126 Data Project, Step 4

Sam Ream, Valeria Lopez, Skyler Yee

Introduction

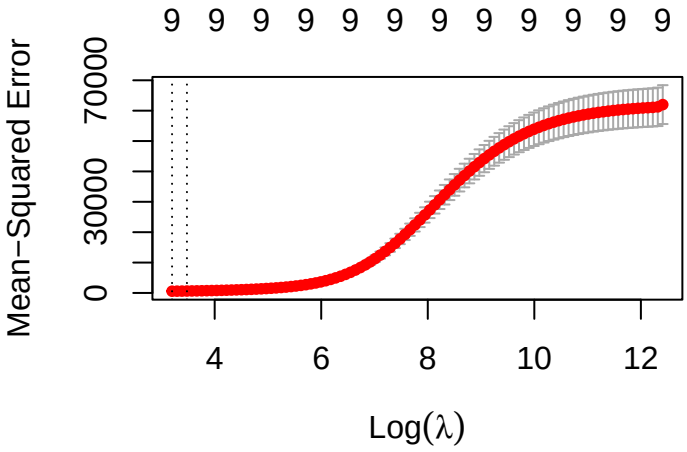
Using the "History of Baseball" data set, we analyzed how our predictors (singles, doubles, triples, home runs, walks, intentional walks, hit by pitches, stolen bases, BMI, and batting hand) affected the runs scores by individual players. We sampled player statistics randomly from games played between 2000-2015, which allowed us to get an accurate representation of the population of all players who played between 2000 and 2015. Using both Ridge Regression and LASSO, we shrunk the size of some predictors to obtain estimates with smaller variance for higher precision.

Ridge Regression

Summary Table of the Ridge Regression

```
##           Length Class     Mode
## a0        100    -none-    numeric
## beta      900    dgCMatrix S4
## df        100    -none-    numeric
## dim         2    -none-    numeric
## lambda    100    -none-    numeric
## dev.ratio 100    -none-    numeric
## nulldev     1    -none-    numeric
## npasses     1    -none-    numeric
## jerr        1    -none-    numeric
## offset      1    -none-    logical
## call        4    -none-    call
## nobs        1    -none-    numeric
```

Optimal Lambda




```
## [1] "The best lamda was found to be: "
```

```
## [1] 24.53741
```

Lasso Regression

Summary Table of the Lasso Regression

```
##           Length Class     Mode
## a0         71    -none-    numeric
## beta      639    dgCMatrix S4
## df         71    -none-    numeric
## dim         2    -none-    numeric
## lambda     71    -none-    numeric
## dev.ratio  71    -none-    numeric
## nulldev     1    -none-    numeric
## npasses     1    -none-    numeric
## jerr        1    -none-    numeric
## offset      1    -none-    logical
## call        4    -none-    call
## nobs        1    -none-    numeric
```

Optimal Lambda

Comparison of our Models

Investigation - Principle Component Analysis

Conclusion