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

ANS 1. **Linear Regression** is a machine learning algorithm based on **supervised learning i.e. past data with labels is used for building the model**.

b. Linear Regression models a target prediction value based on **independent variables**.

c. It is mostly used for finding out the relationship between variables and forecasting i.e. relates the dependent variables and the independent variables.

The equation for the linear regression is as follows:

Y = m1x1+m2x2+…. +c

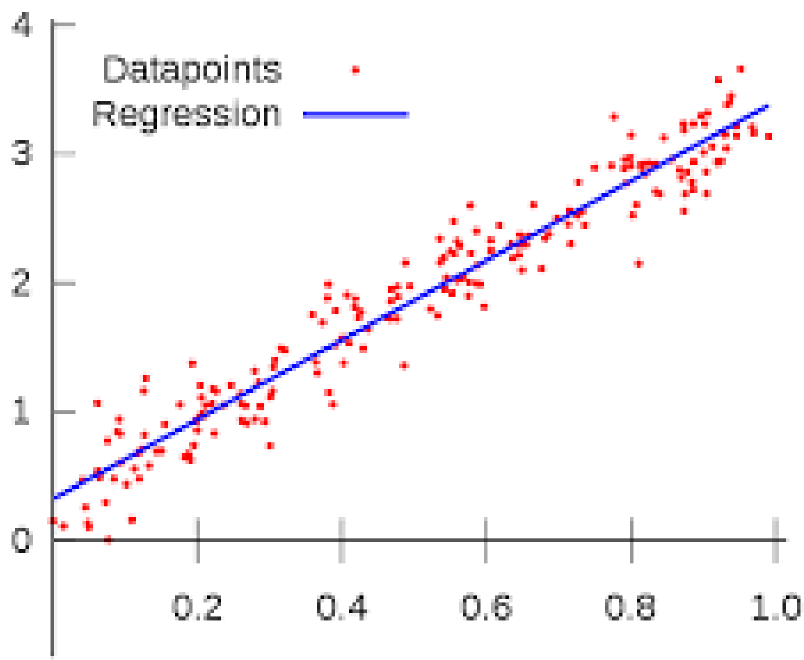


Fig:

In the above figure the blue points are indicating the actual points and the red line is the linear regression line

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

ANS 2: Following are the assumptions regarding linear regression:

1. **Error terms have *constant variance* (homoscedasticity):** The variance should not increase (or decrease) as the error values change. Also, the variance should not follow any pattern as the error terms change.

2. **Error terms are *normally distributed* with mean zero:** The assumption of normality is made, as it has been observed that the error terms generally follow a **normal distribution with mean equal to zero** in most cases.

3. **There is a *linear relationship* between X and Y:** X and Y should display some sort of a linear relationship; otherwise, there is no use of fitting a linear model between them.

4. **Error terms are *independent* of each other:** The error terms should not be dependent on one another**.**

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

ANS 3. Correlation coefficient is a a statistical relationship between two variables and is a numerical measure of some type of correlation meaning. The variables may be two columns of a given data set of observations, often called a sample, or two components of a multivariate variable with a known distribution.

