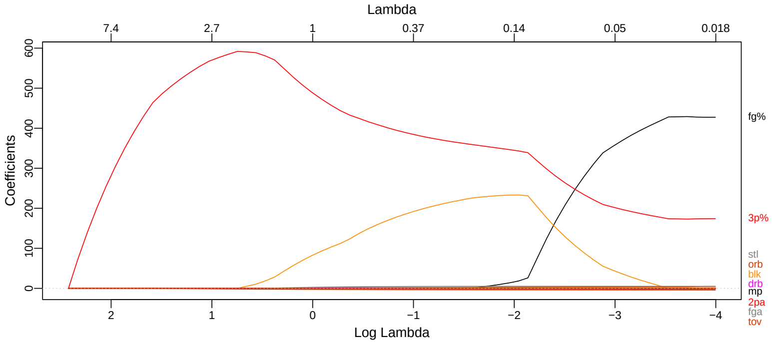
Sum of Betas against Lambda

```
# plot against values of lambda
plot(sum_beta ~ lasso$lambda,
     pch=16,
     col="dodgerblue1",
     ylab = expression(sum(beta)),
     xlab = expression(lambda))
```



Coefficient estimates against Lambda

```r
library(plotmo)
plot_glmnet(lasso)
```



What Lambda to Choose?

- Outside the scope of the class, but known as K-fold cross validation
  - In summary, minimizes the mean squared error to find the right lambda
- `glmnet` allows us to implement this calculation

```
lasso.cv <- cv.glmnet(x = as.matrix(nba_data),
                      y = winning)

lasso.cv$lambda.min
```

```
## [1] 0.2997425
```

Coefficient Estimates

```
coef(lasso,
     s = lasso.cv$lambda.min)
```

```
## 24 x 1 sparse Matrix of class "dgCMatrix"
##                       s1
## (Intercept) -127.4574159
## g                .
## mp               .
## fg               .
## fga           -1.6450347
## fg%              .
## 3p               .
## 3pa              .
## 3p%          373.9583001
## 2p               .
## 2pa           -0.4660265
## 2p%          206.0433940
## ft               .
## fta            0.3110077
## ft%              .
## orb            2.0083773
## drb              .
## trb            1.6439070
## ast              .
## stl            4.8840741
## blk            0.6392855
## tov           -3.5783640
## pf               .
## pts              .
```

Lasso Summary and Comments

- Lasso is an estimation technique, not an identification technique
  - Useful for obtaining more consistent estimates, as was matching, but not sufficient for causal identification
- Tackles two problems: high-dimensionality and overfitting
  - Better than ridge regression by making penalties easier to understand: keep or drop
  - However, can induce bias if OLS assumptions are met
- Other implementations of Lasso
  - Urminsky, Hansen, and Chernozhukov (2016) have a method known as double selection lasso, which attempts to control for differential treatment assignment