1. **Explain the linear regression algorithm in detail.**

   Ans:

   - *Linear Regression* is the statistical method based on supervised method to determine the strength and character between one dependent variable and series of independent variables.
   - Linear regression is used in many different fields, including finance, economics, and psychology, to understand and predict the behavior of a particular variable. For example, in finance, linear regression might be used to understand the relationship between a company's stock price and its earnings, or to predict the future value of a currency based on its past performance.
   - The equation of a linear regression (LR) is in the form of a straight line i.e.:

     $$y = mx + c$$

     y = dependent variable (output or response variable)
     x = independent variable (predictor variable)
     m = slope of the line
     c = intercept on the y-axis
   - The null hypothesis for LR is $\beta 1 = 0$, where $\beta 1$ is coefficient of predictor variable which if failed to reject then model is insignificant. The regression line is the best fit line for our model.

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

   Ans:

   - The residuals or errors are normally distributed and have mean(μ) as 0.
   - They are independent of each other.
   - They maintain homoscedasticity (i.e., constant in variance).

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

   Ans:

   - A *correlation coefficient* is a statistical measure of the degree to which changes to the value of predictor variable predict change to the value of response variable and ranges from -1 to 1.
   - The *coefficient of determination* is the R squared value which represents the proportion of the variance in the dependent variable that is predicted from the independent variable. It ranges from 0 to 1.

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 analyzing it and the effect of outliers on statistical properties.
   - Anscombe's quartet tells us about the importance of visualizing data before applying various algorithms to build models.
   - This suggests the data features must be plotted to see the distribution of the samples that can help you identify the various anomalies present in the data (outliers, diversity of the data, linear separability of the data, etc.).
   - Moreover, the linear regression can only be considered a fit for the data with linear relationships and is incapable of handling any other kind of data set.

5. **What is Pearson's R?**

   Ans:

   - The Pearson correlation coefficient (r) is a way of measuring a linear correlation. It ranges from –1 to 1 that measures the strength and direction of the relationship between two variables i.e., the dependent and independent variables.
   - The 0 < coefficient <1 shows positive correlation.
   - The coefficient = 0 shows no relationship.
   - The -1 < coefficient <0 shows negative correlation.

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

   Ans:

   - In machine learning, *feature scaling* refers to putting the feature values into the same range. Scaling is extremely important for the algorithms considering the distances between observations like k-nearest neighbors.
   - Scaling just affects the coefficients, and not any other parameters, such as t-statistic, F-statistic, p-values and R-squared.
   - In *Normalization (min-max) scaling*, we map the feature values into the [0, 1] range:

     x = (x – min(x))/(max(x) – min(x))

- In *Standardization scaling*, we don't enforce the data into a definite range. Instead, we transform to have a mean(μ) of 0 and a standard deviation(sd) of 1:

  x = (x - μ(x))/sd(x)

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

   Ans:

   - If there is perfect correlation, then *VIF = infinity*. This shows a perfect correlation between two independent variables.
   - In the case of perfect correlation, we get R2 =1, which lead to ***VIF=1/(1-R2)*** to infinity. To solve this problem, we need to drop one of the variables from the dataset which is causing this perfect multicollinearity.
   - 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:

   - *The Gauss–Markov theorem* or simply Gauss 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.
   - 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. **Explain the gradient descent algorithm in detail.**

   Ans:

   - *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.
   - In LR, the cost function = Residual sum of squares (RSS) which has to be minimized.
   - **RSS = $(y_{act} - y_{pred})^2$**

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

Ans:

- *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.
- A Q–Q plot is used to compare the shapes of distributions, providing a graphical view of how properties such as location, scale, and skewness are similar or different in the two distributions.