Activity–Simple Linear Regression

Learning outcomes: Simple Linear Regression

After completing this exercise, you should be able to understand and perform below tasks.

1. Compute covariance and correlations
2. Building predictive regression model using linear regression technique
3. Understanding results of regression output and evaluating them
4. Evaluation of error metrics
5. Applying the models on un-seen data
   a. Splitting data into train and test data sets
   b. Build the model on train data
   c. Apply on test data and get the predictions
   d. Study the error metrics

I. Find the covariance of the eruption duration and waiting time in the data set faithful (built-in dataset in R). Observe if there is any linear relationship between the two variables.

II. On the same data faithful, compute the correlation coefficient of eruptions and waiting. What does it signify?

III. **Business Problem:** Predicting the price of Used Toyota Corolla Automobiles

A large Toyota car dealership offers purchasers of new Toyota cars the option to buy their used car as part of a trade-in. In particular, a new promotion promises to pay high prices for the used Toyota Corolla cars for purchasers of a new car. The dealers then sells the used car for a small profit. To ensure a reasonable profit, the dealer needs to be able to predict the price that the dealership will get for the used cars. For that reason, data were collected on all previous sales of used Toyota Corollas at the dealership. The goal is to predict the price of a used Toyota Corolla based on its age. Age is given in months

Steps:
Statement: Consider the Toyota corolla cars data for analysis "Toyota_SimpleReg.csv". "Price" is the target attribute. Fit a linear regression model. Review all the statistics.

1. Import the data into R
2. Check the summary and structure of the data;
3. Perform any data preprocessing steps required.
4. Plot the data and find the correlation between the attributes.
   To build a simple linear regression model, use X = Age_06_15 and Y = price.
5. Fit the linear regression

   LinReg<-lm(<dependent variable>~<independent variable>,data=<dataset name>)

6. Check the coefficient values

   ```
   coefficients(LinReg)
   ```

7. Estimate dependent variable value for given independent variable value

   ```
   #Estimate price, if Age_06_15= 60
   ```

```
as.numeric(coefficients(LinReg)[1]+coefficients(LinReg)[2]* 60)

#Predict with the help of ‘predict’ function

testdata = data.frame(Age_06_15 =60)
testdata
predict(LinReg, testdata)
```

8. Find the confidence limits value for the predicted value at 0.95 significance level

```
predict(LinReg, testdata, interval="confidence",level=0.95)
```

9. Summarize statistics of the linear regression model

```
summary(LinReg)
```

10. Get the predictions and confidence limits on the data set

```
Pred<-data.frame(predict(LinReg, data, interval="confidence" ,level=0.95))
names(Pred)
```

11. Plotting data, fitted line and confidence limits

```
plot(data$ Age_06_15, data$price)
points(data$ Age_06_15,Pred$fit,type="l", col="red", lwd=2)
points(data$ Age_06_15,Pred$lwr,pch="-", col="blue", lwd=6)
points(data$ Age_06_15,Pred$upr,pch="-", col="blue", lwd=6)
```

12. Residual plot of the model

```
par(mfrow=c(2,2))
plot(LinReg)
```

13. Building the model on train data and evaluating on test data

```
#Split the data into train and test data sets
rows=seq(1,nrow(data),1)
set.seed(123)
trainRows=sample(rows,(70*nrow(data))/100)
train = data[trainRows,]
test = data[-trainRows,]

# building the model on train data
LinReg1<-lm(Price~Age_06_15,data=train)
summary(LinReg1)

#Error Metrics on train data & test data
library(DMwR)
regr.eval(train$price, LinReg$fitted.values)
#Error verification on test data
Pred<-predict(LinReg,test)
regr.eval(test$price, Pred)
```

IV. Assignment :

Statement :

Consider the dataset Data.csv consists of the countrywise 'BigMacPrice' and wages. Predict 'NetHourlyWage' based on the 'BigMacPrice'

a. Estimate dependent variable value for given independent variable value
Estimate price, if 'BigMacPrice'= 3.33

b. Split the data into train and test datasets and build the model on train data.

c. Error metrics evaluation on train data and test data