Best-SubsetSelection.R

vijaykalmath

2022-01-04

```
# Subset Selection Methods - Best Subset Selection.

library(ISLR)

names(Hitters)
```

```
##  [1] "AtBat"     "Hits"      "HmRun"     "Runs"      "RBI"       "Walks"
##  [7] "Years"     "CAtBat"    "CHits"     "CHmRun"    "CRuns"     "CRBI"
## [13] "CWalks"    "League"    "Division"  "PutOuts"   "Assists"   "Errors"
## [19] "Salary"    "NewLeague"
```

```
dim(Hitters)
```

```
## [1] 322  20
```

```
# Removing NA from Hitters

Hitters = na.omit(Hitters)

# Dimensions after removing NA.
dim(Hitters)
```

```
## [1] 263  20
```

```
# the Regsubsets function is part of the leaps library

library(leaps)


# calling regsubsets

regfit.full = regsubsets(Salary~.,Hitters)


summary(regfit.full)
```

```
## Subset selection object
## Call: regsubsets.formula(Salary ~ ., Hitters)
## 19 Variables  (and intercept)
```

```
##            Forced in Forced out
## AtBat          FALSE      FALSE
## Hits           FALSE      FALSE
## HmRun          FALSE      FALSE
## Runs           FALSE      FALSE
## RBI            FALSE      FALSE
## Walks          FALSE      FALSE
## Years          FALSE      FALSE
## CAtBat         FALSE      FALSE
## CHits          FALSE      FALSE
## CHmRun         FALSE      FALSE
## CRuns          FALSE      FALSE
## CRBI           FALSE      FALSE
## CWalks         FALSE      FALSE
## LeagueN        FALSE      FALSE
## DivisionW      FALSE      FALSE
## PutOuts        FALSE      FALSE
## Assists        FALSE      FALSE
## Errors         FALSE      FALSE
## NewLeagueN     FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##          AtBat Hits HmRun Runs RBI Walks Years CAtBat CHits CHmRun CRuns CRBI
## 1  ( 1 ) " "   " "  " "   " "  " " " " " "   " "   " "    " "   " "    " "   "*"
## 2  ( 1 ) " "   "*"  " "   " "  " " " " " "   " "   " "    " "   " "    " "   "*"
## 3  ( 1 ) " "   "*"  " "   " "  " " " " " "   " "   " "    " "   " "    " "   "*"
## 4  ( 1 ) " "   "*"  " "   " "  " " " " " "   " "   " "    " "   " "    " "   "*"
## 5  ( 1 ) "*"   "*"  " "   " "  " " " " " "   " "   " "    " "   " "    " "   "*"
## 6  ( 1 ) "*"   "*"  " "   " "  " " " " "*"   " "   " "    " "   " "    " "   "*"
## 7  ( 1 ) " "   "*"  " "   " "  " " " " "*"   " "   "*"    "*"   "*"    " "   " "
## 8  ( 1 ) "*"   "*"  " "   " "  " " " " "*"   " "   " "    " "   "*"    "*"   " "
##          CWalks LeagueN DivisionW PutOuts Assists Errors NewLeagueN
## 1  ( 1 ) " "    " "     " "       " "     " "     " "    " "
## 2  ( 1 ) " "    " "     " "       " "     " "     " "    " "
## 3  ( 1 ) " "    " "     " "       "*"     " "     " "    " "
## 4  ( 1 ) " "    " "     "*"       "*"     " "     " "    " "
## 5  ( 1 ) " "    " "     "*"       "*"     " "     " "    " "
## 6  ( 1 ) " "    " "     "*"       "*"     " "     " "    " "
## 7  ( 1 ) " "    " "     "*"       "*"     " "     " "    " "
## 8  ( 1 ) "*"    " "     "*"       "*"     " "     " "    " "
```

```
summary(regfit.full)$which
```

```
##   (Intercept) AtBat  Hits HmRun  Runs   RBI Walks Years CAtBat CHits CHmRun
## 1        TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 2        TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 3        TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 4        TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 5        TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 6        TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  FALSE FALSE  FALSE
## 7        TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE   TRUE  TRUE   TRUE
## 8        TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  FALSE FALSE   TRUE
##   CRuns  CRBI CWalks LeagueN DivisionW PutOuts Assists Errors NewLeagueN
## 1 FALSE  TRUE  FALSE   FALSE     FALSE   FALSE   FALSE  FALSE      FALSE
```

```