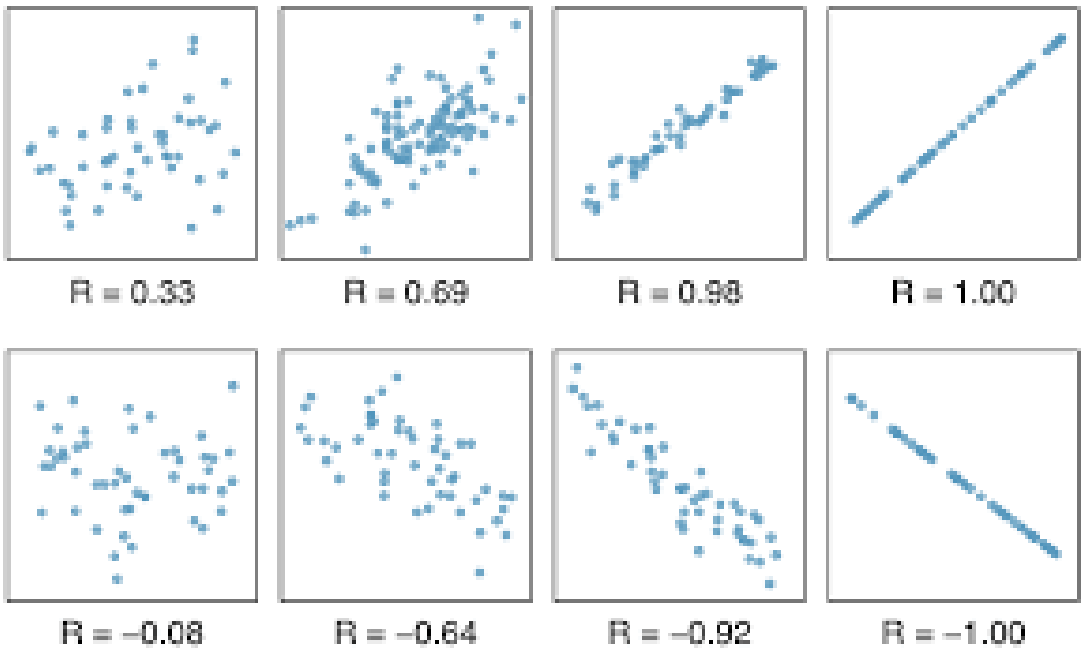


Fig: Here R is the a measure of correlation

Coefficient of determination is also denoted by R square : is the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

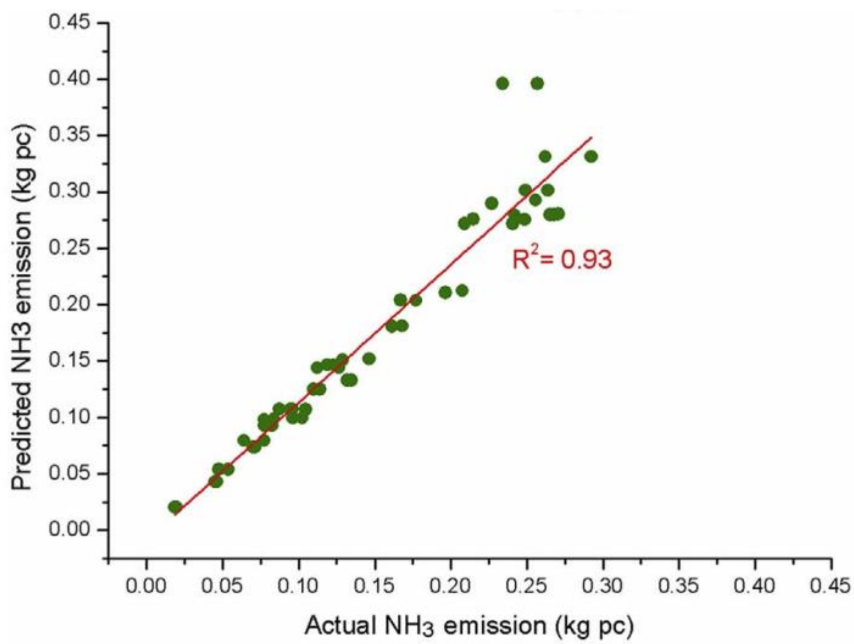
formula:

$$\boldsymbol{R^2 = MSS/TSS = (TSS - RSS)/TSS,}$$

where *MSS* is the model sum of squares, which is the sum of the squares of the prediction from the linear regression minus the mean for that variable;

TSS is the total sum of squares associated with the outcome variable, which is the sum of the squares of the measurements minus their mean; and

RSS is the residual sum of squares, which is the sum of the squares of the measurements minus the prediction from the linear regression.

