sriram\_final.R

ram

2022-08-06

datacsv <- read.csv("~/Downloads/final project/datacsv.csv")  
library(tidyr)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(latticeExtra)

## Loading required package: lattice

library(psych)  
library(caret)

## Loading required package: ggplot2

##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

## The following object is masked from 'package:latticeExtra':  
##   
## layer

library(magrittr)

##   
## Attaching package: 'magrittr'

## The following object is masked from 'package:tidyr':  
##   
## extract

library(leaps)  
library(gvlma)  
library(knitr)  
library(glmnet)

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

## Loaded glmnet 4.1-4

library(knitr)  
library(Matrix)  
  
  
  
  
cars <-datacsv <- read.csv("~/Downloads/final project/datacsv.csv")  
names(cars) <- c("mpg","cylinders","displacement","horsepower","weight","acceleration","model\_year","origin","car\_name")  
str(cars)

## 'data.frame': 398 obs. of 9 variables:  
## $ mpg : num 18 15 18 16 17 15 14 14 14 15 ...  
## $ cylinders : int 8 8 8 8 8 8 8 8 8 8 ...  
## $ displacement: num 307 350 318 304 302 429 454 440 455 390 ...  
## $ horsepower : chr "130" "165" "150" "150" ...  
## $ weight : int 3504 3693 3436 3433 3449 4341 4354 4312 4425 3850 ...  
## $ acceleration: num 12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...  
## $ model\_year : int 70 70 70 70 70 70 70 70 70 70 ...  
## $ origin : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ car\_name : chr "chevrolet chevelle malibu" "buick skylark 320" "plymouth satellite" "amc rebel sst" ...

cars$horsepower <- as.numeric(cars$horsepower)

## Warning: NAs introduced by coercion

cars$horsepower[cars$horsepower=="?"] <- NA  
cars$cylinders <- as.numeric(cars$cylinders)  
kable(head(cars),format = "pandoc", caption = "Data Preview")

Data Preview

| mpg | cylinders | displacement | horsepower | weight | acceleration | model\_year | origin | car\_name |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 18 | 8 | 307 | 130 | 3504 | 12.0 | 70 | 1 | chevrolet chevelle malibu |
| 15 | 8 | 350 | 165 | 3693 | 11.5 | 70 | 1 | buick skylark 320 |
| 18 | 8 | 318 | 150 | 3436 | 11.0 | 70 | 1 | plymouth satellite |
| 16 | 8 | 304 | 150 | 3433 | 12.0 | 70 | 1 | amc rebel sst |
| 17 | 8 | 302 | 140 | 3449 | 10.5 | 70 | 1 | ford torino |
| 15 | 8 | 429 | 198 | 4341 | 10.0 | 70 | 1 | ford galaxie 500 |

new\_data <- select(cars,mpg,displacement,horsepower,weight,acceleration)  
new\_data <- na.omit(new\_data)  
kable(summary(new\_data), format="pandoc", caption = "Summery Statistics")

Summery Statistics

|  | mpg | displacement | horsepower | weight | acceleration |
| --- | --- | --- | --- | --- | --- |
|  | Min. : 9.00 | Min. : 68.0 | Min. : 46.0 | Min. :1613 | Min. : 8.00 |
|  | 1st Qu.:17.00 | 1st Qu.:105.0 | 1st Qu.: 75.0 | 1st Qu.:2225 | 1st Qu.:13.78 |
|  | Median :22.75 | Median :151.0 | Median : 93.5 | Median :2804 | Median :15.50 |
|  | Mean :23.45 | Mean :194.4 | Mean :104.5 | Mean :2978 | Mean :15.54 |
|  | 3rd Qu.:29.00 | 3rd Qu.:275.8 | 3rd Qu.:126.0 | 3rd Qu.:3615 | 3rd Qu.:17.02 |
|  | Max. :46.60 | Max. :455.0 | Max. :230.0 | Max. :5140 | Max. :24.80 |

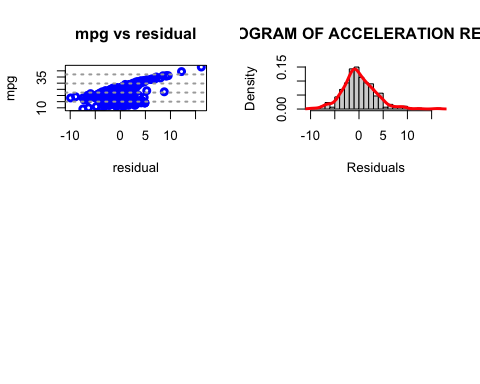
cars.first <- new\_data[1:300,]  
cars.last <- new\_data[301:398,]  
  
#mpg vs displacement  
  
Linear\_Reg.displace <- lm(mpg~displacement, data=cars.first)  
summary(Linear\_Reg.displace)

