1. Explain the linear regression algorithm in detail.

Linear Regression is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.

2. What are the assumptions of linear regression regarding residuals?
   Ans:

Homoscedastic is assumption in linear regression which explains **residuals** are equal across the **regression** line.

3. What is the coefficient of correlation and the coefficient of determination?
   Ans:

A correlation coefficient is a numerical measure of some type of correlation, meaning a statistical relationship between two variables. Whereas **Coefficient of Determination** is the square of Coefficeint of **Correlation**. R square or coeff. of **determination** shows percentage variation in y which is explained by all the x variables together

4. Explain the Anscombe's quartet in detail.
   Ans:
   Anscombe's quartet comprises four datasets that have nearly identical simple statistical properties, yet appear very different when graphed. Each dataset consists of eleven (x,y) points. They were constructed in 1973 by the statistician Francis Anscombe to demonstrate both the importance of graphing data before analysing it and the effect of outliers on statistical properties

5. What is Pearson's R?
   Ans:

Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or correlation, as well as the direction of the relationship.

6. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?
   Ans:

Scaling is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or values or units. ... So, we use Feature **Scaling** to bring all values to same magnitudes and thus, tackle this issue.

Normalization usually means to scale a variable to have a values **between** 0 and 1, while **standardization** transforms data to have a mean of zero and a standard deviation of 1.

7. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
   Ans:

An **infinite VIF value** indicates that the corresponding variable may be expressed exactly by a linear combination of other variables (which show an **infinite VIF** as well)

8. What is the Gauss-Markov theorem?
   Ans:

In statistics, the **Gauss–Markov theorem** states that the ordinary least squares (OLS) estimator has the lowest sampling variance within the class of linear unbiased estimators, if the errors in the linear regression model are uncorrelated, have equal variances and expectation value of zero.

9. Explain the gradient descent algorithm in detail.
   Ans:

Gradient Descent is the most common optimization algorithm in *machine learning* and *deep learning*. It is a first-order optimization algorithm. This means it only takes into account the first derivative when performing the updates on the parameters. On each iteration, we update the parameters in the opposite direction of the gradient of the objective function *J(w)* w.r.t the parameters where the gradient gives the direction of the steepest ascent. The size of the step we take on each iteration to reach the local minimum is determined by the learning rate α. Therefore, we follow the direction of the slope downhill until we reach a local minimum.

10. What is a Q-Q plot? Explain the use and importance of a Q-Q plot in linear regression.
    Ans:

The quantile-quantile (q-q) plot is a graphical technique for determining if two data sets come from populations with a common distribution. A q-q plot is a plot of the quantiles of the first data set against the quantiles of the second data set.

A q-q plot is a plot of the quantiles of the first data set against the quantiles of the second data set. By a quantile, we mean the fraction (or percent) of points below the given value. That is, the 0.3 (or 30%) quantile is the point at which 30% percent of the data fall below and 70% fall above that value