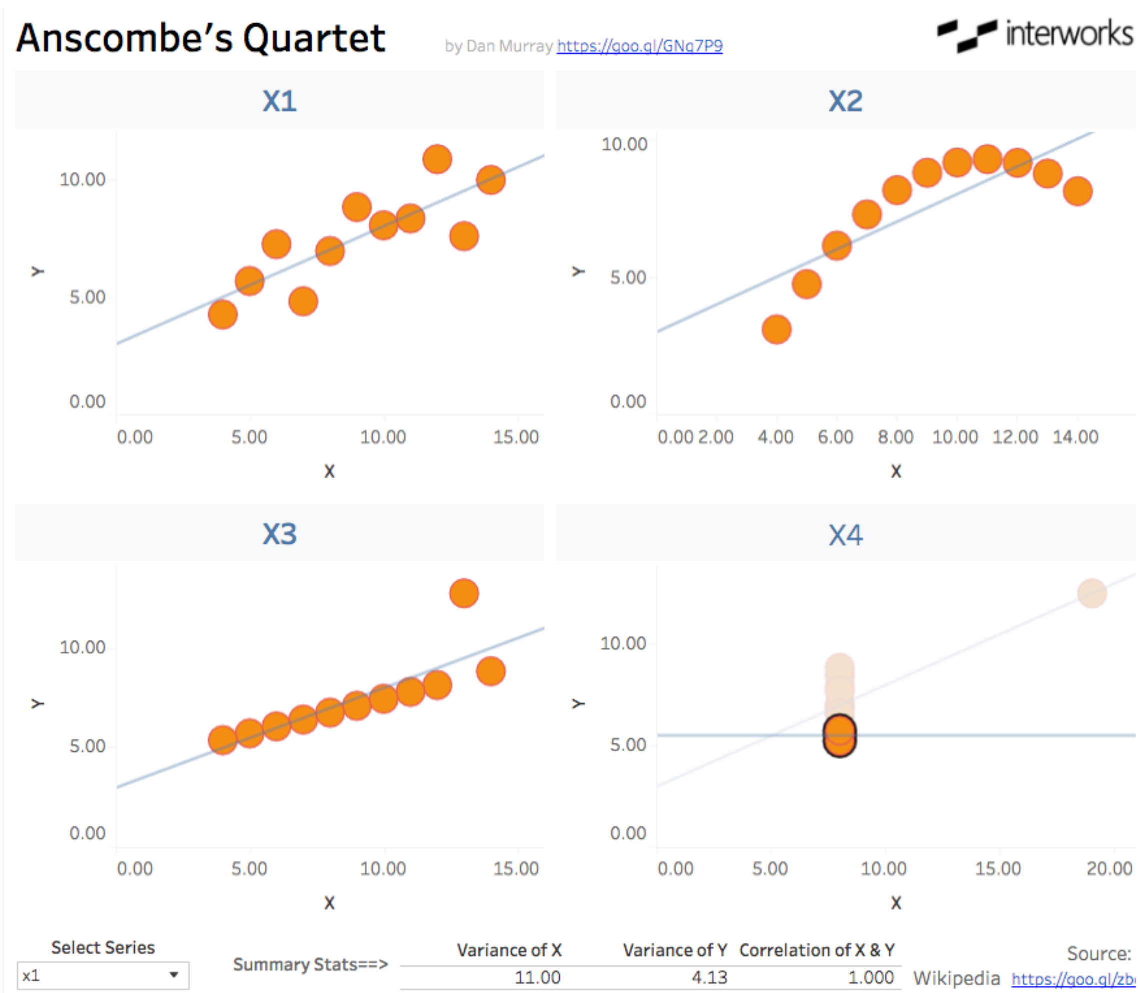


Predicted NH3 emission (kg pc)
Actual NH3 emission (kg pc)
R^2= 0.93
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45

QUES 4. Explain the Anscombe’s quartet in detail.

