

Lasso_Regression.R

vijaykalmath

2022-01-04

```
# Ridge Regression and Lasso Regression

# GLMNET package is used for Lasso and Ridge Regression

# set seed for reproducibility
set.seed(1)

library(ISLR) # For the data
library(glmnet)

## Loading required package: Matrix

## Loaded glmnet 4.1-3

# Omitting NA
Hitters = na.omit(Hitters)

# Model Matrix produces matrix with 19 predictors but also transforms any qualitative variables into dummy variables
x = model.matrix(Salary~.,data=Hitters)[,-1]

y = Hitters$Salary

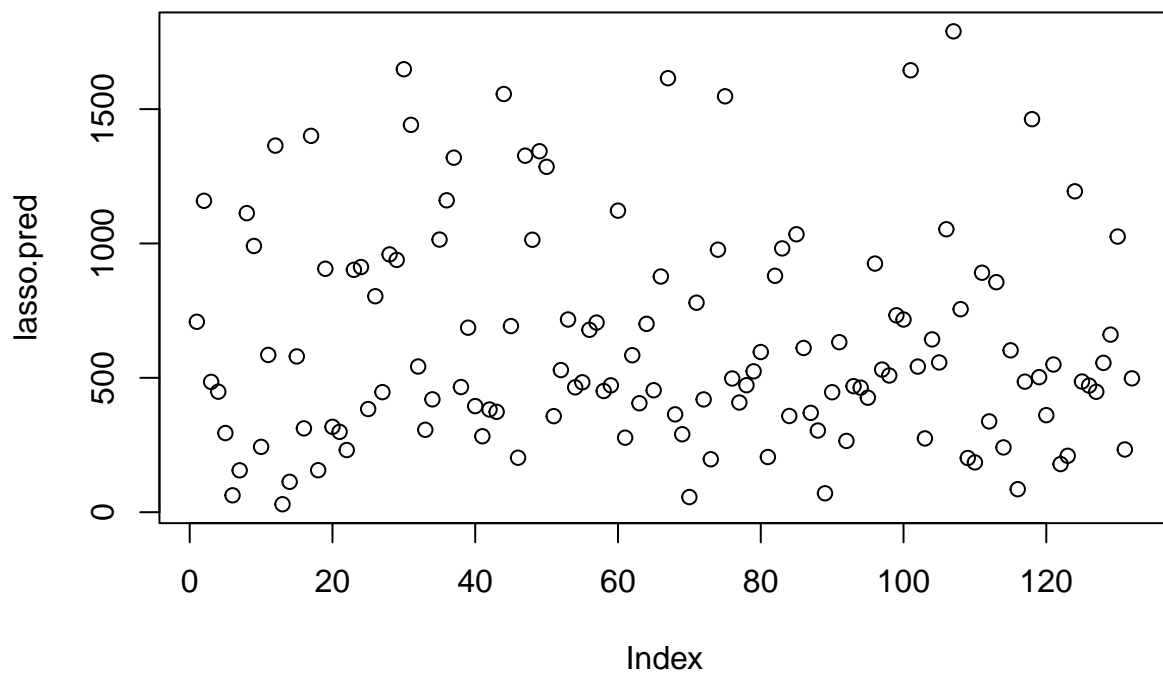
# Creating Exponential space gridspace for lambda or regularization factor
grid= 10^seq(10,-2,length=100)

# Define Train and test set
train <- sample(1:nrow(x), nrow(x) / 2)
test <- (-train)

# Alpha = 0 -> Ridge Regression , Alpha = 1 -> Lasso Regression
lasso.mod <- glmnet(x[train, ], y[train], alpha = 1,
                    lambda = grid)

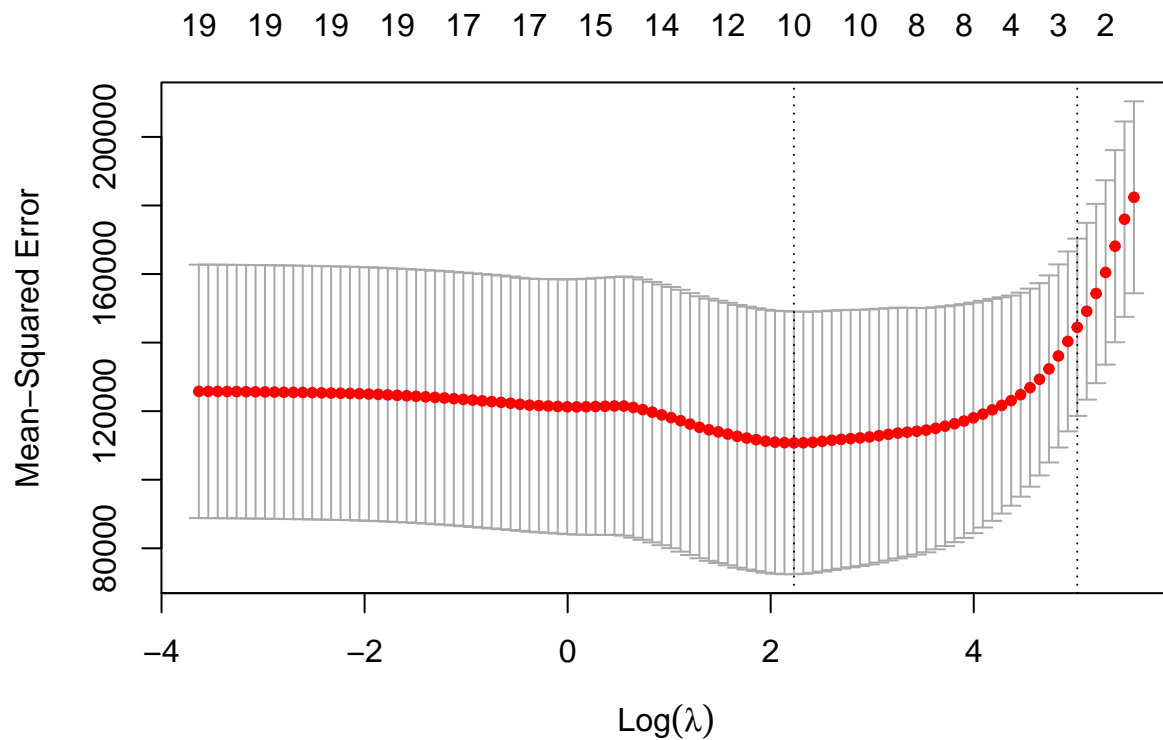
lasso.pred <- predict(lasso.mod, s = 4, newx = x[test, ])

plot(lasso.pred)
```



```
set.seed(1)

# Gridsearch for best lambda
cv.out = cv.glmnet(x[train,],y[train],alpha=1)
plot(cv.out)
```



```
bestlam = cv.out$lambda.min
```

```