## 2 FALSE  TRUE  FALSE   FALSE     FALSE   FALSE   FALSE  FALSE      FALSE
## 3 FALSE  TRUE  FALSE   FALSE     FALSE    TRUE   FALSE  FALSE      FALSE
## 4 FALSE  TRUE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 5 FALSE  TRUE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 6 FALSE  TRUE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 7 FALSE FALSE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 8  TRUE FALSE   TRUE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
```

```
summary(regfit.full)$rss
```

```
## [1] 36179679 30646560 29249297 27970852 27149899 26194904 25906548 25136930
```

```
# Set number of variables to 19

regfit.full19 = regsubsets(Salary~.,data=Hitters,nvmax=19)


regfit.full19.summary = summary(regfit.full19)

regfit.full19.summary$which
```

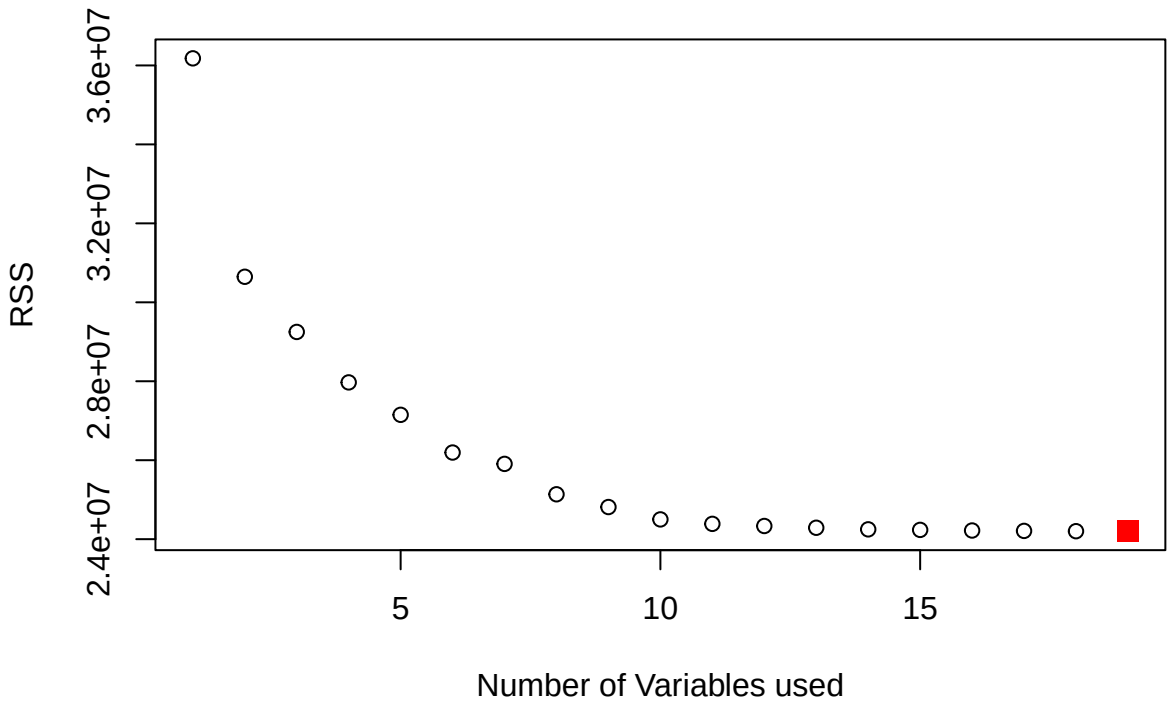
```
##    (Intercept) AtBat  Hits HmRun  Runs   RBI Walks Years CAtBat CHits CHmRun
## 1         TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 2         TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 3         TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 4         TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 5         TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE
## 6         TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  FALSE FALSE  FALSE
## 7         TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE   TRUE  TRUE   TRUE
## 8         TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE  FALSE FALSE   TRUE
## 9         TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE   TRUE FALSE  FALSE
## 10        TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE   TRUE FALSE  FALSE
## 11        TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE FALSE   TRUE FALSE  FALSE
## 12        TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE   TRUE FALSE  FALSE
## 13        TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE   TRUE FALSE  FALSE
## 14        TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE   TRUE FALSE  FALSE
## 15        TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE   TRUE  TRUE  FALSE
## 16        TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE   TRUE  