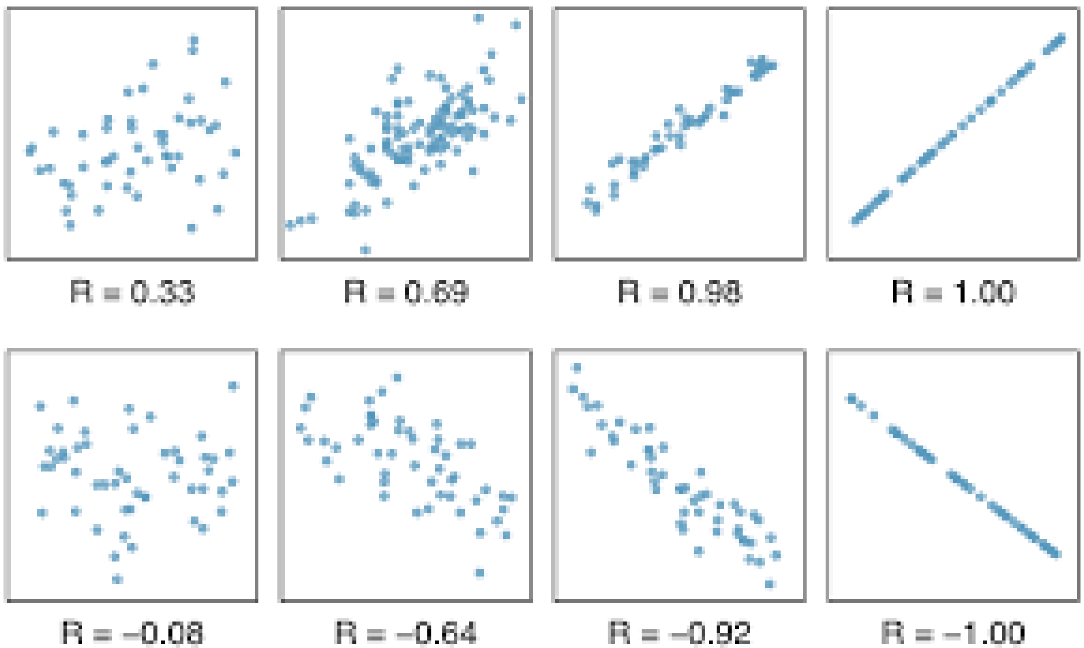
ANS 4 : Anscombe's quartet comprises four data sets that have nearly identical simple descriptive statistics, yet have very different distributions and appear very different when graphed.

Each dataset consists of eleven (*x*, *y*) points. They were constructed to demonstrate both the importance of graphing data before analyzing it and the effect of outliers and other influential statistics on statistical properties. The impression among statisticians that "numerical calculations are exact, but graphs are rough.



QUES 5. What is Pearson’s R?

ANS 5. Pearson’s R is a measure of the linear correlation between two variables *X* and *Y*. It has a value between +1 and −1, where 1 is total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation.



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

ANS 6: Scaling is a step of Data Pre-Processing which is applied to independent variables or features of data. It is performed because it helps to normalize the data within a particular range. Sometimes, it also helps in speeding up the calculations in an algorithm.

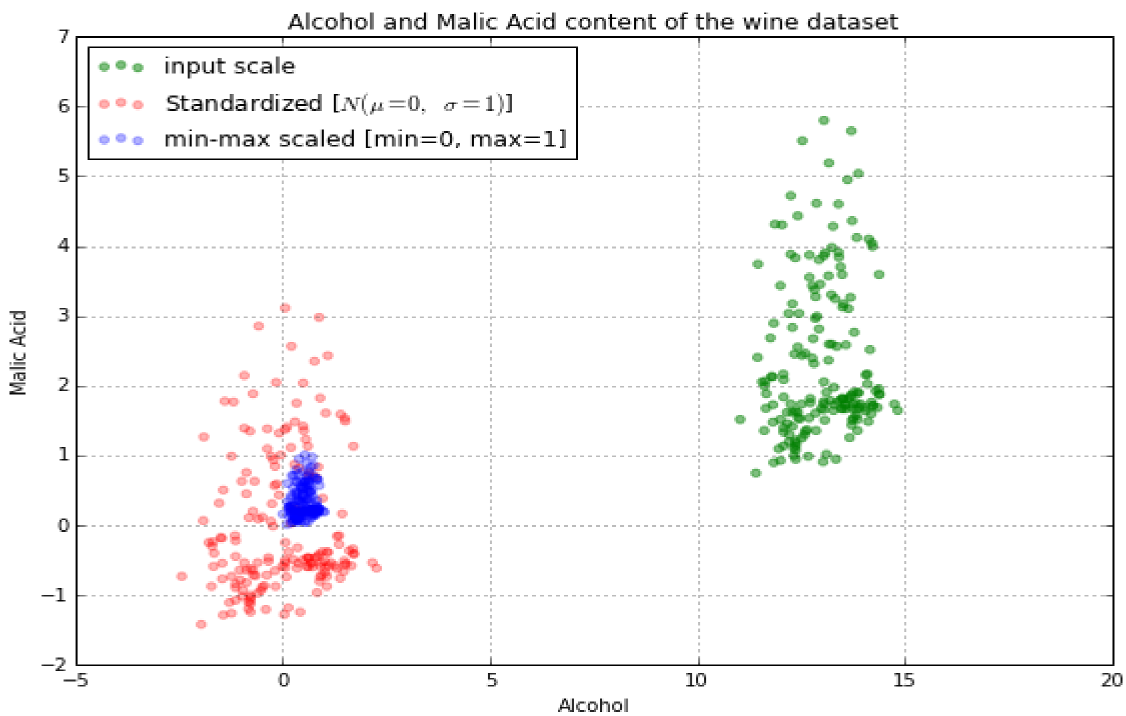
Normalization is of two types Standardization and MIN MAX SCALING

