

350 presentation

Xiaoliang ZHANG

2019/11/12

```
#install.packages('CombMSC')
library(CombMSC)

##
## Attaching package: 'CombMSC'

## The following object is masked from 'package:stats':
##
##     BIC

#install.packages('leaps')
library(leaps)

SH=read.csv("~/Desktop/STAT 350/Project/ShanghaiPM20100101_20151231.csv")

# remove NA in the response
SH=na.omit(SH)
dim(SH)

## [1] 21436    17

sum(is.na(SH))

## [1] 0

regfit.full=regsubsets(PM_US.Post~., data=SH)
summary(regfit.full)

## Subset selection object
## Call: regsubsets.formula(PM_US.Post ~ ., data = SH)
## 19 Variables (and intercept)
##
## Forced in Forced out
## No FALSE FALSE
## year FALSE FALSE
## month FALSE FALSE
## day FALSE FALSE
## hour FALSE FALSE
## season FALSE FALSE
## PM_Jingan FALSE FALSE
## PM_Xuhui FALSE FALSE
## DEWP FALSE FALSE
## HUMI FALSE FALSE
## PRES FALSE FALSE
## TEMP FALSE FALSE
## cbwdNE FALSE FALSE
## cbwdNW FALSE FALSE
## cbwdSE FALSE FALSE
## cbwdSW FALSE FALSE
## Iws FALSE FALSE
## precipitation FALSE FALSE
## Iprec FALSE FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##
## No year month day hour season PM_Jingan PM_Xuhui DEWP HUMI PRES
## 1 ( 1) " " " " " " " " " " " " " " " " " " " " " "
## 2 ( 1) " " " " " " " " " " " " " " " " " " " " "
## 3 ( 1) " " " " " " " " " " " " " " " " " " " "
## 4 ( 1) " " " " " " " " " " " " " " " " " " " "
## 5 ( 1) " " " " " " " " " " " " " " " " " " " "
## 6 ( 1) " " " " " " " " " " " " " " " " " " " "
## 7 ( 1) " " " " " " " " " " " " " " " " " " " "
## 8 ( 1) " " " " " " " " " " " " " " " " " " " "
##
## TEMP cbwdNE cbwdNW cbwdSE cbwdSW Iws precipitation Iprec
## 1 ( 1) " " " " " " " " " " " " " " " " " " " "
## 2 ( 1) " " " " " " " " " " " " " " " " " " " "
## 3 ( 1) " " " " " " " " " " " " " " " " " " " "
## 4 ( 1) " " " " " " " " " " " " " " " " " " " "
## 5 ( 1) " " " " " " " " " " " " " " " " " " " "
## 6 ( 1) " " " " " " " " " " " " " " " " " " " "
## 7 ( 1) " " " " " " " " " " " " " " " " " " " "
## 8 ( 1) " " " " " " " " " " " " " " " " " " " "
```

```
# nvmax defines the maximum size of subsets to examine
# default is nvmax = 8
regfit.full=regsubsets(PM_US.Post~.,data=SH, nvmax=17)
