Ridge Regression and the Lasso

Contents

1 Ridge Regression

```
library(ISLR2)
Hitters = na.omit(Hitters)
```

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

```
head(x)
```

```
##                   AtBat Hits HmRun Runs RBI Walks Years CAtBat CHits CHmRun
## -Alan Ashby         315   81     7   24  38    39    14   3449   835     69
## -Alvin Davis        479  130    18   66  72    76     3   1624   457     63
## -Andre Dawson       496  141    20   65  78    37    11   5628  1575    225
## -Andres Galarraga   321   87    10   39  42    30     2    396   101     12
## -Alfredo Griffin    594  169     4   74  51    35    11   4408  1133     19
## -Al Newman          185   37     1   23   8    21     2    214    42      1
##                   CRuns CRBI CWalks LeagueN DivisionW PutOuts Assists Errors
## -Alan Ashby         321  414    375       1         1     632      43     10
## -Alvin Davis        224  266    263       0         1     880      82     14
## -Andre Dawson       828  838    354       1         0     200      11      3
## -Andres Galarraga    48   46     33       1         0     805      40      4
## -Alfredo Griffin    501  336    194       0         1     282     421     25
## -Al Newman           30    9     24       1         0      76     127      7
##                   NewLeagueN
## -Alan Ashby                1
## -Alvin Davis               0
## -Andre Dawson              1
## -Andres Galarraga          1
## -Alfredo Griffin           0
## -Al Newman                 0
```

```
dim(x)
```

```
## [1] 263  19
```

```
library(glmnet)
```

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

```
## Loaded glmnet 4.1-4
```

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

```
dim(coef(ridge.mod))
```

```
## [1]  20 100
```

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

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

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

```
##   (Intercept)         AtBat          Hits         HmRun          Runs
## 407.356050200   0.036957182   0.138180344   0.524629976   0.230701523
##           RBI         Walks         Years        CAtBat         CHits
##   0.239841459   0.289618741   1.107702929   0.003131815   0.011653637
##        CHmRun         CRuns          CRBI        CWalks       LeagueN
##   0.087545670   0.023379882   0.024138320   0.025015421   0.085028114
##     DivisionW       PutOuts       Assists        Errors    NewLeagueN
##  -6.215440973   0.016482577   0.002612988  -0.020502690   0.301433531
```

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

```
## [1] 6.360612
```

```
ridge.mod$lambda[60]
```

```
## [1] 705.4802
```

```
coef(ridge.mod)[, 60]
```

```
##  (Intercept)         AtBat          Hits         HmRun          Runs           RBI
##  54.32519950   0.11211115   0.65622409   1.17980910   0.93769713   0.84718546
##        Walks         Years        CAtBat         CHits        CHmRun         CRuns
##   1.31987948   2.59640425   0.01083413   0.04674557   0.33777318   0.09355528
##         CRBI        CWalks       LeagueN     DivisionW       PutOuts       Assists
##   0.09780402   0.07189612  13.68370191 -54.65877750   0.11852289   0.01606037
##       Errors    NewLeagueN
##  -0.70358655   8.61181213
```

```
sqrt(sum(coef(ridge.mod)[-1, 60]^2))
```

```
## [1] 57.11001
```

```
# We can use predict() function to obtain the ridge regression coeff for a new value of \lambda=50:
predict(ridge.mod, s=50, type = "coefficients")[1:20,]
```

```
##   (Intercept)         AtBat          Hits         HmRun          Runs
##  4.876610e+01 -3.580999e-01  1.969359e+00 -1.278248e+00  1.145892e+00
##           RBI         Walks         Years        CAtBat         CHits
##  8.038292e-01  2.716186e+00 -6.218319e+00  5.447837e-03  1.064895e-01
##        CHmRun         CRuns          CRBI        CWalks       LeagueN
##  6.244860e-01  2.214985e-01  2.186914e-01 -1.500245e-01  4.592589e+01
##     DivisionW       PutOuts       Assists        Errors    NewLeagueN
## -1.182011e+02  2.502322e-01  1.215665e-01 -3.278600e+00 -9.496680e+00
```

```
predict(ridge.mod, s=0.01, type = "coefficients")[1:20, ]
```

```
##   (Intercept)         AtBat          Hits         HmRun          Runs
##  164.11321606   -1.97386151    7.37772270    3.93660219   -2.19873625
##           RBI         Walks         Years        CAtBat         CHits
```

```
##   -0.91623008    6.20037718   -3.71403424   -0.17510063    0.21132772
##        CHmRun         CRuns          CRBI        CWalks       LeagueN
##    0.05629004    1.36605490    0.70965516   -0.79582173   63.40493257
##     DivisionW       PutOuts       Assists        Errors    NewLeagueN
## -117.08243713    0.28202541    0.37318482   -3.42400281  -25.99081928
```

```
coef(ridge.mod)[, 100]
```

```
##   (Intercept)         AtBat          Hits         HmRun          Runs
##  164.11321606   -1.97386151    7.37772270    3.93660219   -2.19873625
##           RBI         Walks         Years        CAtBat         CHits
##   -0.91623008    6.20037718   -3.71403424   -0.17510063    0.21132772
##        CHmRun         CRuns          CRBI        CWalks       LeagueN
##    0.05629004    1.36605490    0.70965516   -0.79582173   63.40493257
##     DivisionW       PutOuts       Assists        Errors    NewLeagueN
## -117.08243713    0.28202541    0.37318482   -3.42400281  -25.99081928
```

```
set.seed(1)
train <- sample(1: nrow(x), nrow(x) / 2)
test <- (-train)
y.test <- y[test]
```

```
ridge.mod <- glmnet(x[train, ], y[train], alpha=0,
                    lambda=grid, thresh = 1e-12)
ridge.pred <- predict(ridge.mod, s = 4, newx = x[test, ])
mean((ridge.pred - y.test)^2)
```

```
## [1] 142199.2
```