

L1Workshop2

ye

6/28/2017

Subset Selection Methods

```
suppressWarnings(library(ISLR))
sum(is.na(Hitters))
```

```
## [1] 59
```

```
Hitters=na.omit(Hitters)
dim(Hitters)
```

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

```
suppressWarnings(library(leaps))
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
## 1 ( 1 ) " " " " " " " " " " " " " " " " " "
## 2 ( 1 ) " " "*" " " " " " " " " " " " " " "
## 3 ( 1 ) " " "*" " " " " " " " " " " " " " "
## 4 ( 1 ) " " "*" " " " " " " " " " " " " " "
## 5 ( 1 ) "*" "*" " " " " " " " " " " " " " "
## 6 ( 1 ) "*" "*" " " " " " " "*" " " " " " " "
```

```

## 7 ( 1 ) " " "*" " " " " " " "*" " " "*" "*" "*" " "
## 8 ( 1 ) "*" "*" " " " " " " "*" " " " " " " "*" "*"
##      CRBI CWalks LeagueN DivisionW PutOuts Assists Errors NewLeagueN
## 1 ( 1 ) "*" " " " " " " " " " " " " " "
## 2 ( 1 ) "*" " " " " " " " " " " " " " "
## 3 ( 1 ) "*" " " " " " " "*" " " " " " "
## 4 ( 1 ) "*" " " " " "*" "*" " " " " " "
## 5 ( 1 ) "*" " " " " "*" "*" " " " " " "
## 6 ( 1 ) "*" " " " " "*" "*" " " " " " "
## 7 ( 1 ) " " " " " " "*" "*" " " " " " "
## 8 ( 1 ) " " "*" " " "*" "*" " " " " " "

regfit.full=regsubsets(Salary~.,data=Hitters,nvmax=19)
reg.summary=summary(regfit.full)
par(mfrow=c(2,2))
plot(reg.summary$rss,xlab="Number of Variables",ylab="RSS",type="l")
plot(reg.summary$adjr2,xlab="Number of Variables",ylab="Adjusted RSq",type="l")
which.max(reg.summary$adjr2)

## [1] 11

points(11,reg.summary$adjr2[11], col="red",cex=2,pch=20)
plot(reg.summary$cp,xlab="Number of Variables",ylab="Cp",type='l')
which.min(reg.summary$cp)

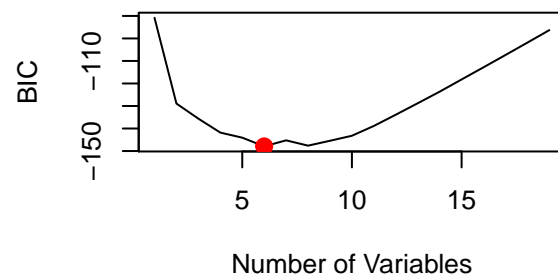
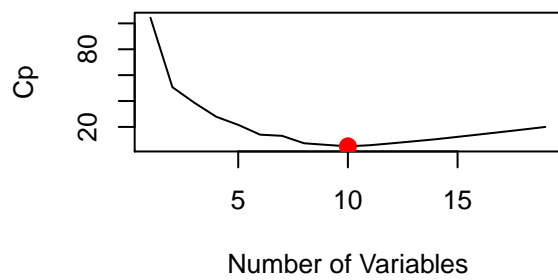
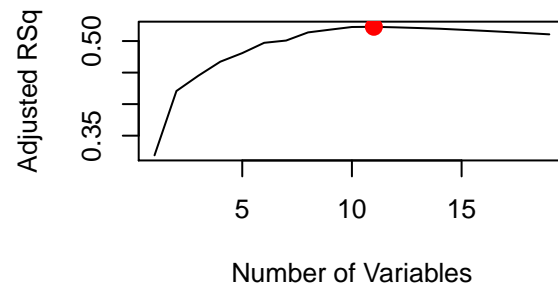
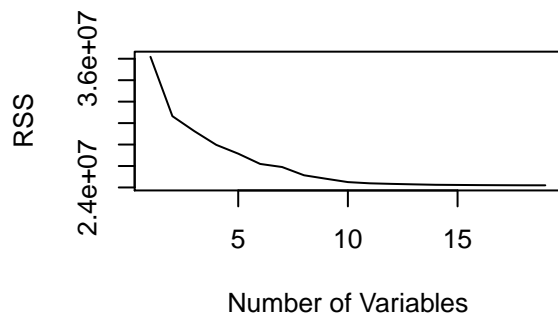
## [1] 10

points(10,reg.summary$cp[10],col="red",cex=2,pch=20)
which.min(reg.summary$bic)

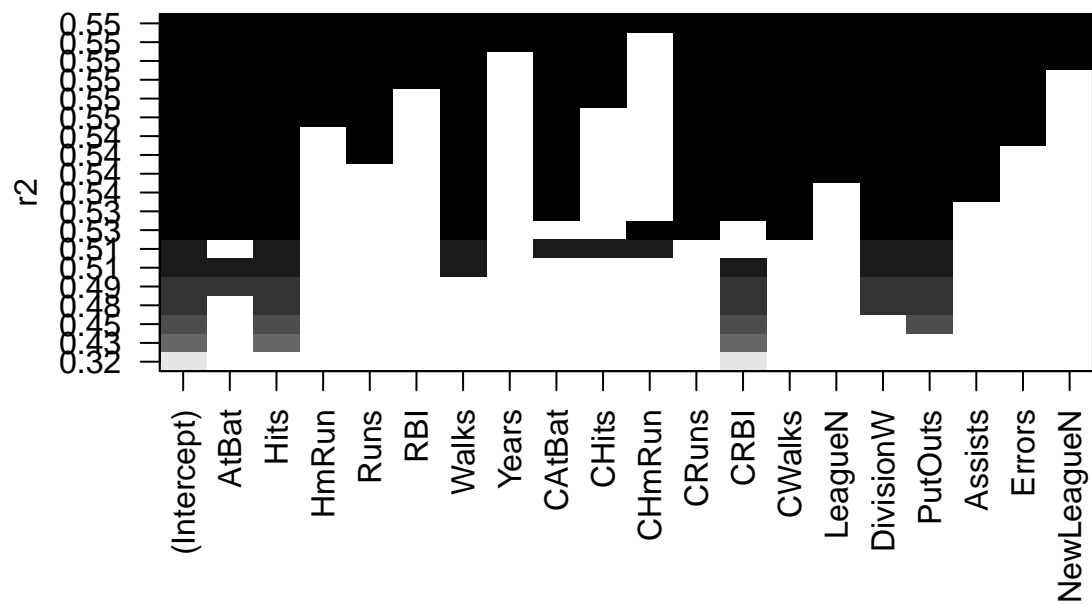
## [1] 6

plot(reg.summary$bic,xlab="Number of Variables",ylab="BIC",type='l')
points(6,reg.summary$bic[6],col="red",cex=2,pch=20)

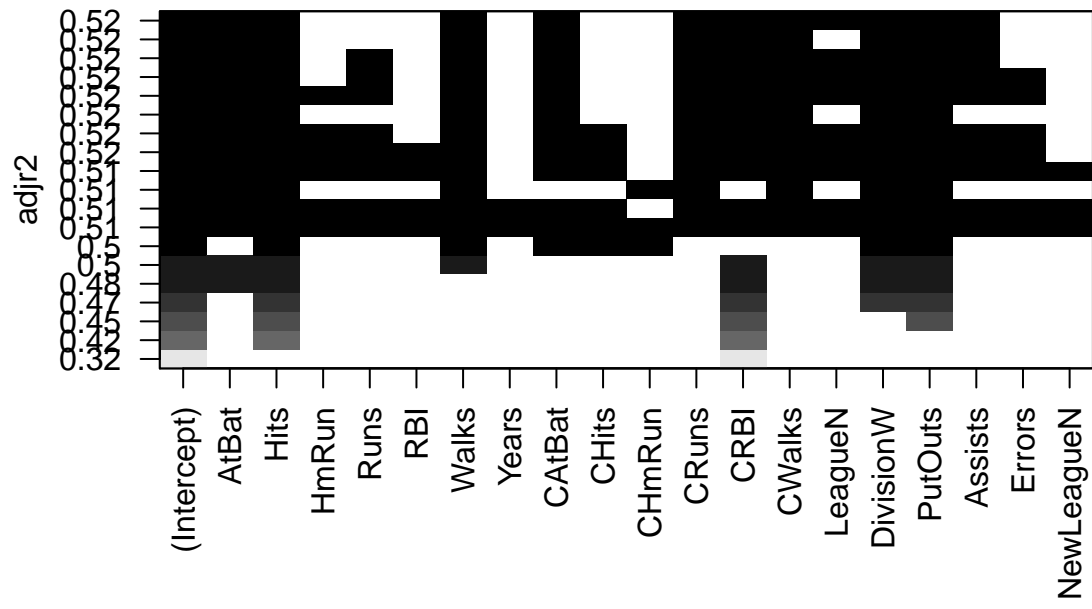
```