TRUE  FALSE
## 17        TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE   TRUE  TRUE  FALSE
## 18        TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE   TRUE  TRUE  FALSE
## 19        TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE   TRUE  TRUE   TRUE
##    CRuns  CRBI CWalks LeagueN DivisionW PutOuts Assists Errors NewLeagueN
## 1  FALSE  TRUE  FALSE   FALSE     FALSE   FALSE   FALSE  FALSE      FALSE
## 2  FALSE  TRUE  FALSE   FALSE     FALSE   FALSE   FALSE  FALSE      FALSE
## 3  FALSE  TRUE  FALSE   FALSE     FALSE    TRUE   FALSE  FALSE      FALSE
## 4  FALSE  TRUE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 5  FALSE  TRUE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 6  FALSE  TRUE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 7  FALSE FALSE  FALSE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 8   TRUE FALSE   TRUE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 9   TRUE  TRUE   TRUE   FALSE      TRUE    TRUE   FALSE  FALSE      FALSE
## 10  TRUE  TRUE   TRUE   FALSE      TRUE    TRUE    TRUE  FALSE      FALSE
## 11  TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE  FALSE      FALSE
```

```
## 12  TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE  FALSE      FALSE
## 13  TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 14  TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 15  TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 16  TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE      FALSE
## 17  TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE       TRUE
## 18  TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE       TRUE
## 19  TRUE  TRUE   TRUE    TRUE      TRUE    TRUE    TRUE   TRUE       TRUE
```

```
regfit.full19.summary$rss
```

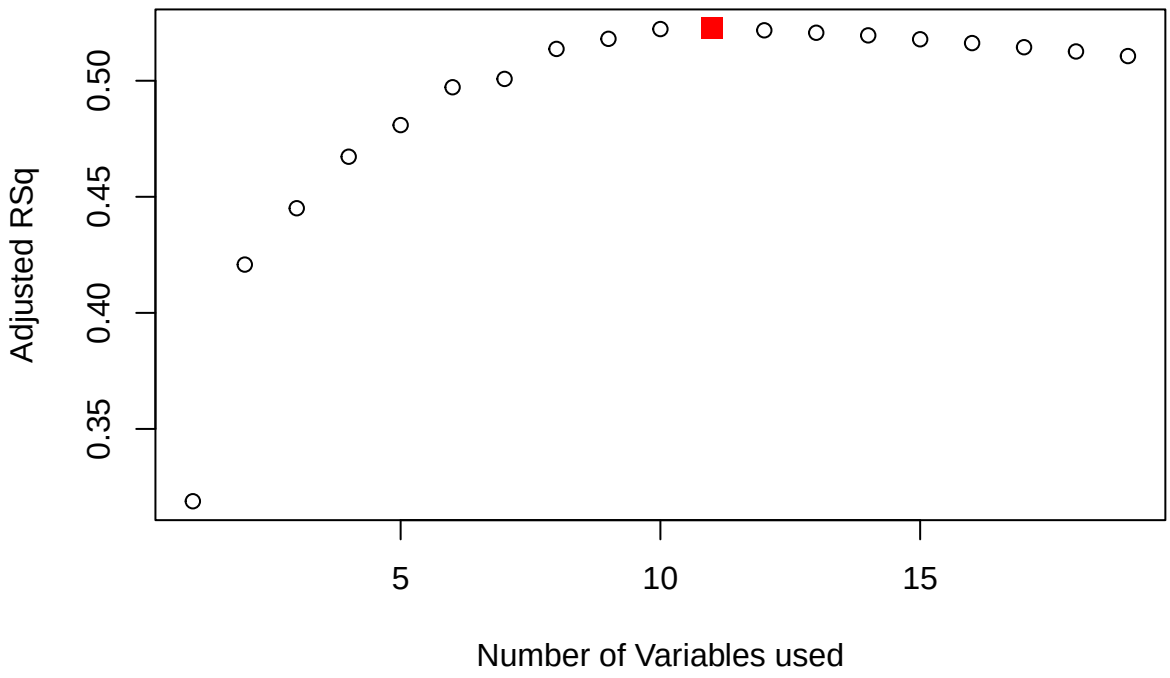
```
##  [1] 36179679 30646560 29249297 27970852 27149899 26194904 25906548 25136930
##  [9] 24814051 24500402 24387345 24333232 24289148 24248660 24235177 24219377
## [17] 24209447 24201837 24200700
```

