20180902_ Batch 47_CSE 7302c_SimpleLinearRegression_Activity
_____

Objective:

In this session, you will learn how to build your first simple linear regression model, to validate the model assumptions and the model, and to interpret the results and the model evaluation metrics.

Key takeaways:

•Building univariate linear predictive model using the function lm()

•Reading diagnostic plots to test the linear regression assumptions

•Identifying influential observations and handling them

•Data Pre-processing

- Binning- Converting a numeric attribute to categorical

- Dummies- Converting a categorical variable to numeric o Standardization- Effect of scaling the data

- Imputation- Handling the missing values

- Train-Test splitting of the data

- Using Correlation plots, correlation values to identify appropriate variables for model building

Problem 1: Lab
Toyota_SimpleReg.csv

- We have provided a dataset with some details about cars. You are expected to build a linear regression model for the age and the price columns, perform all the relevant steps mentioned above.

Problem 2: Assignment
CustomerData.csv

- A large child education toy company (company name is confidential, and data is masked) which sells edutainment tablets and gaming systems both online and in retail stores in the US wanted to analyze the customer data. They are operating from last 15 years and maintaining all transactional information data. The given data 'CustomerData.csv' is a sample of customer level data extracted and processed for the analysis from various set of transactional files. Using this data, they want us to understand the life time value of each customer (LTV). This will enable them to design marketing strategies and customize the product offerings. The objective of activity is building a regression model to predict the customer revenue based on one factor that influences revenue the most.

- Read the given "CustomerData.csv" data into R.

- Are there any missing values in the data? If there are, then impute using central imputation method.

- The target for this problem is "Total Revenue generated".

- Select one most influencing numeric variable as predictor for predicting the revenue generated. How would you do this?

- Build a linear regression model to predict the target with the selected variable.

- Check the residual plots and report your observations.