```
# Rudy Meza
# Machine Learning in R
# Michael Chang
# Spring
# ‘***Lab and Quiz***'
'********************************Machine Learning*************************************'
# ***Lab and Quiz***'
# '********************************Machine
Learning****************************************'
# **********************************  Lab 3.6*****************************************'
install.packages("ISLR")
install.packages("car")
library (MASS)
library (ISLR)
library(car)



# Boston DataSet from ISLR

fix(Boston )
names(Boston )

lm.fit =lm(medv~lstat)
# This command wont work because R does not recognize the variables being called...Need to either
attach or call by dataset and $.
#
# Boston$medv

lm.fit =lm(medv~lstat ,data=Boston )
attach (Boston )
 lm.fit =lm(medv~lstat)

 # By typing lm.fit, we get basic information for the model. We can also get more detail from
summary(lm.fit)


 lm.fit

 summary(lm.fit)

 names(lm.fit)

 coef(lm.fit)

 confint (lm.fit)
 # predict fundtion used to can be used to prduce condience intervals, as well as prediction intervals for
our Y and X variable.
 predict (lm.fit ,data.frame(lstat=c(5 ,10 ,15) ),
interval = 'confidence')

 predict (lm.fit ,data.frame(lstat=c(5 ,10 ,15) ),
interval = 'prediction')

# Next, commands to plot and our regression line.

plot(lstat ,medv)
abline (lm.fit)

# The abline() function can be used to draw any line, not just the least
# squares regression line. To draw a line with intercept a and slope b, we
# type abline(a,b). The lwd=3 command causes the width of the
# regression line to be increased by a factor of 3; this works for the plot()
# and lines() functions also.We can also use the pch option to create different
# plotting symbols.

abline (lm.fit ,lwd =3)
 abline (lm.fit ,lwd =3, col ="red ")
plot(lstat ,medv ,col ="red ")
plot(lstat ,medv ,pch =20)
plot(lstat ,medv ,pch ="+")
plot (1:20 ,1:20, pch =1:20)

# Diagonostic Plots can be achevied by plot function but more convienient to view using the par()
function.

par(mfrow =c(2,2))
plot(lm.fit)

 # When trying to identify outliers, one problem that can arise is when there is a potential outlier that
influences the regression model to such an extent that the estimated regression function is "pulled"
towards the potential outlier,
 # so that it isn't flagged as an outlier using the standardized residual criterion. To address this issue,
studentized residuals offer an alternative criterion for identifying outliers
plot(predict (lm.fit), residuals (lm.fit))
 plot(predict (lm.fit), rstudent (lm.fit))

par(mfrow =c(2,2))
plot(lm.fit)

plot(hatvalues (lm.fit ))
which.max (hatvalues (lm.fit))

# 3.6.3 Multiple Linear Regression
```

```
lm.fit =lm(medv~lstat+age ,data=Boston )
 summary (lm.fit)

 lm.fit =lm(medv~.,data=Boston )
 summary (lm.fit)


# vif() function, part of the car package, can be used to compute variance inflation
# factors. Most VIF?s are low to moderate for this data.

vif(lm.fit)

# age variable has a high p-vaule, which is not significant. We want to remove this from our model.

 lm.fit =lm(medv~.-age,data=Boston )
 summary (lm.fit)

 # we can use update function as well
lm.fit1=update (lm.fit , ~.-age)

# 3.6.4 Interaction Terms
# The syntax lstat:black tells R to include an interaction term between
# lstat and black. The syntax lstat*age simultaneously includes lstat, age,
# and the interaction term lstat×age as predictors; it is a shorthand for
# lstat+age+lstat:age.
summary (lm(medv~lstat *age ,data=Boston ))

# 3.6.5 Non-linear Transformations of the Predictors

lm.fit2=lm(medv~lstat +I(lstat ^2))
summary (lm.fit2)

lm.fit =lm(medv~lstat)
anova(lm.fit ,lm.fit2)

par(mfrow=c(2,2))
plot(lm.fit2)

lm.fit5=lm(medv~poly(lstat ,5))
summary (lm.fit5)

summary (lm(medv~log(rm),data=Boston ))

# 3.6.6 Qualitative Predictors

fix( Carseats )
names(Carseats )
```

```
# Given a qualitative variable such as Shelveloc, R generates dummy variables
# automatically. Below we fit a multiple regression model that includes some
# interaction terms.

lm.fit =lm(Sales~.+ Income :Advertising +Price :Age ,data=Carseats )
summary (lm.fit)

attach (Carseats )
contrasts (ShelveLoc )

# 3.6.7 Writing Functions

LoadLibraries

LoadLibraries()

LoadLibraries=function() {
  library (ISLR)
  library (MASS)
  print("The libraries have been loaded.")
  }

LoadLibraries
LoadLibraries()

# Week 2 Quiz for R
# Question 1


lm.fit =lm(medv~lstat)
summary(lm.fit)


predict (lm.fit ,data.frame(lstat=c(5 ,10 ,15) ),
interval = "prediction")

25.053

# Question 2

 lm.fit =lm(medv~.-age,data=Boston )
 summary (lm.fit)

indus

# Question 3 and 4
summary (A<-lm(medv~lstat *age ,data=Boston ))
summary (B<-lm(medv~lstat *black ,data=Boston ))

# Question 5

non_log =lm(medv~rm,data=Boston )
 summary (non_log)
log_t=lm(medv~log(rm),data=Boston )
summary(log_t)

# Question 6

lm.fit =lm(Sales~.+ Income :Advertising +Price :Age ,data=Carseats )
summary (lm.fit)
contrasts = list(ShelveLoc=contr.treatment(c("Bad", "Good", "Medium"), base = 3))
contrasts
summary(contrasts)

# shelvelocMedium
```