CSE 7202c – Statistics and Probability in Decision Modeling

Activity Sheet-Time Series Analysis & Regression

Learning outcomes:

After completing this exercise, you should be able to understand and perform below tasks.
1. Building predictive Time Series & Regression model
2. Evaluation of error metrics
3. Applying the models on un-seen data
a. Splitting data into train and test data sets
b. Comparing the error metrics
4. Interpretation of the results

Problem Statement Time Series:

1. A Global F&B Food chain manufactures an item which is then circulated to all its merchandize\branches stores for sale. The company wants to come up with the best estimate to number to be manufactured weekly, You have been given day wise units production for 2 years, you are expected to build a model which can forecast the weekly sales from the given data. Use the appropriate frequency and preprocessing steps to arrive at weekly forecast (Example Jan month 2nd week forecast is X). Come up with appropriate error metric and finalize your model.

Problem Statement Regression:

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 specifications.

Steps:
1. Consider the Toyota corolla cars data for analysis
2. In excel, fit a linear regression model. Review all the statistics.
3. Import the data into R
4. Check the summary and structure of the data and any preprocessing required.
5. Plot data and find the correlation between the attributes. To build a simple linear regression model, use X = Age_06_15 and Y = price.
6. Fit the linear regression
LinReg<-lm(<dependent variable>~<independent variable>,data=<dataset name>)
7. Check the coefficient values
coefficients(LinReg)
8. Estimate dependent variable value for given independent variable value
#Estimate price, if 'age' = 66

```
as.numeric(coefficients(LinReg)[1]+coefficients(LinReg)[2]* 4.5)
#With the help of ‘predict’ function
testdata = data.frame('age_06_15' =4.5)
testdata
predict(LinReg, testdata)
```
9. Summarize inferential 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. Error metrics evaluation on train data and 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,]
```
CSE 7202c – Statistics and Probability in Decision Modeling
Inspire…Educate…Transform. 3

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