Penalized Regression

Section 5

Sima Biondi

Spring 2025

Gov 51: Data Analysis and Politics

Housekeeping

- Problem Set I: we're grading now

Housekeeping

- Problem Set I: we're grading now
- Project deadlines

Housekeeping

- Problem Set I: we're grading now
- Project deadlines
  - ↪ April 3: 1-pager that outlines the research question and introduce data source(s)

Housekeeping

- Problem Set I: we're grading now
- Project deadlines
  - ↪ April 3: 1-pager that outlines the research question and introduce data source(s)
  - ↪ April 10: 1-pager describing results

Roadmap

Last week: how to estimate

Roadmap

Last week: how to estimate

$\rightarrow$ Loss functions

Roadmap

Last week: how to estimate

→ Loss functions

→ Limits of OLS: not all regression are useful regressions

Roadmap

Last week: how to estimate

- → Loss functions
- → Limits of OLS: not all regression are useful regressions
- → Solution: penalized regressions

Roadmap

Last week: how to estimate

- $\rightarrow$ Loss functions
- $\rightarrow$ Limits of OLS: not all regression are useful regressions
- $\rightarrow$ Solution: penalized regressions

Roadmap

Last week: how to estimate

- $\rightarrow$ Loss functions
- $\rightarrow$ Limits of OLS: not all regression are useful regressions
- $\rightarrow$ Solution: penalized regressions

In this section, we look at the nuts and bolts of actual estimation with penalized regression

Agenda

- Lasso
- 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

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

Lasso

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

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

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

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
- The advantage of lasso is the ease of interpretation of its penalty: keep or drop

Lasso intuition

- Intuition from lecture: The size of the boxed term, $\lambda|\widetilde{\beta}|$, determines the penalty size, or how stringent lasso is

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)

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

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

What factors contribute to winning basketball?

Lassoing some lambdas

```
#load libraries: tidyverse, rvest, ggthemes
#load data
bballref <-
    read_html("https://www.basketball-reference.com/
    leagues/ NBA_2024.html")
standings <- bballref |> html_elements("#per_game-team
    tbody td") |>
    html_text()
statnames <- bballref |>
    html_elements("#per_game-team thead .center+
        .center") |>
    html_text()
nbadf <- matrix(standings, ncol = 24, byrow = TRUE) |>
    as_tibble()
standings2 <- bballref |>
  html_elements("#advanced-team tbody
      .right:nth-child(4) , #advanced-team tbody th+
      .left") |> html_text()
```

Lassoing some lambdas

```r
# data cleaning
wins <- matrix(standings2, ncol = 2, byrow = TRUE) |>
  as_tibble() |>
  rename("Team" = "V1",
         "Wins" = "V2")

nba <- nbadf |>
  left_join(wins,
            by = "Team") |>
  mutate(across(!"Team", as.numeric)) |>
  mutate(winpct = Wins/G) |>
  rename_with(tolower)
```

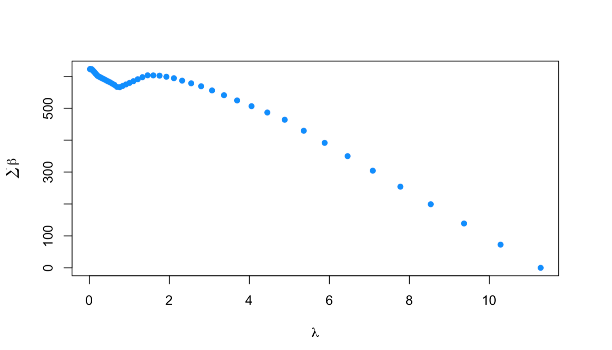
Lassoing some lambdas

```
# set up
library(glmnet)
set.seed(02138)
nba_data <- as.matrix(nba[, -c(1,25,26)])
winning <- as.matrix(nba[, c(25)])

## Lassoing some Lambdas
lasso <- glmnet(x = nba_data, y = winning)
# sum absolute values of the betas
sum_beta <- colSums(abs(lasso$beta))
# plot against values of lambda
plot(sum_beta ~ lasso$lambda, pch=16, col="blue",
      ylab = expression(sum(beta)),
      xlab = expression(lambda))
```

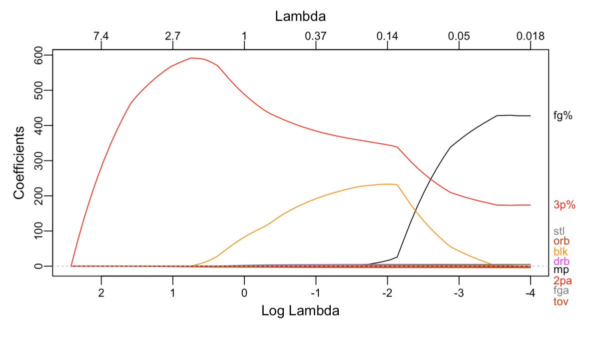
Lassoing some lambdas

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



Lasso coefficients

```
## Lasso Coefficients
library(plotmo)
plot_glmnet(lasso)
```



Lasso coefficients: what lambda to choose?

Outside the scope of the class, but known as K-fold cross validation

- Short summary: K-fold cross validation (CV) 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
```

Issues with Lasso

- Researchers can choose λ - theoretically, infinite models and infinite results, so opportunity to cherry pick

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

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
  - In short, if all our OLS assumptions are met, regularization and shrinkage methods will bias our estimates

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
    - 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

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
  - 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

Lasso summary and comments

- Lasso is an estimation technique, not an identification technique

Lasso summary and comments

- Lasso is an estimation technique, not an identification technique
    - Useful for obtaining more consistent estimates but not sufficient for causal identification

Lasso summary and comments

- Lasso is an estimation technique, not an identification technique
  - Useful for obtaining more consistent estimates but not sufficient for causal identification
- Tackles two problems: high-dimensionality and overfitting

Lasso summary and comments

- Lasso is an estimation technique, not an identification technique
    - Useful for obtaining more consistent estimates 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

Lasso summary and comments

- Lasso is an estimation technique, not an identification technique
    - Useful for obtaining more consistent estimates 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

Lasso summary and comments

- Lasso is an estimation technique, not an identification technique
    - Useful for obtaining more consistent estimates 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

Lasso summary and comments

- Lasso is an estimation technique, not an identification technique
    - Useful for obtaining more consistent estimates 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
    - Some methods attempt to control for differential treatment assignment