DATA SCIENCE FOR ENGINEERS

Week-7

R Questions:

1. What is the use of set.seed()?

   Answer:

   The **set**. **seed**() function **sets** the starting number used to generate a sequence of random numbers – it ensures that you get the same result if you start with that same **seed** each time you run the same process.

2. While doing the regression on R studio, I am getting only intercept table and not all the information?

   Answer:

   When you use the summary(model_name), it displays all the information.

Performance Measures

1. What is ROC?

   Answer:

   A receiver operating characteristic curve, or **ROC** curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.

2. What is Sensitivity?

   Answer:

   Sensitivity (also called the true positive rate, or the recall in some fields) measures the proportion of actual positives which are correctly identified as such (e.g., the percentage of sick people who are correctly identified as having the condition), and is complementary to the false negative rate. Sensitivity= true positives/(true positive + false negative)

3. What is specificity?

   Answer:

   Specificity (also called the true negative rate) measures the proportion of negatives which are correctly identified as such (e.g., the percentage of healthy people who are correctly identified as not having the condition), and is complementary to the false positive rate.

   Specificity=true negatives/(true negative + false positives)

4. What is Maximum Likelihood Estimation?

   Answer:

   In statistics, maximum likelihood estimation (MLE) is a method of estimating the parameters of a probability distribution by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable.

5. How to increase the performance of Logistic regression model ?

   Answer:

   To increase the performance of the model

   - Treat missing and outlier values
   - Feature selection
   - Cross validation
   - Parameter tuning

Cross Validation

1. Is there any real-world problem where we can use the k- fold cross validation?

   Answer:

   Basically, if you have less data samples you can apply k - fold cross validation. It is a resampling procedure used to evaluate on machine learning models. If you have any real time data, you can apply and see how it works.