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

- Based on **supervised learning**, mainly used for forecasting and find out relationship between variables.
- Regression aims at finding relationship between **Dependent (Target) and Independent (Predictor)** variables.
- Assumes input and output **variables possesses linear relation** i.e., output variable is combination of input variable.
- Used to predict values in **continuous range** instead of **categories**.
- Regression is mainly of **2 types**: Simple Linear Regression and Multiple Linear Regression
- **Simple linear Regression**: finding **best fit line** that explains relationship between dependent and independent **variables linearly**. Here number of independent variables is one. Represented as below:

  y = m*x + c, where m is the slope and c is the intercept.

  Algorithm aim is to find the best value of m and c, so that y can be predicted for a value of x (within the training set range).

  It tries to minimize the cost function (RSS- Residual sum of squares) using Ordinary least square method, this can be done either using Gradient descent or Differentiation method.

- **Multiple regression**: finding **relationship** between **multiple independent variables** with dependent (**response) variable**.

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

Residuals are the **error terms** and is formulated as:

Residuals = measured value – Predicted value

And denoted as, $\mathbf{e_i = y_i - y_p}$

- **Normality**: it says, the error terms must be **distributed normally** i.e., residuals must not be skewed.
- **Homoscedasticity**: having **same variance** or equality of variance.
- **Independence**: there is **no pairwise covariance** which means error terms are independent of each other.
- **Zero mean**: normally distributed **error terms around zero**. That is error terms does not increase or decrease across values.

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

a. **Coefficient of correlation** is the linear relationship that calculates direction and strength between two variables, denoted as 'R'. The value of R can lie between -1 to 1. Well, 1 indicates that the variables are perfectly correlated means increasing and decreasing together. And -1 indicates they opposes each other which means one variable tends to

increase and other tends to decrease. Whereas, 0 means there is no correlation amongst them.

- For example, value of R is 0.9 and 0.1, where R = 0.9 describes strong positive correlation and R= 0.1 depicts weak correlation.

b. Coefficient of determination can be said as square of coefficient of correlation. It defines how good the model is. It is the variation in y-axis caused due to variation in x collectively and denoted as R-square. Since it is square of R so the value can never be negative and lies between 0 and 1. The value 1 indicates higher variation which is better one whereas 0 indicates lower variation.

- So, R-square 0 indicates that y(target) variable cannot be predicted by the X (predictor variables) whereas 1 indicate that predication is very good. Value between 0-1 indicates the extend to which value can be predicted without error.
- For example, if R-square is 0.845 and 0.765, then higher value which is 0.845 is better than 0.765.

4. Explain the Anscombe's quartet in detail.

- Anscombe's quartet explains with 4 data set, which have similar summary statistics such as mean, variance, correlation Coefficient etc, that we should not consider them similar unless we plot a graph of these data and analyze them visually.
- Graphing the data on a plot like scatter plot shows the behavior of the data points and potential outliers which are important to get the complete picture of data set.

5. What is Pearson's R?

- Measures the strength of linear relationship between two variables, such that it is correct to draw a line representing all the data points.
- Its range is between – 1 and +1, +1 means that when one variable increases other variable also increases. -1 signifies opposite of this.
- R of 0 indicates that two variables are not related and a linear line representing them cannot be drawn.
- Below is the equation to calculate the Pearson's R for respective variable set x and y.

$$\frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{\sqrt{\left(\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right)}\sqrt{\left(\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}\right)}}$$

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

Scaling: It is way of data normalization which squeezes data within some range which helps to speedup calculations. Since machine learning algorithm performs on Euclidean distance for computation of data points. Since data set may contains values of different magnitude, range and

units that varies highly. As a result, high magnitude features would weight more than the low ones.

Scaling helps to ease out the variables interpretation and avoid any misleading inferences due to this.

To avoid this situation, we need to normalize data to standard same level for fair computation.

Main difference between normalized scaling and standardized scaling:

a. **Normalized Scaling also referred as min-max scaling,** is rescaling of features in particular range of [0,1]. It is formulated as follows:

$$\text{Min-max} = x - \text{Min}(x) / (\text{Max}(x) - \text{Min}(x))$$

b. **Standardization Scaling** is referred as rescaling of features with zero mean and 1 standard deviation which means unit variance. It is formulated as:

$$\text{Standardization} = x - \text{mean}(x) / (\text{S.D}(x))$$

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

VIF refers to Variance Inflation Factors which defines the correlation extent in a model between the predictors. It helps to recognize collinearity or multicollinearity in a model. It is evaluated as,

$$VIF = 1/ (1 - R^2)$$

Where, R^2 is squared R which defines coefficient of determination.

If Value of R is 1 than VIF can be infinite, this shows that variable is highly correlated with other variables and can be removed from the model (as its effect can be captured by other variables combination).

8. What is the Gauss-Markov theorem?

Gauss-Markov Theorem, name was derived from **Carl Friedrich Gauss** and **Andrey Markov,** and states that a model exists, if in a linear regression model,

- Errors are **uncorrelated**,
- Errors has equal **variance** and zero **expected value**,
- Coefficient of best linear unbiased estimator (BLUE) is the result of estimator of Ordinary Least Square.

The '**best'** means model with lowest variance. Assumptions of Gauss-Markov theorem are perfectly met rarely but still depicts what would be **ideal conditions** like. **Assumptions** of Gauss-Markov theorem are as follows:

- Error term's **expected value** must be 0 (Zero).
- **Collinearity,** independent variables are highly correlated.

- Regression equation which says that independent variable not dependent on dependent variables, refer to **Exogeneity**
- **Homoscedasticity** means having same variance (values of all independent variable are same).

9. Explain the gradient descent algorithm in detail.

- Gradient descent or steepest descent is a first order optimization algorithm proposed in 1847 by Cauchy. It is first order optimization because first derivative is considered where gradient gives steepest ascent's direction.
- Well, it is an iterative approach to minimize the function by proceeding in the negative direction or steepest descent. Both magnitude and direction are counted as its characteristics.
- Particularly in linear regression gradient descent can be used to minimize the cost function (Residual sum of squares) such that we get optimal coefficients.
- It follows the algorithm to compute partial derivates with respect to each coefficient and check values where the slope is very small (ideally zero).

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

Q-Q plot is referred as **Quantile- Quantile plot** which gives **graphical representation** of two sets of probability distribution, which plots quantiles against each other and check the distribution Example: whether normal or exponential.

- If sets of quantiles possesses same distribution then roughly straight line is plotted.
- It takes sample of data, and sorts them ascendingly which is then plotted against quantiles.
- It is used in comparing shapes of distributions, and data collection.
- It is more powerful than histogram used for two data samples and most widely used. No need to take pair of values as it makes comparison of two different quantiles.
- Thus, it is not necessary to bin values for Q-Q plot to work.
- Q-Q plot thus can help to check if data set is from a specified distributed based upon whether most of the points lie on a line or not.