

126 Data Project, Step 4

Sam Ream, Valeria Lopez, Skyler Yee

```
##
## Call:
## lm(formula = RUNS ~ HOME_RUNS + SINGLES + WALKS + STOLEN_BASES,
##     data = batting)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -124.140   -8.110   -0.528    6.956   88.759
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.781728    1.280636   -0.61    0.542
## HOME_RUNS      0.977261    0.031560   30.96 <2e-16 ***
## SINGLES        0.400909    0.007306   54.87 <2e-16 ***
## WALKS          0.268561    0.013396   20.05 <2e-16 ***
## STOLEN_BASES  0.490476    0.028537   17.19 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.6 on 495 degrees of freedom
## Multiple R-squared:  0.9925, Adjusted R-squared:  0.9925
## F-statistic: 1.642e+04 on 4 and 495 DF,  p-value: < 2.2e-16
```

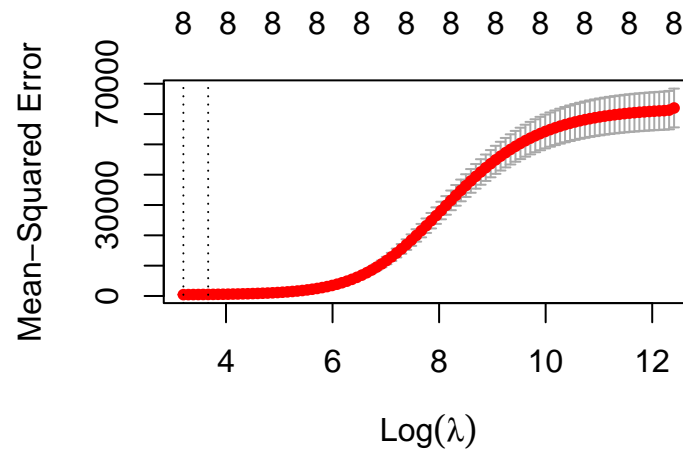
Ridge Regression

Fit the Ridge Regression Model

```
##           Length Class      Mode
## a0         100   -none-   numeric
## beta        800 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
```

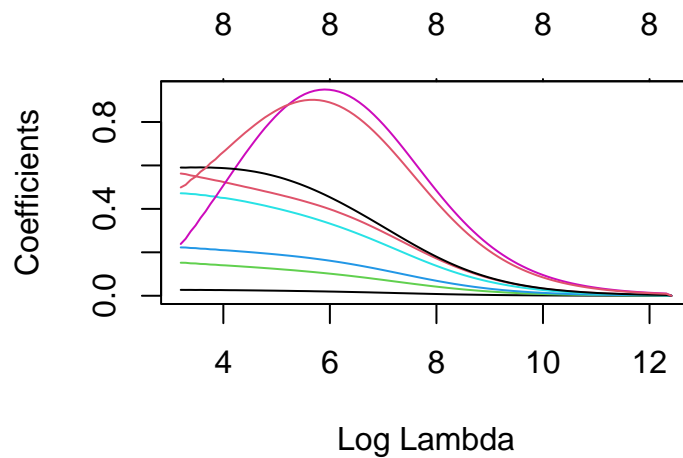
Choose an Optimal Value for Lambda

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



The lambda value that minimizes the test MSE is 24.53741.

Analyze Final Model



```
## [1] 0.9931578
```

The R-Squared is 0.9914, so the best model explains 99.14% of the variation in the response values.

LASSO

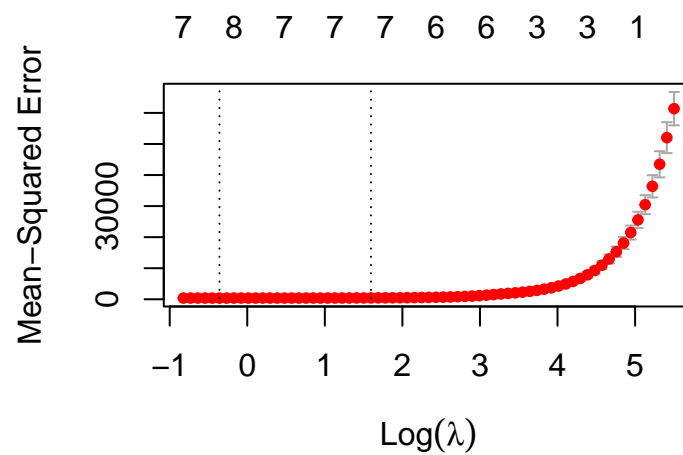
Fit the Lasso Regression Model

```
##          Length Class      Mode
## a0         69   -none-   numeric
## beta       552 dgMatrix S4
## df          69   -none-   numeric
```

```
## dim      2    -none-   numeric
## lambda   69    -none-   numeric
## dev.ratio 69    -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
```