##   
## Call:  
## lm(formula = mpg ~ displacement, data = cars.first)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.9282 -2.0043 -0.5401 1.9737 16.1501   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 31.352035 0.435875 71.93 <2e-16 \*\*\*  
## displacement -0.048913 0.001809 -27.04 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.412 on 298 degrees of freedom  
## Multiple R-squared: 0.7104, Adjusted R-squared: 0.7094   
## F-statistic: 731.1 on 1 and 298 DF, p-value: < 2.2e-16

par(mfrow=c(2,2))  
residual <- Linear\_Reg.displace$residuals  
plot(cars.first$mpg~residual,lwd=3, col="blue",main="mpg vs residual", xlab="residual",ylab = "mpg")  
grid(NA, 5, lwd = 2,col = "darkgray")  
  
hist(residual,prob=T,breaks=20,main="HISTOGRAM OF ACCELERATION RESIDUALS",xlab="Residuals")  
lines(density(residual),col="red",lwd=3)  
  
  
#mpg vs horsepower  
  
Linear\_Reg.horse\_power <- lm(mpg~horsepower, data=cars.first)  
summary(Linear\_Reg.horse\_power)

##   
## Call:  
## lm(formula = mpg ~ horsepower, data = cars.first)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.8442 -2.7816 -0.3376 2.4948 14.2360   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 34.903508 0.648037 53.86 <2e-16 \*\*\*  
## horsepower -0.125824 0.005455 -23.07 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.8 on 298 degrees of freedom  
## Multiple R-squared: 0.641, Adjusted R-squared: 0.6397   
## F-statistic: 532 on 1 and 298 DF, p-value: < 2.2e-16

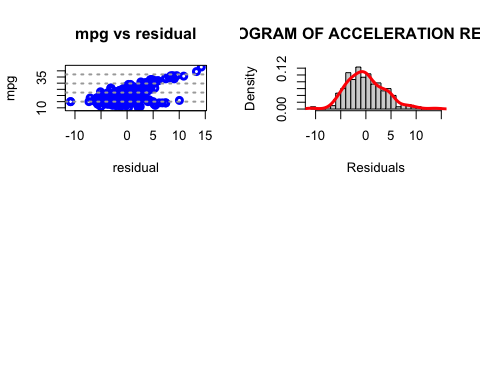
par(mfrow=c(2,2))



residual <- Linear\_Reg.horse\_power$residuals  
plot(cars.first$mpg~residual,lwd=3, col="blue",main="mpg vs residual", xlab="residual",ylab = "mpg")  
grid(NA, 5, lwd = 2,col = "darkgray")  
  
hist(residual,prob=T,breaks=20,main="HISTOGRAM OF ACCELERATION RESIDUALS",xlab="Residuals")  
lines(density(residual),col="red",lwd=3)  
  
#mpg vs acceleration  
  
Linear\_Reg.acc <- lm(mpg~acceleration, data=cars.first)  
summary(Linear\_Reg.acc)

##   
## Call:  
## lm(formula = mpg ~ acceleration, data = cars.first)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.202 -4.126 -1.012 3.268 16.154   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.0012 1.8352 2.725 0.00681 \*\*   
## acceleration 1.0379 0.1183 8.770 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.654 on 298 degrees of freedom  
## Multiple R-squared: 0.2052, Adjusted R-squared: 0.2025   
## F-statistic: 76.91 on 1 and 298 DF, p-value: < 2.2e-16

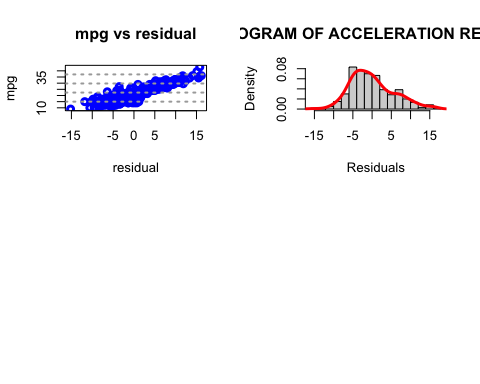
par(mfrow=c(2,2))



residual <- Linear\_Reg.acc$residuals  
plot(cars.first$mpg~residual,lwd=3, col="blue",main="mpg vs residual", xlab="residual",ylab = "mpg")  
grid(NA, 5, lwd = 2,col = "darkgray")  
  
hist(residual,prob=T,breaks=20,main="HISTOGRAM OF ACCELERATION RESIDUALS",xlab="Residuals")  
lines(density(residual),col="red",lwd=3)  
  
  
  
#mpg vs weight  
  
  
Linear\_Reg.wei <- lm(mpg~weight, data=cars.first)  
summary(Linear\_Reg.wei)