MIN MAX scaling usually means to scale a variable to have a value between 0 and 1,

i.e**. xn = x-xmin/xmax-xmin** while standardization transforms data to have a mean of zero and a standard deviation of 1.

This standardization is called a **z-score**, and data points can be standardized with the following formula:

X = xi – xbar / s



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

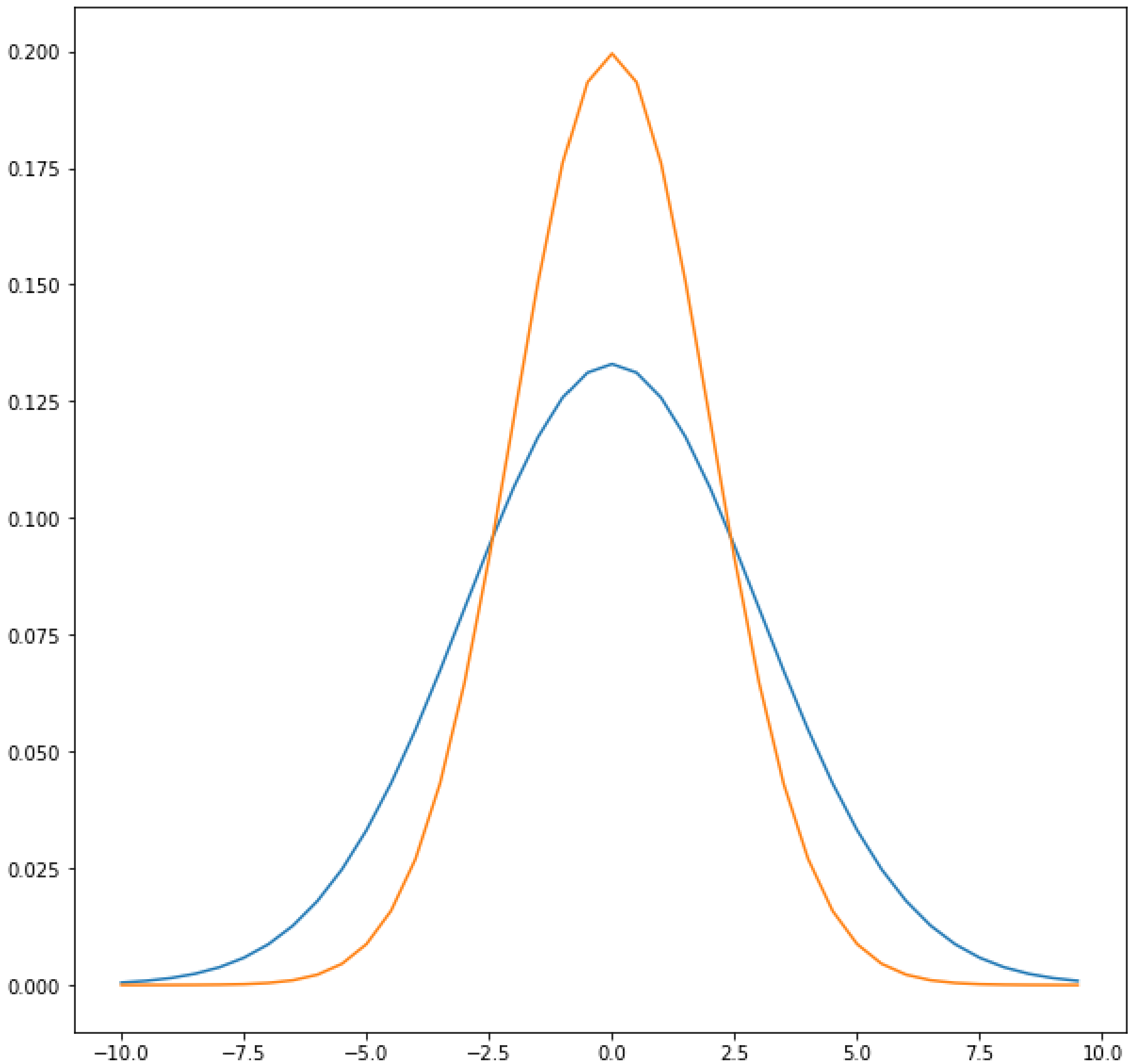
ANS 7. If there is perfect correlation, then **VIF** = **infinity**. A large value of **VIF** indicates that there is a correlation between the variables. If the **VIF** is 4, this means that the variance of the model coefficient is inflated by a factor of 4 due to the presence of multicollinearity.

