DATA SCIENCE FOR ENGINEERS

Week-6

R Questions

1. How do we get Summary of the model using Linear regression?

   Answer:

   The summary(model) command will provide all the values call, residual value, coefficients.

2. How to normalize the data using the scaling function?

   Answer:

   data[“column”] <- scale(data[“column”])

   You can replace the names according to your data frame and column names.

3. logisfit<-glm(formula = crashTest_1$CarType~.,family = 'binomial', data = crashTest_1)

   Error in eval(family$initialize) : y values must be 0 <= y <= 1.

   Getting error while using this code.

   Answer:

   Convert the car type to 0 and 1 before fitting the model

Multiple Linear Regression

1. When do we use Multiple Linear Regression?

   Answer:

   Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables.

2. What is P value?

   Answer:

   In statistics, the **p-value** is the probability of obtaining results as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct.

3. When does Multicollinearity exist?

   Answer:

   Multicollinearity exists when two or more of the predictors in a regression model are moderately or highly correlated.

4. What is Residual?

   Answer:

   Residuals. The difference between the observed value of the dependent variable (y) and the predicted value (ŷ) is called the **residual** (e). Each data point has one **residual**. **Residual** = Observed value - Predicted value.

Outliers

5. How do we detect outliers?

   Answer:

   Using Z-scores to **Detect Outliers**

   Z-scores can quantify the unusualness of an observation when your data follow the normal distribution. Z-scores are the number of standard deviations above and below the mean that each value falls.