HOMEWORK 3

Group 5 (Shubham Pal, Pratik, Sunil Joshi)

1. **Using loops write a program in R to print out the first 15 digits of the Fibonnaci sequence. In a Fibonnaci sequence the next digit is the sum of the previous two digits. 0, 1, 1, 2, 3, 5, 8, 13….**

**Solution:**

#Printing Fibonacci series of first 15 digits

n <- 15

fibo <- numeric(n)

fibo[0]= 1

fibo[1]= 1

for(i in 3:n){

fibo[i] = fibo[i-1]+fibo[i-2]

print(paste("number is",fibo[i]))

}

2.A)  **Why should the data be partitioned into training and validation sets? What will the training set be used for? What will the validation set be used for?**

**Solution**: Data should be partitioned into training and validation sets and especially, the huge data sets to build models where one can be used to show relation between predictor variables while other one can be used to validate that predicted output. Training sets are used to build or train models, find regression coefficients for predictor variables and finding residuals. Validation sets are used for evaluation and validation of model fit on training data sets and also to it computes error.

2.B) **Fit a multiple linear regression model to the median house price (MEDV) as a function of CRIM, CHAS, and RM. Write the equation for predicting the median house price from the predictors in the model.**

**Solution:** Required regression model

Housing\_df <- read.csv("Desktop/data mining/BostonHousing.csv")

Housing\_df <- Housing\_df[1:500,]

selected.var <- c(1,4,6,13)

#Partition Data

set.seed(3)

train.index <- sample(c(1:500),350)

train.df <- Housing\_df[train.index,selected.var]

valid.df <- Housing\_df[-train.index,selected.var]

#Use of regression Model

Housing.lm <- lm(MEDV ~ .,train.df)

options(scipen = 999)

summary(Housing.lm)

Required Equation:

**MEDV = -23.6071 - 0.2613\* (CRIM) + 2.8866\*(CHAS) + 7.5081\*(RM)**

2.C) MEDV = -23.6071 - 0.2613\* (CRIM) + 2.8866\*(CHAS) + 7.5081\*(RM)

MEDV = -23.6071 - 0.2613\* (0.1) + 2.8866\*(0) + 7.5081\*(6)

MEDV = 21.4157

Median House price is $21,415.7

2.D i)There are several variables that measure the levels of industrialization, which are expected to be positively correlated, these include INDUS, NOX and TAX.We expect a positive correlation between INDUS , NOX and TAX as areas that have high proportion of non-retail business tend to have high taxes and high pollution.

2.D ii) RAD and TAX have high value of correlation (0.910288) and can be removed for avoiding potential redundancy and can cause multicollinearity.

3. **Use a while loop to investigate the number of terms required before the product 1⋅2⋅3⋅4⋅… reaches above 10 million**.

**Solution:**

m <- 1

j <- 0

while(factorial(m) < 10000000) {

m = m+1

j = j+1

}

print (paste("Number of terms required before product reaches above 10 million", (j-1)))