$$VIF_i = \frac{1}{1 - R_i^2}$$

Formula: Clearly from the formula above when R2 reaches 1 i.e. 100% variance in the data is being explained the variable thus we get VIF as INFINITY.

QUES 8. What is the Gauss-Markov theorem?

Ans 8. The Gauss–Markov theorem states that in a linear regression model in which the errors are uncorrelated have equal variances and expectation value of zero, the best linear unbiased estimator of the coefficients is given by the ordinary least square (OLS) estimator.



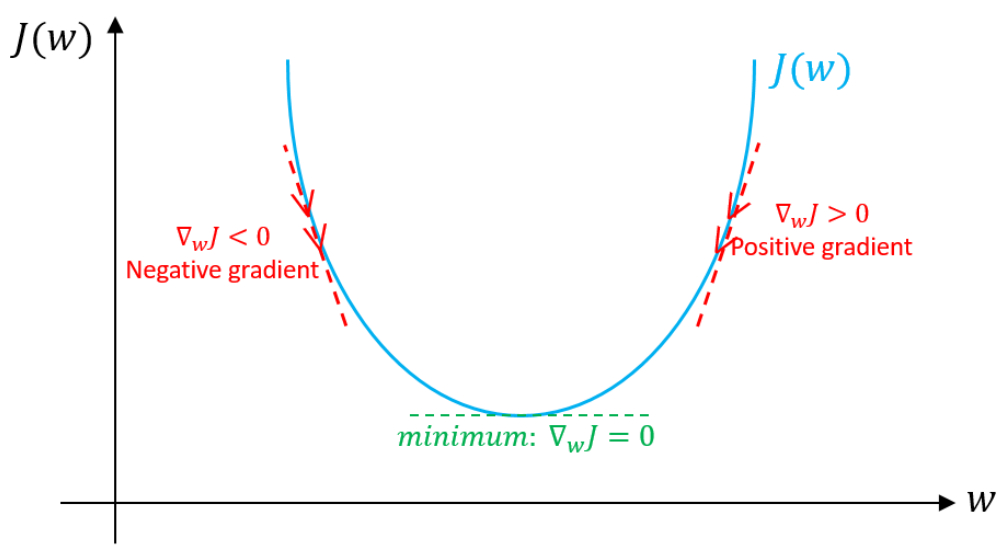
QUES 9. Explain the gradient descent algorithm in detail.

Ans 9. Gradient descent is a first order iterative optimization algorithm for finding the local minimum of a function.

Gradient descent is relatively slow close to the minimum: technically, its asymptotic rate of convergence is inferior to many other methods i.e. For poorly conditioned convex problems, gradient descent increasingly 'zigzags' as the gradients point nearly orthogonally to the shortest direction to a minimum 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**. Gradient descent is also known as **steepest descent**.

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.



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

ANS 10. The quantile-quantile (q-q) plot is a graphical technique for determining if two data sets come from populations with a common distribution. A q-q plot is a plot of the quantiles of the first data set against the quantiles of the second data set. By a quantile, we mean the fraction (or percent) of points below the given value. That is, the 0.3 (or 30%) quantile is the point at which 30% percent of the data fall below and 70% fall above that value.

The advantages of the q-q plot are:

a. The sample sizes do not need to be equal.

b. Many distributional aspects can be simultaneously tested. For example, shifts in location, shifts in scale, changes in symmetry, and the presence of outliers can all be detected from this plot. For example, if the two data sets come from populations whose distributions differ only by a shift in location, the points should lie along a straight line that is displaced either up or down from the 45-degree reference line.