#HW10:

Use the R program code below to generate all possible and stepwise selection procedures on the 6 predictors provided in the lmod statement.  
Submit the following output and summary into Canvas.  
generated output (graphic output is not needed)  
a short paragraph discussing the model you would choose  
#--------------------------------------------------------------------------------#  
library(faraway) #this command brings in a library of regression functions

#install.packages("olsrr") #install olsrr if package has not been installed on your computer  
library(olsrr)

data(fat,package="faraway")  
#Can we predict body fat using only easy-to-record measurements?

#use the variables specified in this model  
lmod <- lm(brozek ~ weight + neck + chest + abdom + hip + thigh, data=fat)

ols\_step\_all\_possible(lmod)  
ols\_step\_forward\_p(lmod,details=TRUE)  
ols\_step\_backward\_p(lmod,details=TRUE)  
ols\_step\_both\_p(lmod,details=TRUE)

##--------------------------------------------------------------------------------##

#HW11:  
Use Ex\_Cross\_Validation.R to perform a 10-fold cross validation to choose between the model that includes weight and abdom with the model that includes weight, abdom, and thigh.

Choose the model by RMSE, by Rsquared, and by MAE criteria. Are the choices consistent?

Submit the output (no graphics needed) and a short paragraph into Canvas summarizing the results.

Use 5-fold cross validation to decide the degree of polynomial to use for a regression of speed on distance needed to stop from the cars data set.  See details below and the Rcode you will need.  This R program, HR12.R is also loaded into Canvas under [Files Loaded after Spring Break](https://rutgers.instructure.com/courses/39556/files/folder/Files%20Loaded%20after%20Spring%20Break).  Just hand in the text output with a sentence explaining your choice of degree - no graphics files are necessary.

cars is a data frame with 50 observations on 2 variables

speed - speed(mph)

dist - stopping distance measured in feet

##--------------------------------------------------------------------------------##

#HW12

Use 5-fold cross validation to decide the degree of polynomial to use for a regression of speed on distance needed to stop from the cars data set.  See details below and the Rcode you will need.  This R program, HR12.R is also loaded into Canvas under [Files Loaded after Spring Break](https://rutgers.instructure.com/courses/39556/files/folder/Files%20Loaded%20after%20Spring%20Break).  Just hand in the text output with a sentence explaining your choice of degree - no graphics files are necessary.

cars is a data frame with 50 observations on 2 variables

speed - speed(mph)

dist - stopping distance measured in feet

------------------------------------------------------------------------------------------------------------

#HW12 Assignment and program to help you

library(faraway)  
#install.packages("caret") #install this package if needed  
library(caret)  
set.seed(13245) #use this seed

head(cars,1L)

attach(cars) #n=50  
# sorting dataset by distance for graphing purposes  
cars <- cars[order(dist),]   
cars

windows(7,7)  
plot(x=cars$dist,y=cars$speed)

##-------------------------------------------------------------##

#The researcher is interested in predicting speed based on knowing stopping distance  
#fit a polynomial to the data - use degree 1, 2, 3, or 4?  
#use cross-validation since overfitting is a concern  
#ASSIGNMENT: use 5-fold cross validation to obtain the choice of degree 1, 2, 3, or 4

#Here is the r-code for a polynomial of degree 4 and plotting the fitted curve  
#You can use the code below and just repeat for a polynomials of degree 1, 2, and 3

#Fit a polynomial of degree 4  
poly4<- lm(speed~dist+I(dist^2)+I(dist^3)+I(dist^4), data=cars)  
summary(poly4) #summary of results from fitting a polynomial of degree 4  
plot(x=cars$dist,y=cars$speed)  
lines(x=cars$dist,y=poly4$fitted, type="l", col="red")

#Compute the cross-validation metrics for degree 4  
# Define training control  
train.control <- trainControl(method = "cv", number = 5)  
# Train the model  
CVpoly4 <- train(speed~dist+I(dist^2)+I(dist^3)+I(dist^4),data = cars, method = "lm",  
trControl = train.control)  
# Summarize the results  
print(CVpoly4)  
##

#HW13  
Use the presidents data set (presidents.csv, found in Files: [presidents.csvPreview the document](https://rutgers.instructure.com/courses/39556/files/8436314/download?wrap=1) ) that shows quarterly approval ratings of US presidents during 30 years starting in 1945.

a) Generate two lowess plots using spans that you choose between 0.05 and 1.  
b) Provide 1 or 2 sentences describing the pattern for the span you choose.  
  
c) Generate two smoothing plots with normal kernel density standard deviation h between .2 and .8  
d) Provide 1 or 2 sentences describing the pattern for the h you choose.

#See useful R code below.

#read in the data which is in a csv file  
presidents <- read.csv(file="Your directory where the presidents.csv data set is stored/presidents.csv",header = TRUE)  
head(presidents)  
str(presidents)

library(faraway)  
#code to use for different spans. The one below uses a span of 0.20  
with(presidents,{  
plot(presidents ~ quarter, col=gray(0.1))  
f <- loess(presidents ~ quarter,span=0.20)  
i <- order(quarter)  
lines(f$x[i],f$fitted[i])  
})

#code to use for different choices of h. The one below uses an h  
library(sm) #install if not already installed  
with(presidents,sm.regression(x=quarter,y=presidents,h=.1))

##--------------------------------------------------------------------------------##

#HW14:

Use the presidents data set from HW13 and the R code below to plot the data and the smoothed simple moving average:

##--------------------------------------------------------------------------------##

library(smooth)  
MovingAverage <- sma(presidents$presidents,order=5,silent=FALSE)  
summary(MovingAverage)

##--------------------------------------------------------------------------------##

a) using 5 data points

b) using 10 data points

c) The first 20 data points for the presidents data are provided below.

Compute the Simple Moving Average using 5 data points at quarter =10.

|  |  |
| --- | --- |
| quarter | presidents |
| 1 | 83 |
| 2 | 87 |
| 3 | 82 |
| 4 | 75 |
| 5 | 63 |
| 6 | 50 |
| 7 | 43 |
| 8 | 32 |
| 9 | 35 |
| 10 | 60 |
| 11 | 54 |
| 12 | 55 |
| 13 | 36 |
| 14 | 39 |
| 15 | 42 |
| 16 | 55 |
| 17 | 69 |
| 18 | 57 |
| 19 | 57 |
| 20 | 51 |

##--------------------------------------------------------------------------------##

#HW15:

The USArrests data set available in R was used to perform k-means clustering.  The data set contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973. Also given is the percent of the population living in urban areas.

A data frame with 50 observations on 4 variables.

|  |  |  |  |
| --- | --- | --- | --- |
| [,1] | Murder | numeric | Murder arrests (per 100,000) |
| [,2] | Assault | numeric | Assault arrests (per 100,000) |
| [,3] | UrbanPop | numeric | Percent urban population |
| [,4] | Rape | numeric | Rape arrests (per 100,000) |

Use  the  results  which can be found here:[HW15\_Figure.pdfPreview the document](https://rutgers.instructure.com/courses/39556/files/8699915/download?wrap=1) and [HW15\_Output.txtPreview the document](https://rutgers.instructure.com/courses/39556/files/8699916/download?wrap=1)

to  answer the following questions.  No additional analyses are required.

a.  What two variables were used in the clustering?

b.  Comment on the ability of Urbancat  to  classify  observations defined by the k-means clustering algorithm.

##--------------------------------------------------------------------------------##