1. Explain the linear regression algorithm in detail?

Linear regression is a statistical method of finding the relationship between independent and dependent variables

In order to fit the best intercept line between the points, we can use (SSE) Sum of Squared Errors or Ordinary Least Squares (OLS).

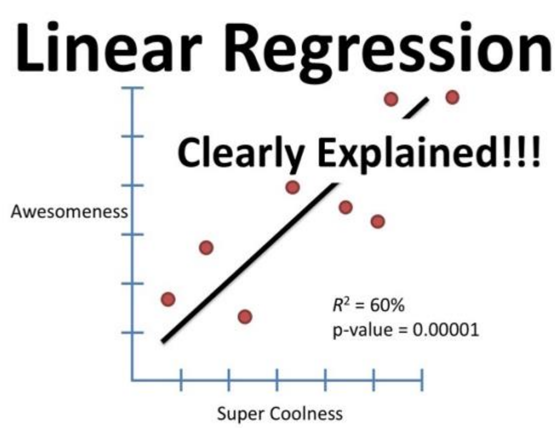
- It is important to understand **if** a linear regression modelling will be **applicable** to the problem you are trying to solve. For example, where linear regression cannot help:

  You decide if a customer will opt out of a subscription, as this is a classification problem.

- Linear regression guarantees **interpolation but not extrapolation**.
- While linear regression can be used for both projection and prediction,
- The business goal is crucial and will decide what path the modelling process should take.
- In the industry, variables that are **more actionable** are valued over others.

  For example,"Views to the platform" and "**Visitors to the platform**", the latter is more actionable, as it is easier to get viewers to the platform than forcing anybody to watch the shows.

- You don't end the modelling process until you are sure that no more significant variables could be added to explain the outcome.



-

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

Assumptions for linear regression:

- *The regression model is linear in parameters*
- Linear regression assumes that there is little or no multicollinearity in the data.
- Linear regression analysis requires all variables to be multivariate normal.

Especially for Residuals:

- *The mean of residuals is zero*
- *Homoscedasticity of residuals or equal variance.*
- *No autocorrelation between residuals.*
- *The X variables and residuals are uncorrelated*

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

Coefficient of correlation *is:* ***"R" value****, which is given in the summary table in the regression output.*
It depicts relation between predictor & target variables.
It's value range lies between: -1 to 1.

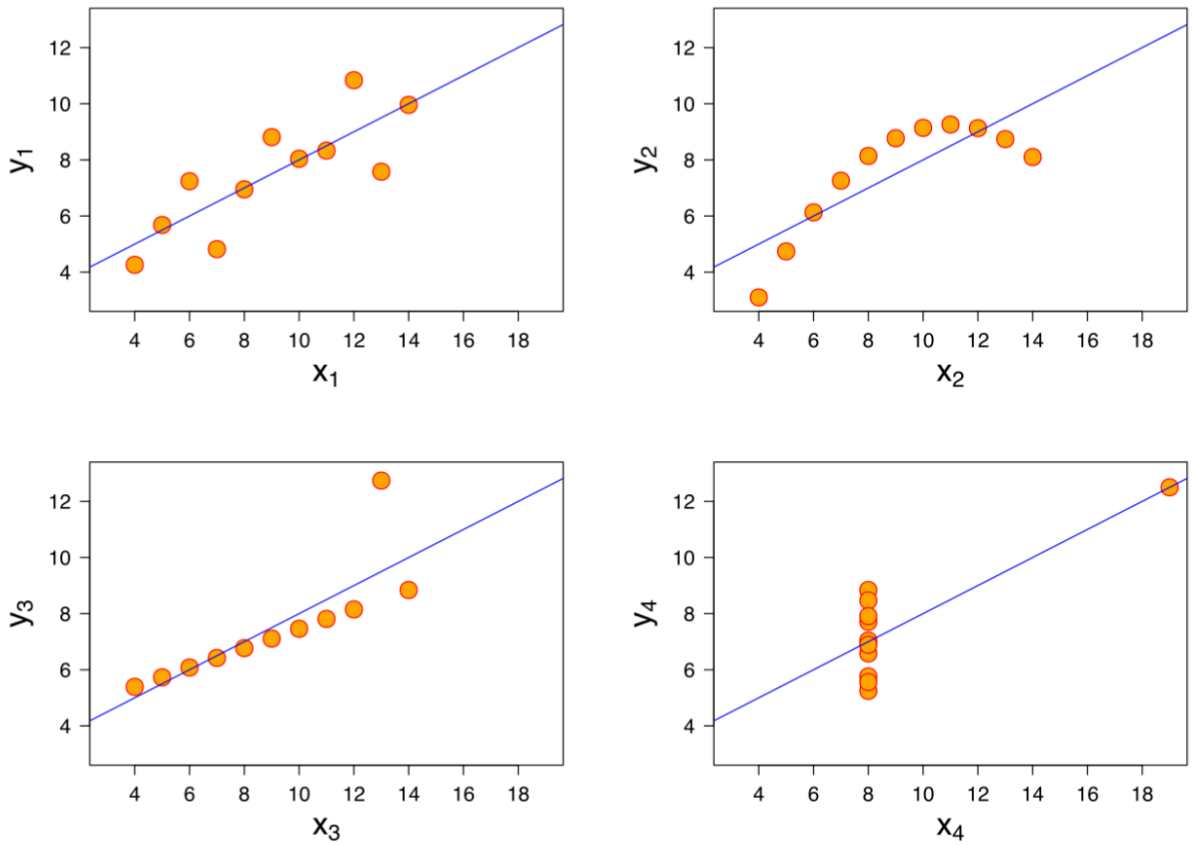
R square *is also called* ***coefficient of determination****.*