(reg.summary=summary(regfit.full))
```

```
## Subset selection object
## Call: regsubsets.formula(PM_US.Post ~ ., data = SH, nvmax = 17)
## 19 Variables (and intercept)
##
## Forced in Forced out
## No FALSE FALSE
## year FALSE FALSE
## month FALSE FALSE
## day FALSE FALSE
## hour FALSE FALSE
## season FALSE FALSE
## PM_Jingan FALSE FALSE
## PM_Xuhui FALSE FALSE
## DEWP FALSE FALSE
## HUMI FALSE FALSE
## PRES FALSE FALSE
## TEMP FALSE FALSE
## cbwdNE FALSE FALSE
## cbwdNW FALSE FALSE
## cbwdSE FALSE FALSE
## cbwdSW FALSE FALSE
## Iws FALSE FALSE
## precipitation FALSE FALSE
## Iprec FALSE FALSE
## 1 subsets of each size up to 17
## Selection Algorithm: exhaustive
##
## No year month day hour season PM_Jingan PM_Xuhui DEWP HUMI PRES
## 1 ( 1) " " " " " " " " " " " " " " " " " " " "
## 2 ( 1) " " " " " " " " " " " " " " " " " " " "
## 3 ( 1) " " " " " " " " " " " " " " " " " " " "
## 4 ( 1) " " " " " " " " " " " " " " " " " " " "
## 5 ( 1) " " " " " " " " " " " " " " " " " " " "
## 6 ( 1) " " " " " " " " " " " " " " " " " " " "
## 7 ( 1) " " " " " " " " " " " " " " " " " " " "
## 8 ( 1) " " " " " " " " " " " " " " " " " " " "
## 9 ( 1) " " " " " " " " " " " " " " " " " " " "
## 10 ( 1) " " " " " " " " " " " " " " " " " " " "
## 11 ( 1) " " " " " " " " " " " " " " " " " " " "
## 12 ( 1) " " " " " " " " " " " " " " " " " " " "
## 13 ( 1) " " " " " " " " " " " " " " " " " " " "
## 14 ( 1) " " " " " " " " " " " " " " " " " " " "
## 15 ( 1) " " " " " " " " " " " " " " " " " " " "
## 16 ( 1) " " " " " " " " " " " " " " " " " " " "
## 17 ( 1) " " " " " " " " " " " " " " " " " " " "
##
## TEMP cbwdNE cbwdNW cbwdSE cbwdSW Iws precipitation Iprec
## 1 ( 1) " " " " " " " " " " " " " " " " " " " "
## 2 ( 1) " " " " " " " " " " " " " " " " " " " "
## 3 ( 1) " " " " " " " " " " " " " " " " " " " "
## 4 ( 1) " " " " " " " " " " " " " " " " " " " "
## 5 ( 1) " " " " " " " " " " " " " " " " " " " "
## 6 ( 1) " " " " " " " " " " " " " " " " " " " "
## 7 ( 1) " " " " " " " " " " " " " " " " " " " "
## 8 ( 1) " " " " " " " " " " " " " " " " " " " "
## 9 ( 1) " " " " " " " " " " " " " " " " " " " "
## 10 ( 1) " " " " " " " " " " " " " " " " " " " "