##   
## Call:  
## lm(formula = mpg ~ weight, data = cars.first)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.2011 -1.9157 -0.0812 1.7341 15.0246   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 40.5619792 0.6461532 62.77 <2e-16 \*\*\*  
## weight -0.0062905 0.0001984 -31.71 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.032 on 298 degrees of freedom  
## Multiple R-squared: 0.7714, Adjusted R-squared: 0.7706   
## F-statistic: 1005 on 1 and 298 DF, p-value: < 2.2e-16

par(mfrow=c(2,2))



residual <- Linear\_Reg.wei$residuals  
plot(cars.first$mpg~residual,lwd=3, col="blue",main="mpg vs residual", xlab="residual",ylab = "mpg")  
grid(NA, 5, lwd = 2,col = "darkgray")  
  
hist(residual,prob=T,breaks=20,main="HISTOGRAM OF ACCELERATION RESIDUALS",xlab="Residuals")  
lines(density(residual),col="red",lwd=3)  
  
  
#predicted values with remaining values  
  
predicted\_displacement <- Linear\_Reg.displace %>% predict(cars.last)  
data.frame( R2 = R2(predicted\_displacement, cars.last$mpg),  
 RMSE = RMSE(predicted\_displacement, cars.last$mpg),  
 MAE = MAE(predicted\_displacement, cars.last$mpg))

## R2 RMSE MAE  
## 1 NA NA NA

predictions\_error <- RMSE(predicted\_displacement, cars.last$mpg)/mean(cars.last$mpg)  
compare\_dis <- as.data.frame(cbind(cars.last$mpg,predicted\_displacement),row=FALSE)  
names(compare\_dis) <- c("observ","predi\_dis")  
kable(head(compare\_dis),format="pandoc", caption = "predicted values of mpg - displacement model")

predicted values of mpg - displacement model

| observ | predi\_dis |
| --- | --- |
| 34.5 | 26.21621 |
| 31.8 | 27.19446 |
| 37.3 | 26.90099 |
| 28.4 | 23.96623 |
| 28.8 | 22.89016 |
| 26.8 | 22.89016 |

predicted\_horsepower <- Linear\_Reg.horse\_power %>% predict(cars.last)  
data.frame( R2 = R2(predicted\_horsepower, cars.last$mpg),  
 RMSE = RMSE(predicted\_horsepower, cars.last$mpg),  
 MAE = MAE(predicted\_horsepower, cars.last$mpg))

## R2 RMSE MAE  
## 1 NA NA NA

predictions\_error <- RMSE(predicted\_horsepower, cars.last$mpg)/mean(cars.last$mpg)  
compare\_hors <- as.data.frame(cbind(cars.last$mpg,predicted\_horsepower),row=FALSE)  
names(compare\_hors) <- c("observ","predi\_dis")  
kable(head(compare\_hors),format="pandoc", caption = "predicted values of mpg - horsepower model")

predicted values of mpg - horsepower model

| observ | predi\_dis |
| --- | --- |
| 34.5 | 26.09584 |
| 31.8 | 26.72496 |
| 37.3 | 26.22166 |
| 28.4 | 23.57936 |
| 28.8 | 20.43376 |
| 26.8 | 20.43376 |

predicted\_weight <- Linear\_Reg.wei %>% predict(cars.last)  
data.frame( R2 = R2(predicted\_weight, cars.last$mpg),  
 RMSE = RMSE(predicted\_weight, cars.last$mpg),  
 MAE = MAE(predicted\_weight, cars.last$mpg))

## R2 RMSE MAE  
## 1 NA NA NA

predictions\_error <- RMSE(predicted\_weight, cars.last$mpg)/mean(cars.last$mpg)  
compare\_wei <- as.data.frame(cbind(cars.last$mpg,predicted\_weight),row=FALSE)  
names(compare\_wei) <- c("observ","predi\_dis")  
kable(head(compare\_wei),format="pandoc", caption = "predicted values of mpg - weight model")

predicted values of mpg - weight model

| observ | predi\_dis |
| --- | --- |
| 34.5 | 27.03751 |
| 31.8 | 27.85526 |
| 37.3 | 27.16331 |
| 28.4 | 23.76647 |
| 28.8 | 24.23825 |
| 26.8 | 23.57776 |

predicted\_acceleration <- Linear\_Reg.acc %>% predict(cars.last)  
data.frame( R2 = R2(predicted\_acceleration, cars.last$mpg),  
 RMSE = RMSE(predicted\_acceleration, cars.last$mpg),  
 MAE = MAE(predicted\_acceleration, cars.last$mpg))

## R2 RMSE MAE  
## 1 NA NA NA

predictions\_error <- RMSE(predicted\_acceleration, cars.last$mpg)/mean(cars.last$mpg)  
compare\_acc <- as.data.frame(cbind(cars.last$mpg,predicted\_acceleration),row=FALSE)  
names(compare\_acc) <- c("observ","predi\_dis")  
kable(head(compare\_acc),format="pandoc", caption = "predicted values of mpg acceleration model")

predicted values of mpg acceleration model

| observ | predi\_dis |
| --- | --- |
| 34.5 | 20.46536 |
| 31.8 | 24.92818 |
| 37.3 | 20.25778 |
| 28.4 | 21.60701 |
| 28.8 | 16.72904 |
| 26.8 | 18.38963 |

