1. Explain the linear regression algorithm in detail.

Machine Learning is a branch of Artificial intelligence that focuses on the development of algorithms and statistical models that can learn from and make predictions on data. Linear regression is also a type of machine-learning algorithm more specifically a supervised machine-learning algorithm that learns from the labelled datasets and maps the data points to the most optimized linear functions. which can be used for prediction on new datasets.

First of we should know what supervised machine learning algorithms is. It is a type of machine learning where the algorithm learns from labelled data. Labeled data means the dataset whose respective target value is already known. Supervised learning has two types:

- **Classification**: It predicts the class of the dataset based on the independent input variable. Class is the categorical or discrete values. like the image of an animal is a cat or dog?
- **Regression**: It predicts the continuous output variables based on the independent input variable. like the prediction of house prices based on different parameters like house age, distance from the main road, location, area, etc.

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

Linear regression relies on several assumptions about the residuals (errors):

1. Linearity: The relationship between the independent and dependent variables should be linear.
2. Independence: The residuals should be independent of each other.
3. Homoscedasticity: The residuals should have constant variance across all levels of the independent variables. This means there should be no patterns when residuals are plotted against predicted values.
4. Normality: The residuals should be approximately normally distributed.

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

- Coefficient of Correlation (r): This measures the strength and direction of the linear relationship between two variables. It ranges from -1 to 1, where:

- $r=1$: Perfect positive correlation,
- $r=-1$: Perfect negative correlation,
- $r=0$: No linear correlation.

- - Coefficient of Determination R^2: This indicates how well the independent variables explain the variability in the dependent variable. It ranges from 0 to 1:
- R^2=1: Perfect fit,
- R^2= 0: The model explains none of the variability.

4. Explain the Anscombe's quartet in detail.

Anscombe's quartet comprises four datasets that have nearly identical statistical properties (mean, variance, correlation, and regression line) but differ significantly in their graphical appearance. This illustrates the importance of visualizing data before relying solely on statistical summaries. The datasets show how outliers, non-linearity, and other patterns can exist even if the summary statistics are similar.

5. What is Pearson's R?

Pearson's R, also known as Pearson's correlation coefficient, measures the linear correlation between two variables. It is calculated as:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

It indicates both the strength and direction of a linear relationship between two variables.

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

Scaling is the process of transforming data so that the features have similar scales. This is important for algorithms like gradient descent and distance-based models, where differences in scales can lead to biased models.

- Normalized Scaling: Transforms the data into a range, typically [0, 1].

- Standardized Scaling: Transforms the data so that it has a mean of 0 and a standard deviation of 1.

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

An infinite VIF indicates perfect multicollinearity, meaning one variable is a perfect linear combination of others. This happens when there is redundant or highly correlated data, making the regression model unstable.

8. What is the Gauss-Markov theorem?

The Gauss-Markov theorem states that in a linear regression model, the ordinary least squares (OLS) estimator is the Best Linear Unbiased Estimator (BLUE) of the coefficients, provided the assumptions of the linear regression model are satisfied.

9. Explain the gradient descent algorithm in detail.

Gradient descent is an optimization algorithm used to minimize the cost function by iteratively updating the model parameters in the opposite direction of the gradient of the cost function.

Steps in Gradient Descent:

1. Initialize the parameters.
2. Calculate the gradient of the cost function with respect to the parameters.
3. Update the parameters using:

$$\theta = \theta - \alpha \cdot \nabla J(\theta)$$

where α is the learning rate.

4. Repeat until convergence.

10. Explain the gradient descent algorithm in detail.

A Q-Q plot (Quantile-Quantile plot) is a graphical tool to assess if a dataset follows a particular distribution, usually the normal distribution. In linear regression, a Q-Q plot of the residuals helps check the normality assumption, indicating if the residuals are normally distributed or if there are deviations.