1. Explain the linear regression algorithm in detail.

   Regression is a method of modelling a target value based on independent predictors. This method is mostly used for forecasting and finding out cause and effect relationship between variables. Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables.

   $Y = b0+b\_1*x$

   Then we use the cost function which would provide us the best possible value for b0 and and b_1. In this way we convert this search problem into minimization problem.By this we would minimize the problem between the predicted and the actual value.

   The difference between the predicted and the actual value . We square the error difference and sum over all data points and divide by the total number.

   Another way of doing that is using gradient descent which uses the back propogation algorithm.

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

   Linearity: The relationship between X and the mean of Y is linear.

   Homoscedasticity: The variance of residual is the same for any value of X.

   Independence: Observations are independent of each other.

   Normality: For any fixed value of X, Y is normally distributed.

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

   Coefficient of correlation is “R” value which is given in the summary table in the Regression output. R square is also called **coefficient of determination**. Multiply R times R to get the R square value. In other words **Coefficient of Determination** is the square of Coefficeint of **Correlation**.

4. Explain the Anscombe’s quartet in detail.

   Anscombe's quartet comprises four datasets

   The first scatter plot (top left) appeared to be a simple linear relationship, corresponding to two variables correlated and following the assumption of normality.

   The second graph (top right) was not distributed normally; while a relationship between the two variables is obvious, it is not linear, and the Pearson correlation coefficient is not relevant. A more general regression and the corresponding coefficient of determination would be more appropriate.

   In the third graph (bottom left), the distribution is linear, but should have a different regression line (a robust regression would have been called for). The calculated regression is offset by the one outlier which exerts enough influence to lower the correlation coefficient from 1 to 0.816.

Finally, the fourth graph (bottom right) shows an example when one high-leverage point is enough to produce a high correlation coefficient, even though the other data points do not indicate any relationship between the variables.

5. What is Pearson's R?
   Strength of linear relation between the numerical variables.

   The Pearson's correlation coefficient varies between -1 and +1 where:

   r = 1 means the data is perfectly linear with a positive slope (i.e., both variables tend to change in the same direction)
   r = -1 means the data is perfectly linear with a negative slope (i.e., both variables tend to change in different directions)
   r = 0 means there is no linear association
   r > 0 < 5 means there is a weak association
   r > 5 < 8 means there is a moderate association
   r > 8 means there is a strong association

6. What is scaling? Why is scaling performed? What is the difference between normalized scaling and standardized scaling?
   In Scaling we transform the data such that the features are within a specific range e.g. [0, 1]. Scaling is important where distance between the features is important.

   The point of normalization is to change your observations so that they can be described as a normal distribution.

   Normal distribution (Gaussian distribution), also known as the **bell curve**, is a specific statistical distribution where a roughly equal observations fall above and below the mean, the mean and the median are the same, and there are more observations closer to the mean.

7. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
   When there is a perfect correlation between the variables then VIF=~Infinity. It simply indicates that variable can be predicted out simply using the linear combination of other variables.

8. What is the Gauss-Markov theorem?
   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.
   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 alpha. Therefore, we follow the direction of the slope downhill until we reach a local minimum.
Let's first see how gradient descent works
1. Initialize weight w and bias *b* to any random numbers.
2. Pick a value for the learning rate α. The learning rate determines how big the step would be on each iteration.
plot the cost function against different values of alpha and pick the value of alpha that is right before the first value

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

    (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.