Assignment4

MengZhang

4/4/2019

```
set.seed(1)
x <- rnorm(100)
noise <- rnorm(100)
```

```
y <- 2+3*x-1*x^2+1*x^3+noise
```

```
install.packages(pkgs='leaps')
```

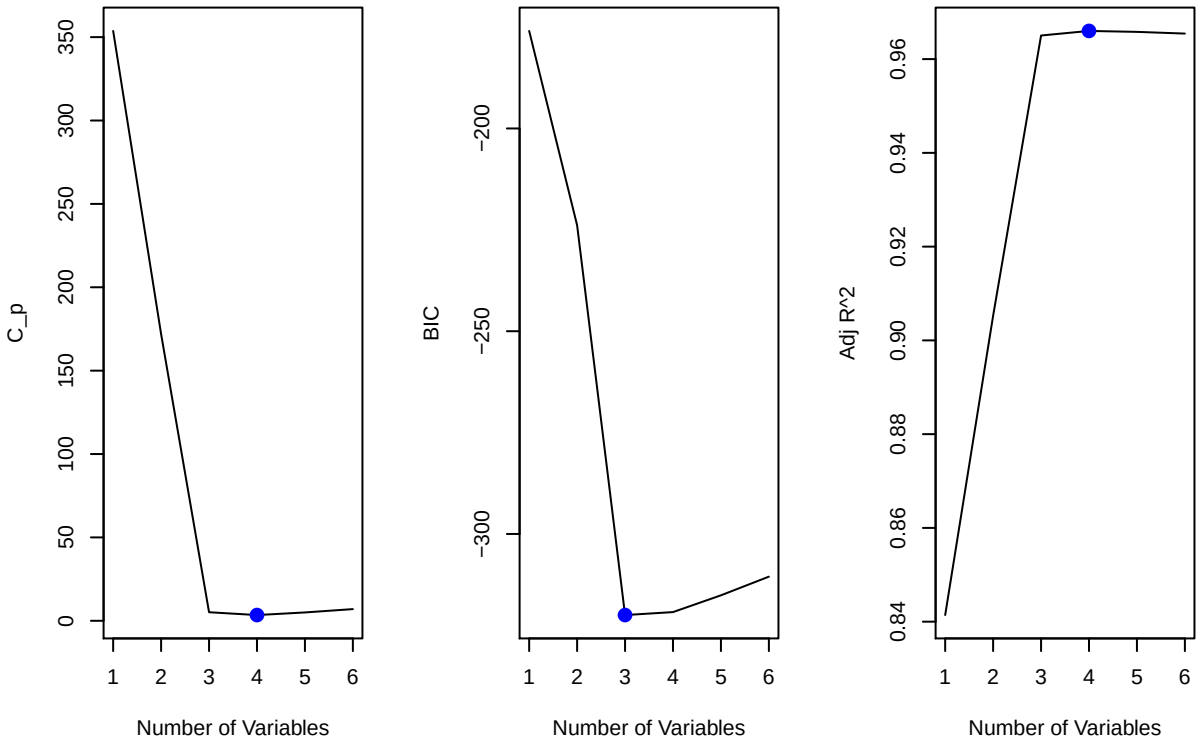
```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/3.5'
## (as 'lib' is unspecified)
```

```
library(leaps)
data.full<-data.frame(y=y,x=x)
regfit.fwd<-regsubsets(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6), data = data.full, nvmax = 6,
reg.summary.fwd<-summary(regfit.fwd)
par(mfrow=c(1,3))

#plot model c_p value for different number of variables.Least value of c_p gives best model
plot(reg.summary.fwd$cp,xlab="Number of Variables",ylab="C_p",type="l")
points(which.min(reg.summary.fwd$cp),reg.summary.fwd$cp[which.min(reg.summary.fwd$cp)],col="blue",cex=2

#plot model BIC value
plot(reg.summary.fwd$bic,xlab="Number of Variables",ylab="BIC",type="l")
points(which.min(reg.summary.fwd$bic),reg.summary.fwd$bic[which.min(reg.summary.fwd$bic)],col="blue",ce

#plot model adj R square. HIgher adj r sqaure gives best model
plot(reg.summary.fwd$adjr2,xlab="Number of Variables",ylab="Adj R^2",type="l")
points(which.max(reg.summary.fwd$adjr2),reg.summary.fwd$adjr2[which.max(reg.summary.fwd$adjr2)],col="bl
```




```
coef(regfit.fwd,which.max(reg.summary.fwd$adjr2))
```

