

1. Explain the linear regression algorithm in detail.

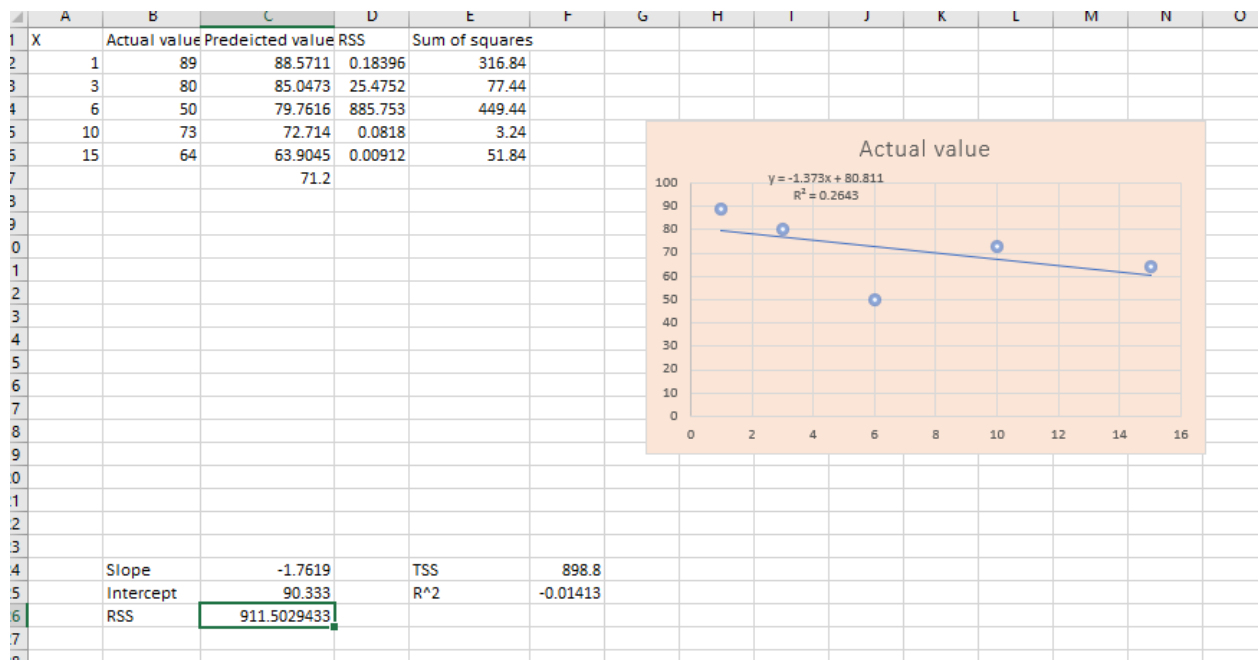
Linear regression is a supervised learning algorithm.

Linear regression is a predictive modeling technique which defines the relationship between dependent variable and single/multiple independent variables.

The algorithm is used to find the '**Best Fit Line**' of relation between any independent variable(s) and dependent variable.

The standard equation of the simple linear regression line is given by the following expression: $Y = \beta_0 + \beta_1 X$

Best fit line can be determined by minimizing the RSS (Residual Sum of Squares) value. Residuals for any data point is found by subtracting predicted value of dependent variable from actual value of dependent variable.



The idea of Simple Linear Regression is finding those parameters α and β for which the error term is minimized. One way to do this is by using OLS (Ordinary Least Squares method).

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

- Normality assumption: It is assumed that the error terms are normally distributed.
- Zero mean assumption: It is assumed that the residuals have a mean value of zero, i.e., the error terms are normally distributed around zero.

- Constant variance assumption: It is assumed that the residual terms have the same (but unknown) variance. This assumption is also known as the assumption of homogeneity or homoscedasticity.
- Independent error assumption: It is assumed that the residual terms are independent of each other, i.e. their pair-wise covariance is zero.

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. R square is also called coefficient of determination.

R square or coeff. of determination shows percentage variation in y which is explained by all the x variables together. Higher the better. It is always between 0 and 1. It can never be negative – since it is a squared value.

Coefficient of Correlation: is the degree of relationship between two variables say x and y. It can go between -1 and 1. 1 indicates that the two variables are moving in unison. They rise and fall together and have perfect correlation. -1 means that the two variables are in perfect opposites. One goes up and other goes down, in perfect negative way.

4. Explain the Anscombe’s quartet in detail.

Anscombe’s Quartet was developed by statistician Francis Anscombe. 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 differs completely when they are graphed.

5. What is Pearson’s R?

The Pearson product-moment correlation coefficient (PPMCC), or the bivariate correlation, [1] is a statistic that measures linear correlation between two variables X and Y. It has a value between +1 and -1

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

When you have a lot of independent variables in a model, a lot of them might be on very different scales which will lead a model with very weird coefficients that might be difficult to interpret. So we need to scale features because of two reasons:

1. Ease of interpretation

2. Faster convergence for gradient descent methods

You can scale the features using two very popular methods:

1. Standardizing: The variables are scaled in such a way that their mean is zero and standard deviation is one.

2. MinMax Scaling: The variables are scaled in such a way that all the values lie between zero and one using the maximum and the minimum values in the data. It is important to note that

scaling just affects the coefficients and none of the other parameters like t-statistic, F statistic, p-values, R-square, etc

7. You might have observed that sometimes the value of VIF is infinite. Why does this happen?
VIF is infinite if there is a perfect correlation between the variables and indicates that the corresponding variable may be expressed exactly by a linear combination of other variables
8. What is the Gauss-Markov theorem?
Gauss–Markov theorem states that the ordinary least squares (OLS) estimator has the lowest sampling variance within the class of linear unbiased estimators, if the errors in the linear regression model are uncorrelated, have equal variances and expectation value of zero.
9. Explain the gradient descent algorithm in detail.
Gradient descent is an optimization algorithm used to find the values of parameters (coefficients) of a function (f) that minimizes a cost function (cost).

The algorithm starts off with initial values for the coefficient or coefficients for the function. These could be 0.0 or a small random value.

$$\text{coefficient} = 0.0$$

The cost of the coefficients is evaluated by plugging them into the function and calculating the cost.

$$\text{cost} = f(\text{coefficient})$$

The derivative of the cost is calculated. We need to know the slope so that we know the direction (sign) to move the coefficient values in order to get a lower cost on the next iteration.

$$\text{delta} = \text{derivative}(\text{cost})$$

Now that we know from the derivative which direction is downhill, we can now update the coefficient values. A learning rate parameter (alpha) must be specified that controls how much the coefficients can change on each update.

$$\text{coefficient} = \text{coefficient} - (\text{alpha} * \text{delta})$$

This process is repeated until the cost of the coefficients (cost) is 0.0 or close enough to zero to be good enough.

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

Q-Q plot is a scatter plot used to determine the type of distribution. The plot can easily tell us if the distribution is normal or not.

After fitting a linear regression model, if the points lie approximately on the line, and if they do not, the residuals and errors are not normal distribution.