## 11 ( 1) " " " " " " " " " " " " " " " " " " " "
## 12 ( 1) " " " " " " " " " " " " " " " " " " " "
## 13 ( 1) " " " " " " " " " " " " " " " " " " " "
## 14 ( 1) " " " " " " " " " " " " " " " " " " " "
## 15 ( 1) " " " " " " " " " " " " " " " " " " " "
## 16 ( 1) " " " " " " " " " " " " " " " " " " " "
## 17 ( 1) " " " " " " " " " " " " " " " " " " " "
```

```
names(reg.summary)

## [1] "which" "rsq" "rss" "adjr2" "cp" "bic" "outmat" "obj"
```

```
reg.summary$rsq

## [1] 0.9250253 0.9429089 0.9467413 0.9472375 0.9475254 0.9475690 0.9476047
## [8] 0.9476534 0.9477323 0.9477793 0.9478152 0.9478570 0.9478918 0.9478956
## [15] 0.9478993 0.9479114 0.9479357
```

```
# plotting RSS, adjusted r^2, Cp and BIC vs Number of Variables
par(mfrow=c(2,2))
plot(reg.summary$rs, xlab="Number of Variables", ylab="RSS", type="l")
plot(reg.summary$adjr2, xlab="Number of Variables", ylab="Adjusted RSq", type="l")

# find the model with the largest adjusted r^2
which.max(reg.summary$adjr2)

## [1] 17
```

```
points(13,reg.summary$adjr2[13], col="red",cex=2,pch=20)
plot(reg.summary$cp,xlab="Number of Variables",ylab="Cp",type='l')

# find the model with the lowest cp
which.min(reg.summary$cp)

## [1] 17
```

```
points(12,reg.summary$cp[12],col="red",cex=2,pch=20)

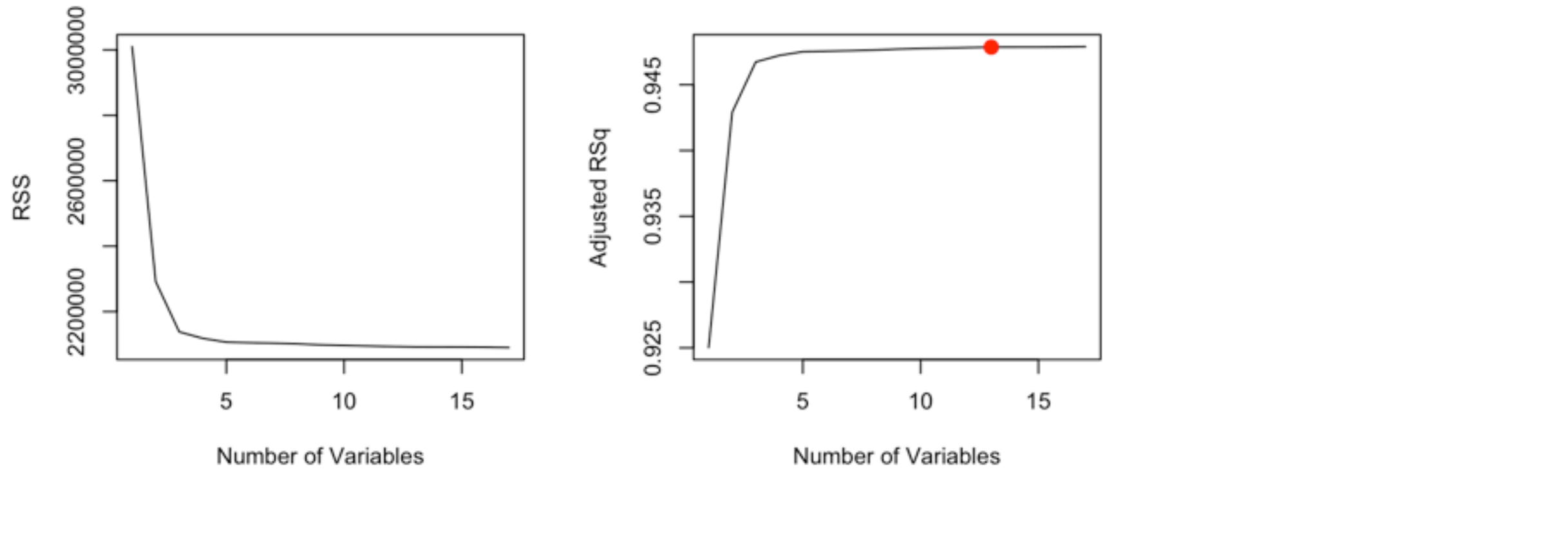
# find the model with the lowest BIC
which.min(reg.summary$bic)

## [1] 13
```

```
plot(reg.summary$bic,xlab="Number of Variables",ylab="BIC",type='l')
points(10,reg.summary$bic[10],col="red",cex=2,pch=20)
```



```
# plotting RSS, adjusted r^2, Cp and BIC for comparison
par(mfrow=c(2,2))
plot(regfit.full,scale="r2")
plot(regfit.full,scale="adjr2")
plot(regfit.full,scale="Cp")
plot(regfit.full,scale="bic")
```



```
# according to BIC, a 6-variable model is the best model
coef(regfit.full,10)
```

```
## (Intercept)          hour      PM_Jingan      PM_Xuhui      HUMI
## 14.16915233      0.10737306      0.63121751      0.21188373     -0.04025392
##          TEMP      cbwdNE      cbwdNW      cbwdSE      cbwdSW
## -0.60854525      3.39145936      3.69288398      3.21505603      2.85235270
## precipitation
## -0.29194702
```