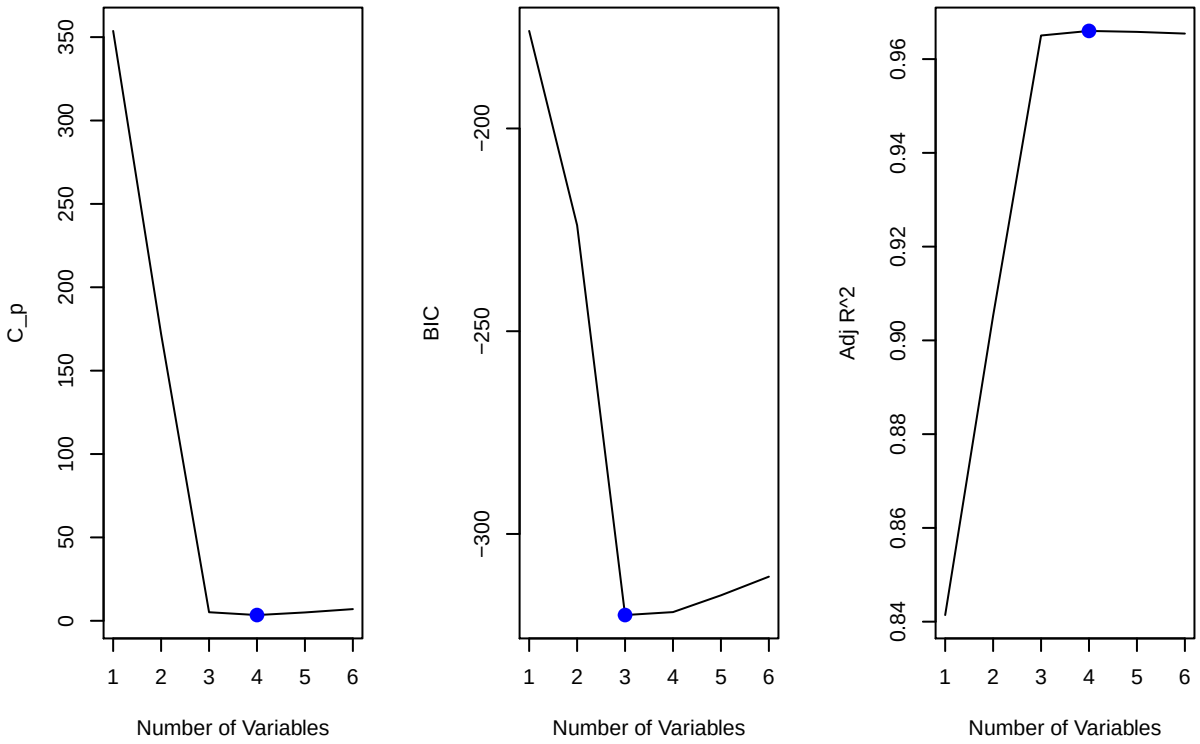
```
## (Intercept)           x      I(x^2)      I(x^3)      I(x^5)
##  2.07200775  3.38745596 -1.15424359  0.55797426  0.08072292
```

```
library(leaps)
data.full<-data.frame(y=y,x=x)
regfit.bwd<-regsubsets(y~x+I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6), data = data.full, nvmax = 6,reall
reg.summary.bwd<-summary(regfit.bwd)
par(mfrow=c(1,3))

#plot model c_p value for different number of variables.Least value of c_p gives best model
plot(reg.summary.bwd$cp,xlab="Number of Variables",ylab="C_p",type="l")
points(which.min(reg.summary.bwd$cp),reg.summary.bwd$cp[which.min(reg.summary.bwd$cp)],col="blue",cex=2

#plot model BIC value
plot(reg.summary.bwd$bic,xlab="Number of Variables",ylab="BIC",type="l")
points(which.min(reg.summary.bwd$bic),reg.summary.bwd$bic[which.min(reg.summary.bwd$bic)],col="blue",ce

#plot model adj R square. HIgher adj r sqaure gives best model
plot(reg.summary.bwd$adjr2,xlab="Number of Variables",ylab="Adj R^2",type="l")
points(which.max(reg.summary.bwd$adjr2),reg.summary.bwd$adjr2[which.max(reg.summary.bwd$adjr2)],col="blu
```




```
coef(regfit.bwd,which.max(reg.summary.bwd$adjr2))
```

```
## (Intercept)           x      I(x^2)      I(x^3)      I(x^5)
##  2.07200775  3.38745596 -1.15424359  0.55797426  0.08072292
```

```
install.packages(pkgs='glmnet')
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/3.5'
## (as 'lib' is unspecified)
```

