1. Explain the Linear Regression in detail.

   In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

   Equation of linear regression is y = m1x1 + m2x2 + ---------- mnxn (Up to n terms)

2. What are assumptions of linear Regression regarding residuals?

   Linear relationship
   Multivariate normality
   No or little multicollinearity
   No auto-correlation
   Homoscedasticity

3. What is Coefficient of correlation and coefficient of Determination?

   A correlation coefficient is a numerical measure of some type of correlation, meaning a statistical relationship between two variables. The variables may be two columns of a given data set of observations, often called a sample, or two components of a multivariate random variable with a known distribution

   In statistics, the coefficient of determination, denoted R^2 or r^2 and pronounced "R squared", is the proportion of the variance in the dependent variable that is predictable from the independent variable.

4. Explain Anscombe's Quartet in detail?

   Anscombe's quartet comprises four data sets that have nearly identical simple descriptive statistics, yet have very different distributions and appear very different when graphed. Each dataset consists of eleven (*x*,*y*) points. They were constructed in 1973 by the statisticianFrancis Anscombe to demonstrate both the importance of graphing data before analyzing it and the effect of outliers and other influential observations on statistical properties. He described the article as

being intended to counter the impression among statisticians that "numerical calculations are exact, but graphs are rough.

5. What is Pearson's R?

   In statistics, the Pearson correlation coefficient, also referred to as Pearson's r, the Pearson product-moment correlation coefficient or the bivariate correlation, is a measure of the linear correlation between two variables X and Y.

6. What is Scaling? Why scaling is performed? What is difference between normalized scaling and standardized scaling?

   In statistics, the Pearson correlation coefficient, also referred to as Pearson's r, the Pearson product-moment correlation coefficient or the bivariate correlation, is a measure of the linear correlation between two variables X and Y

   The goal of scaling is to change the values of numeric columns in the dataset to use a common scale, without distorting differences in the ranges of values or losing information. scaling is also required for some algorithms to model the data correctly.

   Normalized scaling gives range [0,1]
   Standardized Scaling gives range [-1, 1]

7. Why VIF Value becomes infinite?

   If R2 == 1 VIF = 1/0 = infinity

8. What is Gauss-markov Theorm?

   n statistics, the Gauss–Markov theorem states that in a linear regression model in which the errors are uncorrelated, have equal variances and expectation value of zero, the best linear unbiased estimator (BLUE) of the coefficients is given by the ordinary least squares (OLS) estimator, provided it

exists. Here "best" means giving the lowest variance of the estimate, as compared to other unbiased, linear estimators. The errors do not need to be normal, nor do they need to be independent and identically distributed (only uncorrelated with mean zero and homoscedastic with finite variance). The requirement that the estimator be unbiased cannot be dropped, since biased estimators exist with lower variance

9. What is Gradient Descent Algorithm in Detail?

   Gradient Descent is an optimization algorithm used for minimizing the cost function in various machine learning algorithms. It is basically used for updating the parameters of the learning model.

10. What is Q-Q Plot? Explain use and importance of Q-Q Plot in Linear Regression?

    In statistics, a Q–Q plot is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other.