```
#          AtBat Hits HmRun Runs RBI Walks Years CAtBat CHits CHmRun CRuns CRBI CWalks LeagueN Division
# 1  ( 1 )  " "   " "  " "   " "  " " " " " "   " "   " "    " "   " "    " "   "*"  " "    " "      " "
# 2  ( 1 )  " "   "*"  " "   " "  " " " " " "   " "   " "    " "   " "    " "   "*"  " "    " "      " "
# 3  ( 1 )  " "   "*"  " "   " "  " " " " " "   " "   " "    " "   " "    " "   "*"  " "    " "      " "
# 4  ( 1 )  " "   "*"  " "   " "  " " " " " "   " "   " "    " "   " "    " "   "*"  " "    " "      "*"
# 5  ( 1 )  "*"   "*"  " "   " "  " " " " " "   " "   " "    " "   " "    " "   "*"  " "    " "      "*"
# 6  ( 1 )  "*"   "*"  " "   " "  " " " " "*"   " "   " "    " "   " "    " "   "*"  " "    " "      "*"
# 7  ( 1 )  " "   "*"  " "   " "  " " " " "*"   " "   "*"    "*"   "*"    " "   " "  " "    " "      "*"
# 8  ( 1 )  "*"   "*"  " "   " "  " " " " "*"   " "   " "    " "   "*"    "*"   " "  "*"    " "      "*"

# We can see in #7 instance , CRBI is not choosen even though it was choosen from #1 to #6

# plotting Values


plot(regfit.full19.summary$rss, xlab = "Number of Variables used",ylab="RSS") +
points(which.min(regfit.full19.summary$rss),min(regfit.full19.summary$rss),col="red",cex=1.5,pch=15)
```



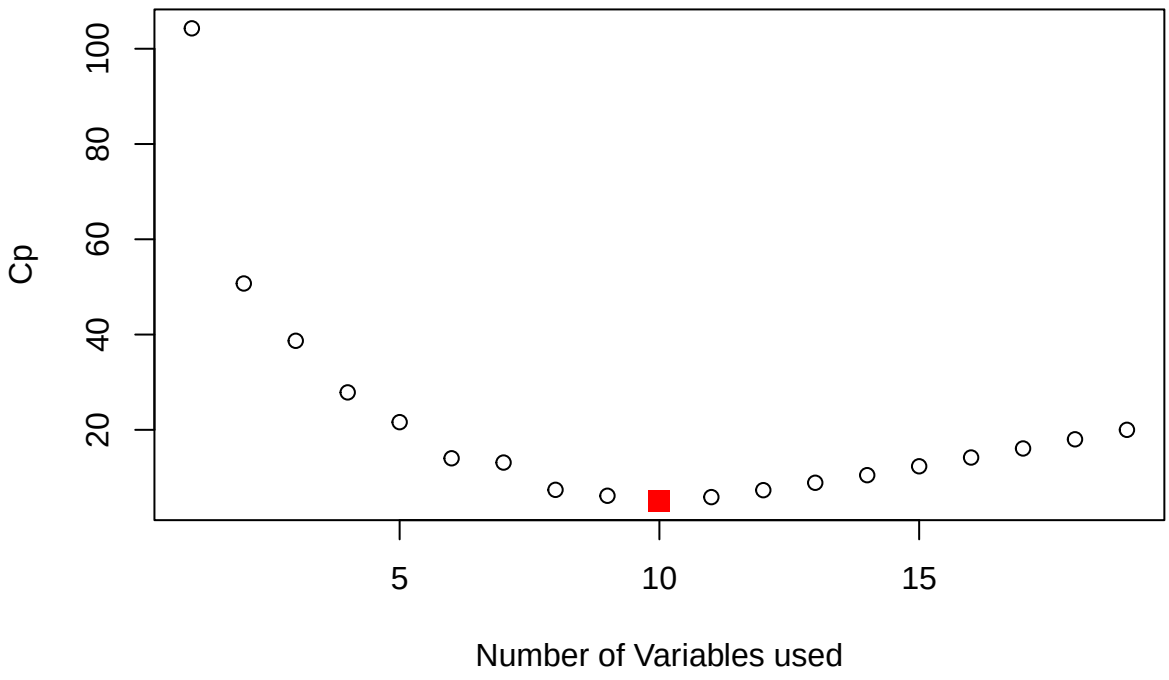

```
## integer(0)
```

```
plot(regfit.full19.summary$adjr2, xlab = "Number of Variables used",ylab="Adjusted RSq") +
points(which.max(regfit.full19.summary$adjr2),max(regfit.full19.summary$adjr2),col="red",cex=1.5,pch=15)
```



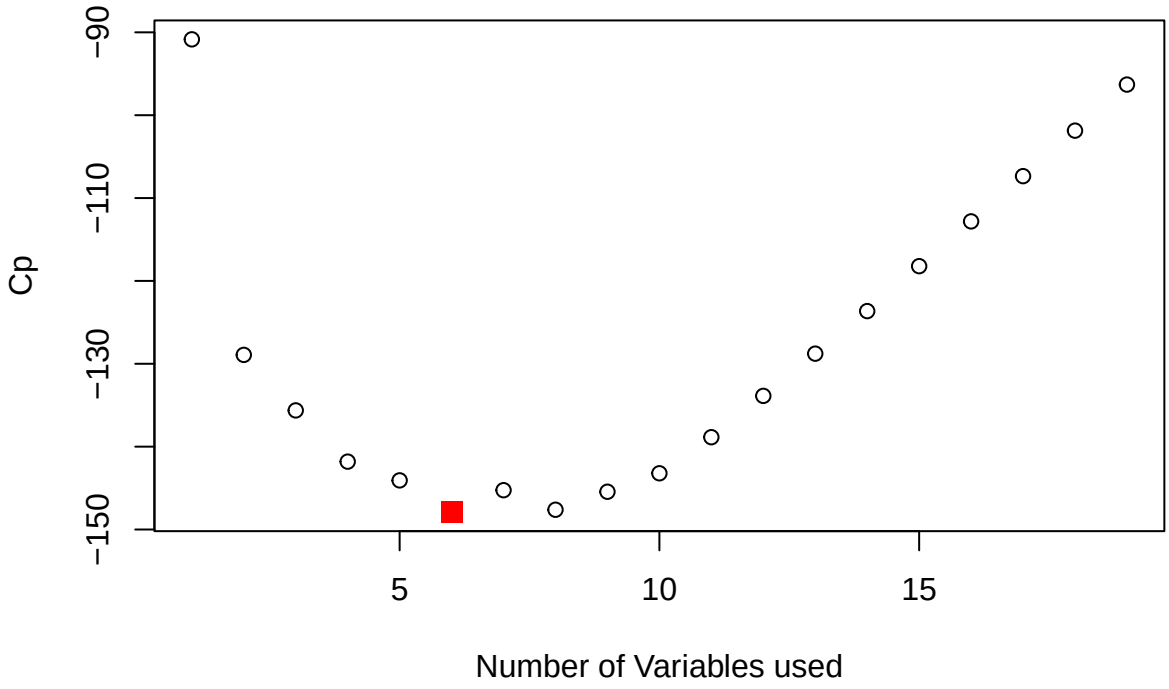

```
## integer(0)
```

```
plot(regfit.full19.summary$cp, xlab = "Number of Variables used",ylab="Cp") +
points(which.min(regfit.full19.summary$cp),min(regfit.full19.summary$cp),col="red",cex=1.5,pch=15)
```




```
## integer(0)
```

```
plot(regfit.full19.summary$bic, xlab = "Number of Variables used",ylab="Cp") +
points(which.min(regfit.full19.summary$bic),min(regfit.full19.summary$bic),col="red",cex=1.5,pch=15)
```




```
## integer(0)
```