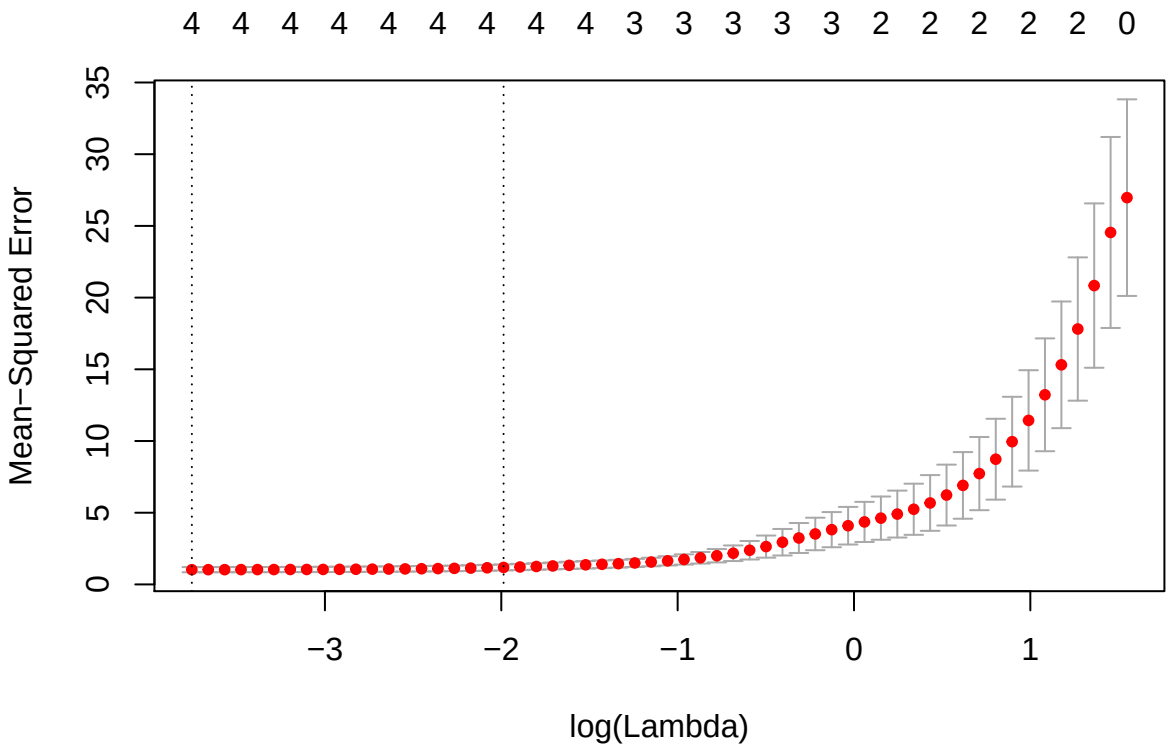
```
library(glmnet)
```

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

```
## Loading required package: foreach
```

```
## Loaded glmnet 2.0-16
```

```
set.seed(1)
xmat<-model.matrix(y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6), data = data.full)[,-1]
set.seed(1)
cv.lasso<-cv.glmnet(xmat,y,alpha=1)
plot(cv.lasso)
```




```
bestlam<-cv.lasso$lambda.min
bestlam
```

```
## [1] 0.02340688
```

```
fit.lasso<-glmnet(xmat,y,alpha=1)
predict(fit.lasso,s=bestlam,type="coefficients")[1:7,]
```

```
## (Intercept)           x      I(x^2)      I(x^3)      I(x^4)      I(x^5)
##  2.05276143  3.30926245 -1.12412051  0.62295673  0.00000000  0.06809596
##      I(x^6)
##  0.00000000
```