Homework 6

Math 189
Spring 2023
Sophia Terenik
Deena Pederson
Gabriel Brewster

2023-05-12

Conceptual Problems

1. Because ridge regression is a shrinkage method, we would find it useful to utilize this method when there are
2.

Application Problems

3a.

```
library(ISLR2)
library(boot)
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-7
```

```
data("Hitters")
```

```
hitters <- subset(Hitters, select = -c(League, Division, NewLeague))
hitters <- hitters[complete.cases(hitters), ]
dim(hitters)
```

```
## [1] 263  17
```

3b.

```
set.seed(123)
train_index <- sample(1:nrow(hitters), round(0.8 * nrow(hitters)))
train <- hitters[train_index, ]
test <- hitters[-train_index, ]
```

3c.

```
set.seed(123)
lm_model <- lm(Salary ~ ., data = train)
summary(lm_model)
```

```
##
## Call:
```

```
## lm(formula = Salary ~ ., data = train)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -789.6 -179.9  -36.2  140.2 1915.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 174.45475   94.28494   1.850  0.06580 .
## AtBat        -1.81331    0.76001  -2.386  0.01800 *
## Hits          5.12107    2.92236   1.752  0.08130 .
## HmRun        -5.02689    6.92000  -0.726  0.46846
## Runs         -1.14996    3.44698  -0.334  0.73903
## RBI           2.78032    3.00883   0.924  0.35661
## Walks         6.37394    2.13035   2.992  0.00313 **
## Years       -13.27666   14.31453  -0.927  0.35483
## CAtBat       -0.28595    0.15817  -1.808  0.07218 .
## CHits         0.75525    0.81379   0.928  0.35453
## CHmRun        1.57799    1.88372   0.838  0.40323
## CRuns         1.37818    0.90003   1.531  0.12735
## CRBI          0.22551    0.84573   0.267  0.79002
## CWalks       -0.63479    0.39721  -1.598  0.11165
## PutOuts       0.21738    0.08754   2.483  0.01388 *
## Assists       0.51573    0.25583   2.016  0.04520 *
## Errors       -6.85930    4.92600  -1.392  0.16538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 316.8 on 193 degrees of freedom
## Multiple R-squared:  0.5819, Adjusted R-squared:  0.5472
## F-statistic: 16.79 on 16 and 193 DF,  p-value: < 2.2e-16
```

3d.

```
set.seed(123)
y_test <- test$Salary
y_pred <- predict(lm_model, newdata = test)
RMSE <- sqrt(mean((y_test - y_pred)^2))
RMSE
```

```
## [1] 373.854
```