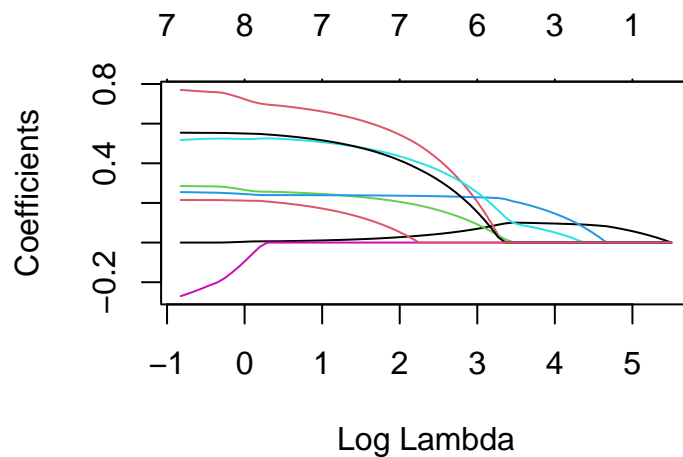
Choose an Optimal Value for Lambda

```
## [1] 0.6988341
```



The lambda value that minimizes the test MSE is 0.364.

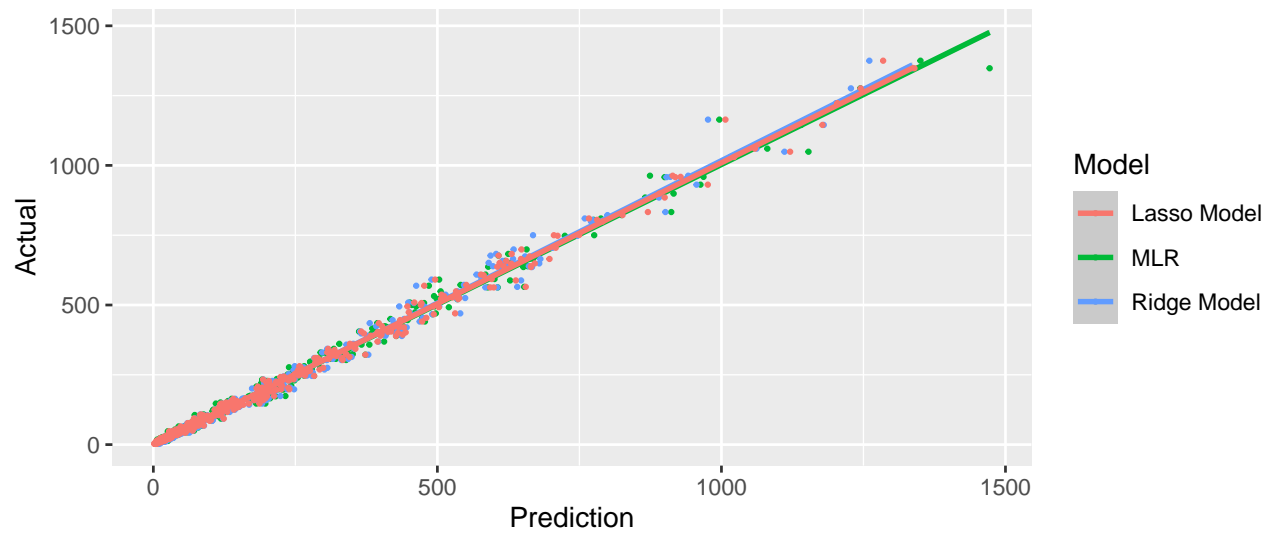
Analyze Final Model



```
## [1] 0.9947245
```

The R-Squared is 0.9936, so the best model explains 99.36% of the variation in the response values.

```
## [1] 0.6988341
```



```
## [1] 6571.305
```

```
## [1] 5954.884
```

```
## [1] 6407.784
```