1. Explain the linear regression algorithm in detail.
   Linear Regression algorithm is used to find correlation formula between one or more independent variables and a dependent variable (also called target variable) and this is possible only when there is linear relationship between the independen variables and the dependent variable.
2. What are the assumptions of linear regression regarding residuals?
   1. Linear relationship between X and Y
   2. Error terms are normally distributed (not X, Y)
   3. Error terms are independent of each other
   4. Error terms have constant variance (homoscedasticity)

3. What is the coefficient of correlation and the coefficient of determination?
   Coefficient of correlation: The correlation coefficient is a statistical measure of the strength of the relationship between the relative movements of two variables. The values range between -1.0 and 1.0. A calculated number greater than 1.0 or less than -1.0 means that there was an error in the correlation measurement.
   Coefficient of determination: This is R square, which is used to assess a LM model. R2 is a number which explains what portion of the given data variation is explained by the developed model. it is represented as: $R^2 = 1 - (RSS / TSS)$. RSS – Residual Summed Square, TSS – Total Summed Square.
4. Explain the Anscombe's quartet in detail.
   It consists of four datasets, having identical descriptive values, but having different in when visualized. Each dataset consists of eleven (x,y) points.
5. What is Pearson's R?
   Pearson's correlation coefficient (r) is a measure of the strength of the association between the two variables. Pearson's correlation coefficient (r) for continuous (interval level) data ranges from -1 to +1.
6. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?
   Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. If feature scaling is not done, then a machine learning algorithm tends to weigh greater values, higher and consider smaller values as the lower values, regardless of the unit of the values.
7. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
   An infinite VIF value indicates that the corresponding variable may be expressed exactly by a linear combination of other variables. If the correlation coefficient R values are equal to 1, then VIF will be infinite

8. What is the Gauss-Markov theorem?
   The Gauss-Markov theorem states that if your linear regression model satisfies the first six classical assumptions, then ordinary least squares (OLS) regression produces unbiased estimates that have the smallest variance of all possible linear estimators. The Gauss-Markov theorem famously states that OLS is BLUE. BLUE is an acronym of "Best Linear Unbiased Estimator"

9. Explain the gradient descent algorithm in detail.

   Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model. Parameters refer to coefficients in Linear Regression and weights in neural networks.

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

    Q Q Plots (Quantile-Quantile plots) are plots of two quantiles against each other. A quantile is a fraction where certain values fall below that quantile. For example, the median is a quantile where 50% of the data fall below that point and 50% lie above it. The purpose of Q Q plots is to find out if two sets of data come from the same distribution. A 45 degree angle is plotted on the Q Q plot; if the two data sets come from a common distribution, the points will fall on that reference line.