1. Explain the linear regression algorithm in detail.

Answer: Linear Regression is one of the machine learning algorithms where the result is predicted by the use of known parameters which are correlated with the output. It is used to predict values within a continuous range rather than trying to classify them into categories. The known parameters are used to make a continuous and constant slope which is used to predict the unknown or the result.

Let's say we have a dataset which contains information about the relationship between 'number of hours studied' and 'marks obtained'. A number of students have been observed and their hours of study along with their grades are recorded. This will be our training data. Our goal is to design a model that can predict the marks if number of hours studied is provided. Using the training data, a regression line is obtained which will give minimum error. This linear equation is then used to apply for a new data. That is, if we give the number of hours studied by a student as an input, our model should be able to predict their mark with minimum error.

Regression analysis will provide you with a relation which can be visualized into a graph in order to make predictions about your data.

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

Answer:

-Linear Relationship between the features and target

- Little or no Multicollinearity between the features

- Homoscedasticity Assumption

- Normal distribution of error terms (The fourth assumption is that the error(residuals) follow a normal distribution)

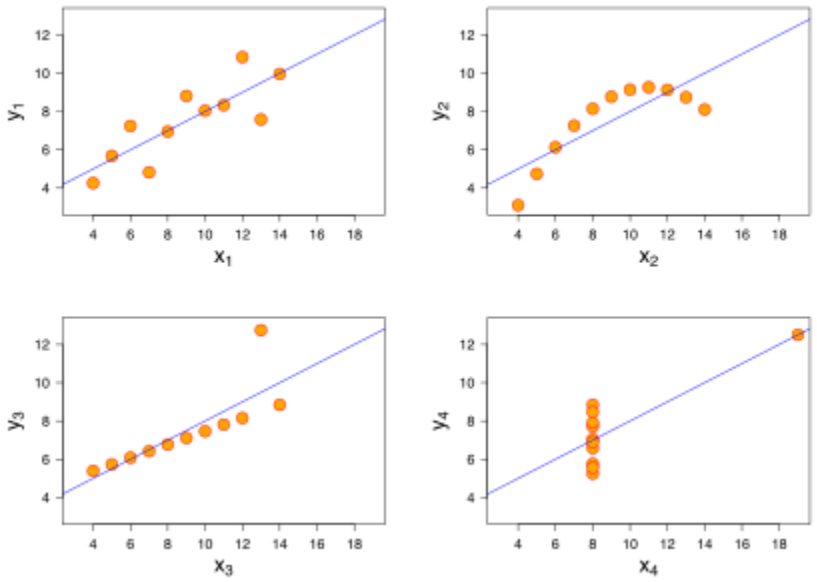
- Little or No autocorrelation in the residuals

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

Answer: 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.

4. Explain the Anscombe's quartet in detail.

Answer: Anscombe's Quartet consists of four data sets, that when examined have nearly the identical statistical properties, yet when graphed the datasets tell a very different story. Each of the datasets in the quartet consists of 11 (x, y) points.



- The first scatter plot (top left) appears to be a simple linear relationship, corresponding to two variables correlated where y could be modelled as gaussian with mean linearly dependent on x.
- The second graph (top right) is 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?

Answer: **Pearson's correlation coefficient** is the test statistics that measures the statistical relationship between two continuous variables. It is based on the method of covariance. It gives information about the magnitude of the correlation, as well as the direction of the relationship.

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

- Answer: Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or values or units. If feature scaling is not done, then a machine learning algorithm tends to weigh greater values, higher and consider smaller values as the lower values, regardless of the unit of the values.

- **Min-Max Normalization:** This technique re-scales a feature or observation value with distribution value between 0 and 1.
- **Standardization:** It is a very effective technique which re-scales a feature value so that it has distribution with 0 mean value and variance equals to 1.

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

Answer: 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). If there is perfect correlation, then VIF = infinity.

8. What is the Gauss-Markov theorem?

Answer: The Gauss-Markov theorem states that if your 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.

9. Explain the gradient descent algorithm in detail.

Answer: Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model. Parameters refer to coefficients in Linear Regression and weights in neural networks.

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

Answer: The Q-Q plot, or quantile-quantile plot, is a graphical tool to help us assess if a set of data plausibly came from some theoretical distribution such as a Normal or exponential. A Q-Q plot is a scatterplot created by plotting two sets of quantiles against one another. If both sets of quantiles came from the same distribution, we should see the points forming a line that's roughly straight. Q-Q plots take your sample data, sort it in ascending order, and then plot them versus quantiles calculated from a theoretical distribution. The number of quantiles is selected to match the size of your sample data. While Normal Q-Q Plots are the ones most often used in practice due to so many statistical methods assuming normality, Q-Q Plots can actually be created for any distribution.