1. Explain the linear regression algorithm in detail.

--→ Linear Regression algorithm is used to find a relationship between dependent and in dependent variables which have a finite number of parameters which define relationship between variables.

Linear Regression Algorithm can be explained as below:
General expression of linear regression is

$$Y = B0 + B1 * X1 + B2 * X2 + \ldots\ldots\ldots\ldots + Bn * Xn$$

Here Y – Dependent Variable and X – dependent variable
B – parameter's to be predicted by algorithm.

We are trying to find parameters B such that there is a minimum error in prediction.

When we have more than one input we can use Ordinary Least Squares to estimate the values of the coefficients.
The Ordinary Least Squares procedure seeks to minimize the sum of the squared residuals. This means that given a regression line through the data we calculate the distance from each data point to the regression line, square it, and sum all of the squared errors together. This is the quantity that ordinary least squares seeks to minimize.

This operation is called Gradient Descent and works by starting with random values for each coefficient. The sum of the squared errors are calculated for each pair of input and output values. A learning rate is used as a scale factor and the coefficients are updated in the direction towards minimizing the error. The process is repeated until a minimum sum squared error is achieved or no further improvement is possible.

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

-→ The assumptions regarding Linear Regression residuals are:

1) Residual error should follow a normal distribution with their mean at 0.
2) Error terms are independent of each other
3) Error terms have constant variance (homoscedasticity).

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

Answer:

Coefficient of Correlation - **Correlation coefficients** are used in statistics to measure how strong a relationship is between two variables. It gives us idea about how much they are related and is it a positive correlation or a negative one.

Coefficient of determination - the coefficient of determination, denoted R^2 or r^2 and pronounced "R squared", is the proportion of the variance in the dependent variable that is predictable from the independent variable.

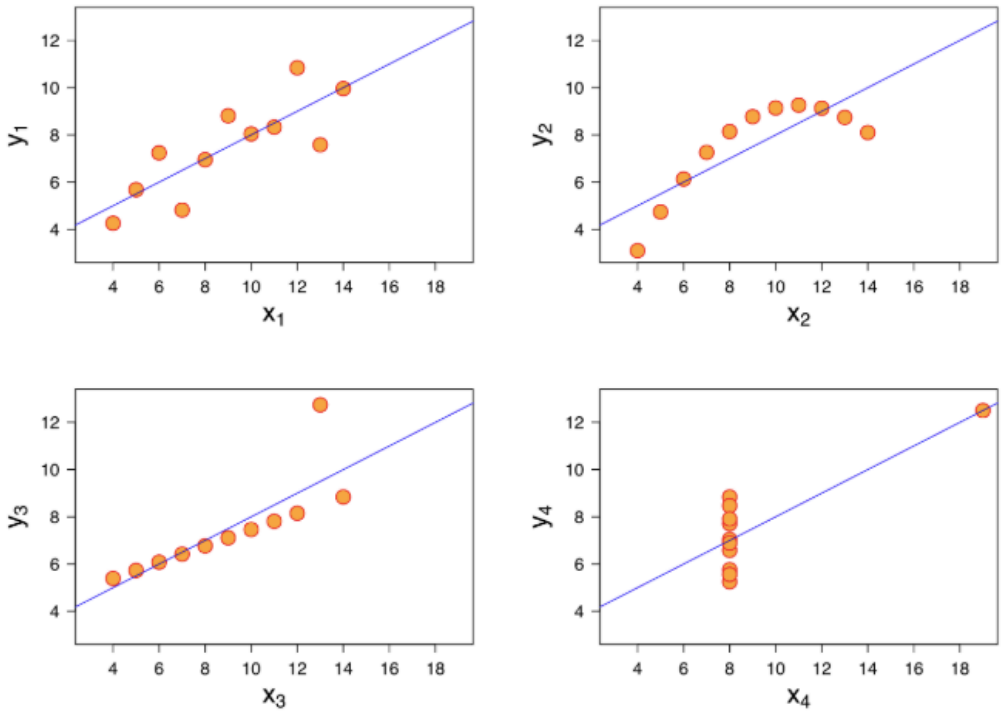
4. Explain the Anscombe's quartet in detail.

Answer:

Anscombe's quartet comprises of graph having similar statistics but a yet have very different distributions and appear very different when graphed.

It comprises four datasets, each containing eleven (x,y) pairs. The essential thing to note about these datasets is that they share the same descriptive statistics But things change completely ,when graphed.

Below shows example of Anscombe's quartet



We can observe details below for this quartet

- Dataset I appears to have clean and well-fitting linear models.
- Dataset II is not distributed normally.
- In Dataset III the distribution is linear, but the calculated regression is thrown off by an outlier.
- Dataset IV shows that one outlier is enough to produce a high correlation coefficient.

5. What is Pearson's R?

Answer:

Pearson's correlation coefficient is the test statistics that measures the statistical relationship, or association, between two continuous variables. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. It gives information about the magnitude of the association, or 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:

One of the most important transformations you need to apply to your data is *feature scaling*. With few exceptions, Machine Learning algorithms don't perform well when the input numerical attributes have very different scales .

There are two common ways to get all attributes to have the same scale: *min-max scaling* and *standardization*.

Min-max scaling (many people call this *normalization*) is quite simple: values are shifted and rescaled so that they end up ranging from 0 to 1. We do this by subtract- ing the min value and dividing by the max minus the min .

Standardization first it subtracts the mean value (so standardized values always have a zero mean), and then it divides by the variance so that the resulting distribution has unit variance.

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

Answer:

VIF (Variance Inflation Factor) - VIF is an index that provides a measure of how much the variance of an estimated regression coefficient increases due to collinearity.

It is given by 〖VIF〗=1/(1-R^2)

So when r is 1 i.e there is a very high correlation when independent VIF tends to go towards infinity.

8. What is the Gauss-Markov theorem?

Answer:

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 squares estimator.

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.
The size of these steps is called the *learning rate*.
High Learning rate may lead to overshooting local minima
Low Learning may lead to taking more time to reach local optima.

A Loss Functions tells us “how good” our model is at making predictions for a given set of parameters. The cost function has its own curve and its own gradients. The slope of this curve tells us how to update our parameters to make the model more accurate.

$$f(m, b) = \frac{1}{N} \sum_{i=1}^{n} (y_i - (mx_i + b))^2$$

Above is an example of Gradient descent. Where it is iterating from 1 to n example calculating difference between predicted and original observation to minimize the error between them.

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

Q–Q (quantile-quantile) plot is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. First, the set of intervals for the quantiles is chosen. A point (x, y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate). Thus the line is a parametric curve with the parameter which is the number of the interval for the quantile.

It helps in observing the distribution and fitting line through datapoints.

When there are two data samples, it is often desirable to know if the assumption of a common distribution is justified. If so, then location and scale estimators can pool both data sets to obtain estimates of the common location and scale. If two samples do differ, it is also useful to gain some understanding of the differences.