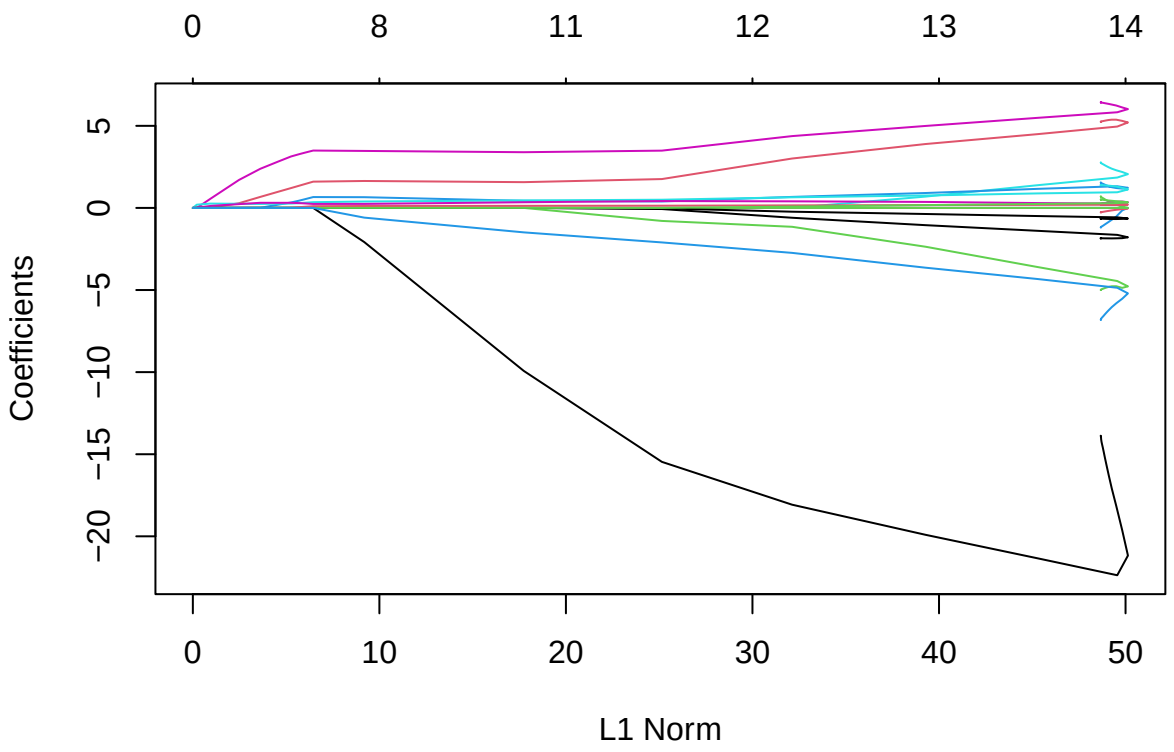
3e.

4a.

```
x <- model.matrix(Salary ~ ., hitters)[, -1]
y <- hitters$Salary
```

```
grid <- 10^seq(10, -2, length = 100)
lasso.mod <- glmnet(x[train_index, ], y[train_index], alpha = 1, lambda = grid)
plot(lasso.mod)
```

```
## Warning in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
## collapsing to unique 'x' values
```



4b.

```
set.seed(123)
cv.out <- cv.glmnet(x[train_index, ], y[train_index], alpha = 1)
bestlam <- cv.out$lambda.min
out <- glmnet(x, y, alpha = 1, lambda = grid)
lasso.coef <- predict(out, type = "coefficients", s = bestlam)[1:17, ]
lasso.coef
```

```
##  (Intercept)        AtBat         Hits        HmRun         Runs          RBI
## -57.89789801   0.00000000   2.01966953   0.00000000   0.00000000   0.00000000
##        Walks        Years       CAtBat        CHits       CHmRun        CRuns
##   2.31756395   0.00000000   0.00000000   0.00000000   0.07145521   0.25558599
##         CRBI       CWalks      PutOuts      Assists       Errors
##   0.36007290   0.00000000   0.23686740   0.00000000  -0.50557651
```

4c.

```
bestlam <- cv.out$lambda.min
lasso.pred <- predict(lasso.mod, s = bestlam, newx = x[-train_index, ])
mean((lasso.pred - y_test)^2)
```

```
## [1] 126670.1
```

5a.

```
grid <- 10^seq(10, -2, length = 100)
ridge.mod <- glmnet(x, y, alpha = 0, lambda = grid)
```

```
ridge.mod$lambda[50]
```

```
## [1] 11497.57
```

5b.

```
coef(ridge.mod)[, 50]
```

```
##   (Intercept)         AtBat          Hits         HmRun          Runs
## 404.259816123   0.036988305   0.138354280   0.524805207   0.231107356
##           RBI         Walks         Years        CAtBat         CHits
##   0.240164596   0.289952867   1.108042412   0.003132353   0.011656495
##        CHmRun         CRuns          CRBI        CWalks       PutOuts
##   0.087564108   0.023393054   0.024142296   0.025034383   0.016492859
##       Assists        Errors
##   0.002627915  -0.020458095
```

5c.

```
sqrt(sum(coef(ridge.mod)[-1, 50]^2))
```

```
## [1] 1.314959
```