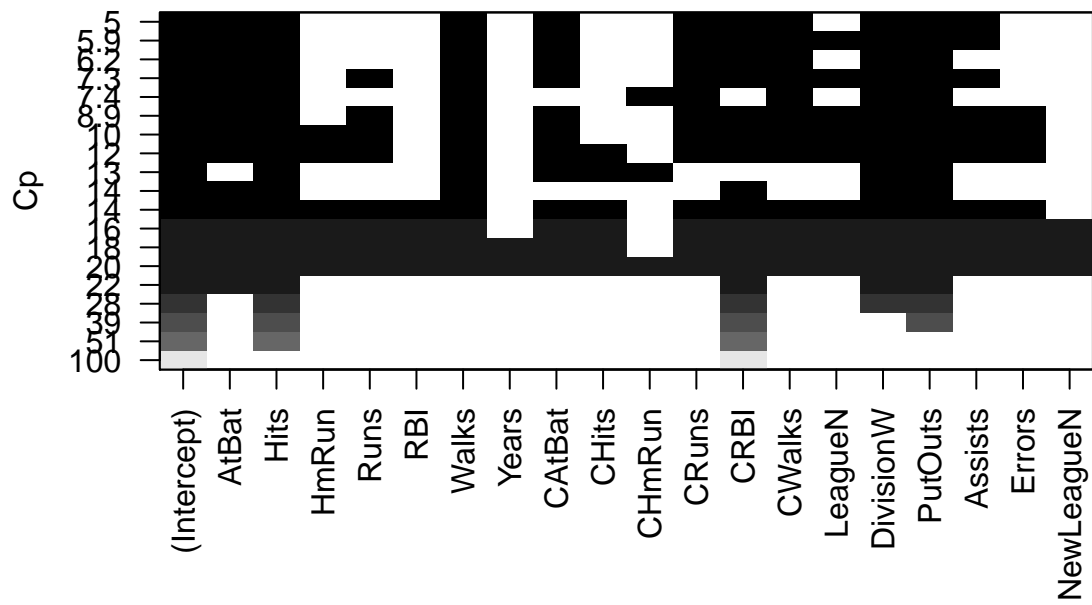
```
par(mfrow=c(1,1))
plot(regfit.full,scale="r2")
```



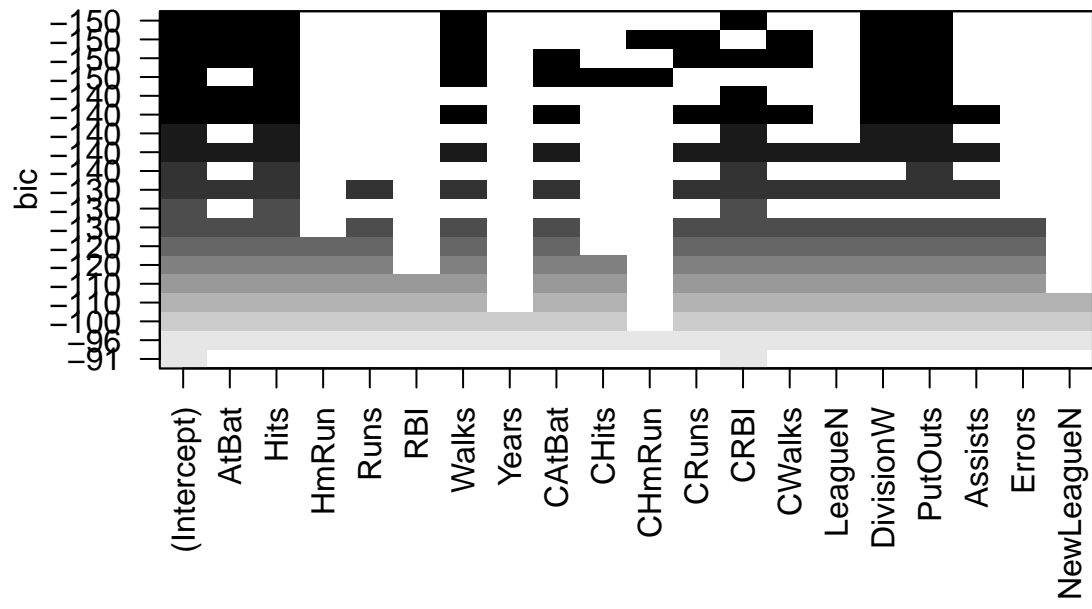
```
plot(regfit.full,scale="adjr2")
```



```
plot(regfit.full,scale="Cp")
```



```
plot(regfit.full,scale="bic")
```



```
coef(regfit.full,6)
```

```
## (Intercept)      AtBat      Hits      Walks      CRBI
##  91.5117981   -1.8685892   7.6043976   3.6976468   0.6430169
##   DivisionW     PutOuts
## -122.9515338    0.2643076
```