Coefficient of Determination is the square of Coefficient of Correlation.
It depicts percentage variation in y is explained by predictor variables.
It should be positive & value range between: 0 to 1.

4. Explain the Anscombe’s quartet in detail.

- *Assume four datasets, each containing 11 data points.*
- *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.*

When we plot these four datasets on an x/y coordinate plane, we can observe that they show the same regression lines as well but each dataset is telling a different story :



5. What is Pearson's R?

Pearson correlation coefficient, *r*:

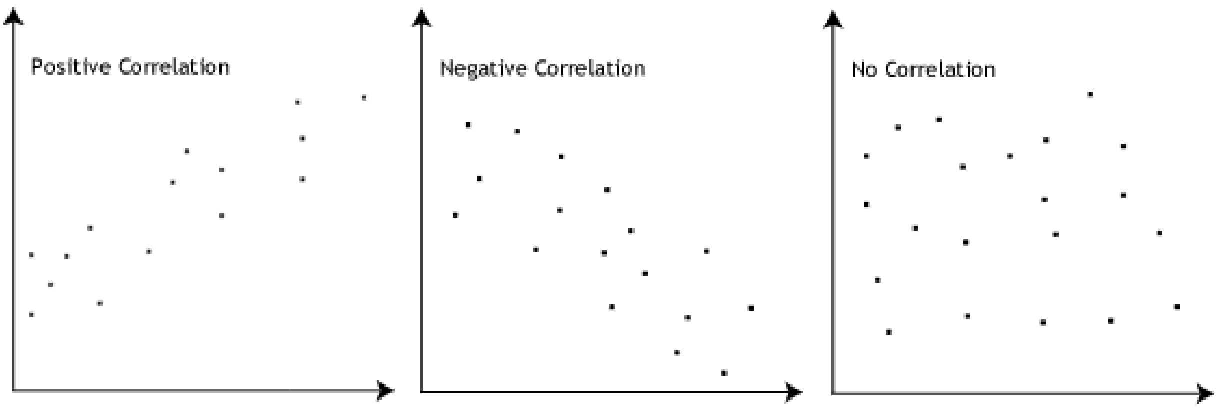
Range of values from +1 to -1.

A value of 0 indicates that there is no association between the two variables.

A value > 0 indicates a positive association,
one variable increases, so does the value of the other variable.

A value < 0 indicates a negative association,
one variable increases, the value of the other variable decreases.

This is shown in the diagram below:



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

- Scaling is a technique to standardize the predictor variable present in the data.
- Scaling performed to handle highly varying magnitudes or values or unit.

- **Min-Max Normalization:** It is a technique which re-scales a feature value into Range of 0 to 1.

  Formula:

$$X_{new} = \frac{X_i - \min(X)}{\max(x) - \min(X)}$$

- **Standardization:** Better technique which re-scales a feature value. so that it has distribution with “0” mean value and variance as 1.

  Formula:

$$X_{new} = \frac{X_i - X_{mean}}{\text{Standard Deviation}}$$

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

Variance Inflation Factor *(VIF) :*

It quantifies the extent of correlation between one predictor and the other predictors in a model.

Formula :

$$VIF = \frac{1}{1 - R^2}$$

From the above equation we can conclude that whenever the R-square value is equal to 1, VIF will be INF. So the extent to which a predictor is correlated with the other predictor variables in a linear regression is important here.

8. What is the Gauss-Markov theorem?

Gauss-Markov Theorem:

Given OLS Assumptions I-V, the OLS estimator is BLUE it means Best Linear Unbaised Estimator.

Linearity: the parameters we are estimating using the OLS method must be themselves linear.

Random: our data must have been randomly sampled from the population.

Non-Collinearity: the regressors being calculated aren't perfectly correlated with each other.

Exogeneity: the regressors aren't correlated with the error term.

Homoscedacity: no matter what values of our regressors might be, the error of the variance is constant.

The **Gauss Markov assumptions** guarantee the validity of ordinary least squares for estimating regression coefficients.

