Calculate Cv value with diffrent polynomial degree regression.

Mehrab Atighi

5/4/2021

# library the packages that we need to solve this exercise

after the library the needed packages we should library our data

library(ISLR)

## Warning: package 'ISLR' was built under R version 4.0.4

library(reshape2)

## Warning: package 'reshape2' was built under R version 4.0.5

library(tibble)

## Warning: package 'tibble' was built under R version 4.0.5

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.3

data(Auto)

Now we want to use leave one out cross validation to calculate Cv with diffrent polynomial degree regression. At the first we can see summary of our data

# Leave-one-out cross-validation :  
dt<-data.frame(Auto$mpg , Auto$horsepower)  
head(dt)

## Auto.mpg Auto.horsepower  
## 1 18 130  
## 2 15 165  
## 3 18 150  
## 4 16 150  
## 5 17 140  
## 6 15 198

Now we make a table and make loops for calculate each MSE with diffrent polynomial degree regression.

all.MSE.table<-tibble(number.of.leave.out.data = rep(c(1:nrow(dt)),10) , degree = rep(c(1:10),each=nrow(dt)) ,MSE =c(1:3920))  
# i show the degree of the polynomial regression function  
# j show the number of data that we will leave out (MSE\_j) for each i  
test<-data.frame(dt$Auto.horsepower , dt$Auto.mpg)  
count=1  
for(i in 1:10){  
 for(j in 1:nrow(dt)){  
 train<-data.frame(train1=dt$Auto.mpg[-j] , train2=dt$Auto.horsepower[-j])  
 fit<-glm( train1 ~poly(train2,degree= i),data = train)  
 new= data.frame(train2 = test$dt.Auto.horsepower[j])  
 all.MSE.table$MSE[count]<-(test$dt.Auto.mpg[j]-predict(fit ,newdata = new ,type="response"))^2  
 count=count+1  
 }  
}

Now we can see the all of the MSE (mean squre Error) with diffrent polynomial degree regression.

all.MSE.table

## # A tibble: 3,920 x 3  
## number.of.leave.out.data degree MSE  
## <int> <int> <dbl>  
## 1 1 1 2.02   
## 2 2 1 1.25   
## 3 3 1 3.07   
## 4 4 1 0.0680  
## 5 5 1 0.708   
## 6 6 1 41.4   
## 7 7 1 81.4   
## 8 8 1 67.1   
## 9 9 1 97.0   
## 10 10 1 26.3   
## # ... with 3,910 more rows

Now we want to calculate Cv with MSE values , then see them.

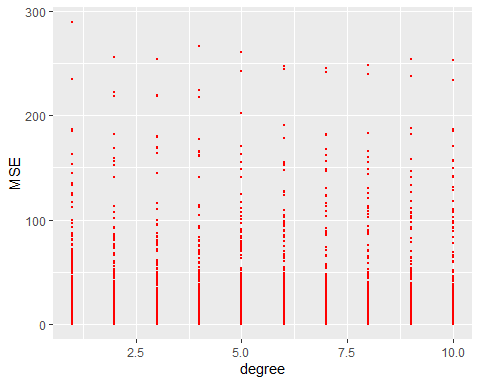
Cv=c(1:10)  
Cv.table<-tibble(Degree =1:10 ,Cv)  
  
for(i in 1:10){  
Cv.table$Cv[i]=mean(all.MSE.table$MSE[which(all.MSE.table$degree==i)])  
}  
  
Cv.table

## # A tibble: 10 x 2  
## Degree Cv  
## <int> <dbl>  
## 1 1 24.2  
## 2 2 19.2  
## 3 3 19.3  
## 4 4 19.4  
## 5 5 19.0  
## 6 6 19.0  
## 7 7 18.8  
## 8 8 19.0  
## 9 9 19.1  
## 10 10 19.5

# drawing plots

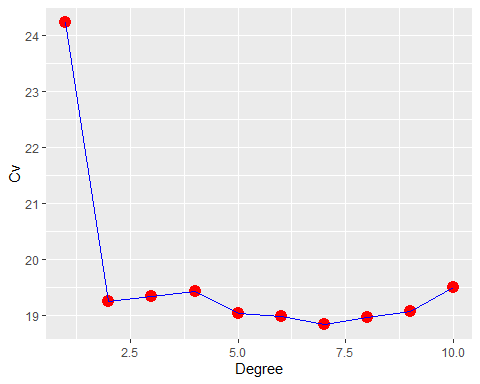
Now we want to draw the plot of y and y hat

ggplot(all.MSE.table , aes(degree,MSE))+  
 geom\_point(colour="Red",cex=0.1)



Now we want to draw the plot of Cv

ggplot(Cv.table,aes(Degree , Cv))+  
 geom\_point(colour = "Red",cex=4) +  
 geom\_line(colour = "Blue")



End.