lasso.pred = predict(lasso.mod,s=bestlam,newx = x[test,])
```

```
mean((lasso.pred - y[test])^2)
```

```
## [1] 516156.5
```

```
out = glmnet(x,y,alpha=1,lambda = grid)
```

```
lasso.coef = predict(out,type="coefficients",s=bestlam)[1:20,]
```

```
lasso.coef
```

```
##      (Intercept)      AtBat      Hits      HmRun      Runs
##      1.27479059   -0.05497143   2.18034583   0.00000000   0.00000000
##           RBI           Walks           Years      CAtBat      CHits
##      0.00000000   2.29192406  -0.33806109   0.00000000   0.00000000
##      CHmRun      CRuns      CRBI      CWalks      LeagueN
##      0.02825013   0.21628385   0.41712537   0.00000000   20.28615023
##      DivisionW      PutOuts      Assists      Errors      NewLeagueN
##     -116.16755870   0.23752385   0.00000000  -0.85629148   0.00000000
```

```
lasso.coef[lasso.coef != 0 ]
```

```
##      (Intercept)      AtBat      Hits      Walks      Years
##      1.27479059   -0.05497143   2.18034583   2.29192406   -0.33806109
##      CHmRun      CRuns      CRBI      LeagueN      DivisionW
##      0.02825013   0.21628385   0.41712537   20.28615023  -116.16755870
##      PutOuts      Errors
##      0.23752385   -0.85629148
```