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

Linear Regression is a machine learning algorithm for Supervised Learning (Here we get previous years data with labels to build the model). This algorithm models a target value which is continuous variable (also known as predictor) based on one or more independent variables by finding a linear relationship between them.

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y (output). Hence, the name is Linear Regression.

Linear regression can be used for prediction as well as forecasting.

Hypothesis function for Linear Regression

$Y = B0 + B1.x$

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

Assumptions of linear regression for residuals:

- *The regression model is linear in the coefficients and the error term.*
- *Error terms are normally distributed.*
- *Error terms are independent of each other.*
- *Error terms have constant variance (homoscedasticity).*

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

Coefficient of correlation:

It is a statistical relationship between two variables. The correlation values are in the range from −1 to +1, where ±1 indicates the strongest possible agreement and 0 the strongest possible disagreement.

Coefficient of determination:

R2 is a number which explains what portion of the given data variation is explained by the developed model. It always takes a value between 0 and 1. In general term, it provides a measure of how well actual outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model, i.e. expected outcomes. Overall, the higher the R-squared, the better the model fits the data.

4. **Explain the Anscombe's quartet in detail.**

Anscombe's Quartet was developed by statistician Francis Anscombe. It comprises four datasets, each containing eleven (x,y) pairs. These datasets share the same descriptive statistics. But things change completely when they are graphed. Each graph tells a different story irrespective of their similar summary statistics.

Anscombe's quartet demonstrate both the importance of graphing data before analyzing it and the effect of outliers and other influential observations on statistical properties.

5. What is Pearson's R?

The Pearson correlation coefficient is a measure of the strength of the linear relationship between two variables. It is referred to as Pearson's correlation or simply as the correlation coefficient. If the relationship between the variables is not linear, then the correlation coefficient does not adequately represent the strength of the relationship between the variables.

The symbol for Pearson's correlation is "ρ" when it is measured in the population and "r" when it is measured in a sample.

Pearson's r can range from -1 to 1. An r of -1 indicates a perfect negative linear relationship between variables, an r of 0 indicates no linear relationship between variables, and an r of 1 indicates a perfect positive linear relationship between variables.

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

Scaling is to change the values of independent variables within a specified range of values.

Scaling is required when there are lot of independent variables in the model and they may be on different scales, for eg, some variables may be in hundreds, others may be in lakhs. Due to this there will be a difficulty in interpreting the model.

Reasons for feature scaling:

- *Ease of interpretation*
- *Faster convergence for gradient descent methods*

The two scaling methods are:

***Normalizing/MinMaxScaling**: The variables are scaled in such a way that all the values lie between 0 and 1 using the maximum and the minimum values in the data.*
x = (x- min(x)) / (max(x) – min(x))

***Standardizing:** The variables are scaled in such a way that their mean is 0 and standard deviation Is 1.*
x = (x- mean(x)) / sd(x)

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

Variance inflation factors show the degree to which a regression coefficient will be affected because of the variable's redundancy with other independent variables. As the squared multiple correlation of any predictor variable with the other predictors approaches unity, the corresponding VIF becomes infinite.

VIF=1/(1-R2)

So when R=1, VIF=infinite

8. What is the Gauss-Markov theorem?

The Gauss-Markov theorem states that if the 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 does not state that these are just the best possible estimates for the OLS procedure, but the best possible estimates for any *linear model estimator.*

The Gauss-Markov theorem states that OLS is BLUE. BLUE is an acronym for ***Best Linear Unbiased Estimator.*** *The definition of "best" refers to the minimum variance or the narrowest sampling distribution. When the model satisfies the assumptions, OLS coefficient estimates follow the tightest possible sampling distribution of unbiased estimates compared to other linear estimation methods.*

9. Explain the gradient descent algorithm in detail.

Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point. If, instead, one takes steps proportional to the positive of the gradient, one approaches a local maximum of that function; the procedure is then known as gradient ascent. The size of step is called Learning rate denoted by alpha.

- *If α is very small, it would take long time to converge and become computationally expensive.*
- *If α is large, it may fail to converge and overshoot the minimum.*

In machine/deep learning terminology, it's the task of minimizing the cost/loss function J(w) parameterized by the model's parameters w ∈R^d. Optimization algorithms (in case of minimization) have one of the following goals:

- *Find the global minimum of the objective function. This is feasible if the objective function is convex, i.e. any local minimum is a global minimum.*
- *Find the lowest possible value of the objective function within its neighborhood. That's usually the case if the objective function is not convex as the case in most deep learning problems.*

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 data sets come from a common distribution, the points will fall on that reference line.

Q-Q plot is useful in linear regression because one of the assumptions in linear regression is that the error terms are normally distributed and this plot is useful to find whether a error term is normally distributed or not. This will help in making precise inferences.