9. Explain the gradient descent algorithm in detail.

- It is an iterative optimization algorithm to find the minimum of a function. Here that function is our Loss Function.
- It is one of the best optimisation algorithms to minimise errors.
- To minimize the cost function the model needs to have the best value of θ_0 and θ_1

We can measure the accuracy of our hypothesis function by using a cost function and the formula is

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$

Gradient Descent for Linear Regression

$$\frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Various variants of gradient descent are defined on the basis of how we use the data to calculate derivative of cost function in gradient descent.

1. Batch Gradient Descent
2. Stochastic Gradient Descent
3. Mini-Batch Gradient Descent

Depending upon the amount of data used, the time complexity and accuracy of the algorithms differs with each other.

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

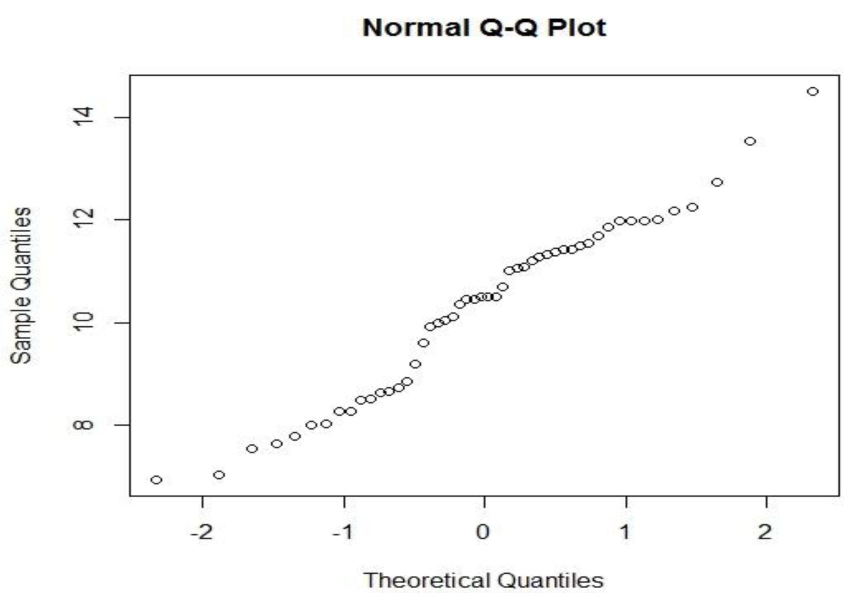
A Q-Q plot also called as quantile-quantile plot help us to know,

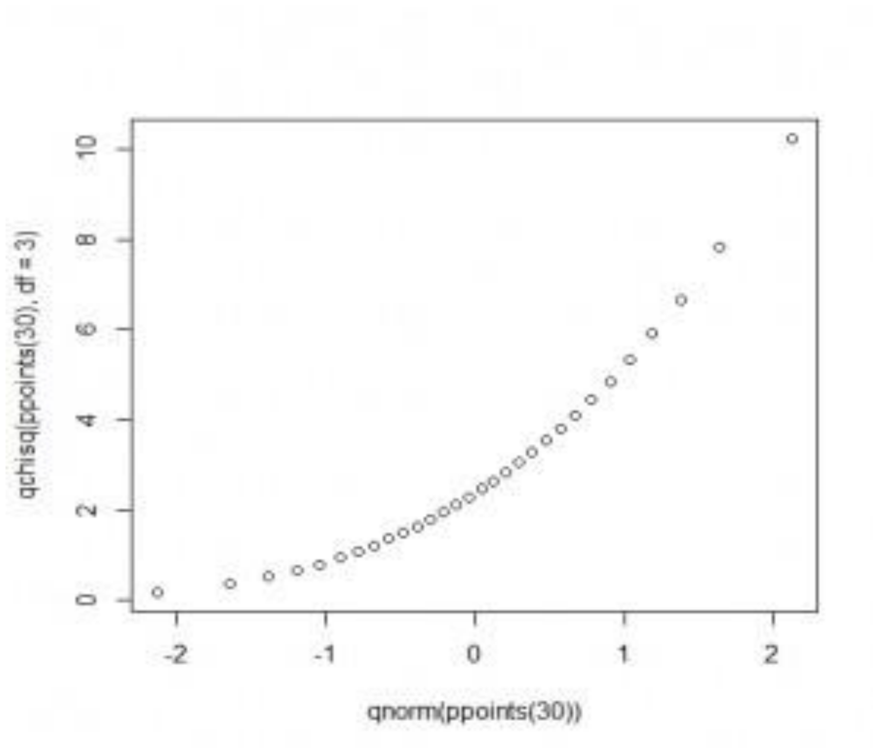
Most people use them in a single, simple way:

fit a linear regression model, check if the points lie on the line or not?

your residuals aren't Normal distribution which means, if a set of data set came from some theoretical distribution such as a Normal or exponential we can easily identify them.

Assume, If points forming a line is straight then it is from Normal distribution.
Here's an example of a Normal Q-Q plot when both sets of quantiles truly come from Normal distributions.





If points form a curve instead of a straight line. Q-Q plots that look like this usually mean your sample data are skewed.