1.Explain the linear regression algorithm in detail.

Ans: Regression is a model (mathematical equation) which predicts the target variable using predictor variables. If the mathematical equation(model) is in the form of a line, then such regressions are called Linear regression.

Simple linear regression line

y is dependent variable or target variable and X1 is independent variables or predictor variables.

Main aim of the linear regression algorithm is to find the best values of so it best fits the data.

Best fit line is the line which best explains the data, and which has least residual sum of squares (RSS)

Error or residual is the difference between actual value and predicted value.

RSS==

To get the best fit line which helps in prediction of target variable using predictors variables. we use least squares method or gradient descent method to get line with low RSS.

Least Squares method:

y= mx + c m is the slope of line and c is the x intercept.

Using ordinary differentiation to condition for RSS to be minimum is

m = n is number of data points

c =

Minimizing Cost function:

This is used to decrease the cost function

= J ()

We can use closed form solution and Iterative form solution. If the number of independent variables is less than 2 then we can prefer closed form solution, which use normal differentiation. But if the number of independent variables increases then iterative form solution gives the solution.

Gradient descent method comes under first order of iterative solution. Newton method is for higher order.

Gradient descent is optimization method used to minimize the cost function by moving in the direction of negative gradient. Learning rate is the step rate at which algorithm is moving towards the negative gradient in each iteration.

| where .

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

Ans: To make inference of population using sample we need to make some assumption in linear regression.

Assumptions regarding residual are

* Residual should be normal distributed with mean centered at zero.
* Residual are independent of each other there should not be any correlation between the residual terms. No, patterns in the scatter plot
* Residual should have constant variance(homoscedasticity), variance should not change as time increases, or value of independent variables increase.

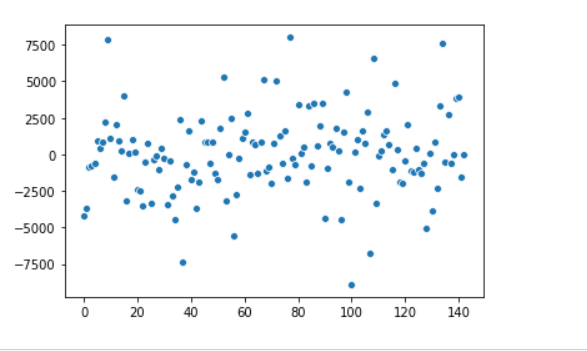
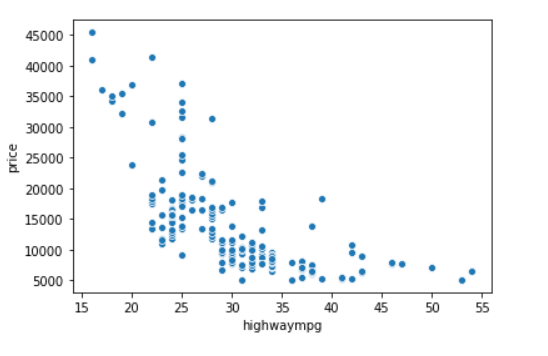
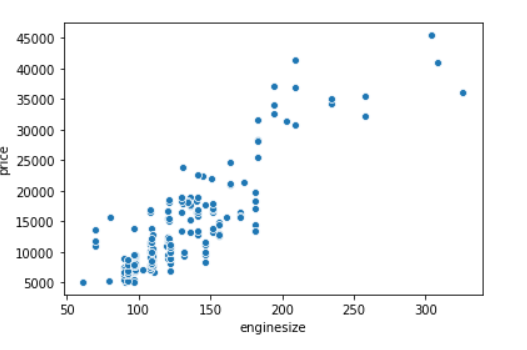
3. What is the coefficient of correlation and the coefficient of determination?

Ans: Coefficient of correlation: It is statistical measure which tells the relation between two continuous variables. It ranges between -1 to 1. It is usually denoted by r

r=1 signifies high positive correlation

r=0 signifies no relation

r=-1 signifies high negative correlation.



Positively correlated Negatively correlated No correlation

Coefficient of determination:

It is defined as square of correlation between predictor variables and target variable. It gives the measure of variance of dependent variable explained using the predictor variables. It is also defined as explained variance by total variance. It ranges from 0 to 1.

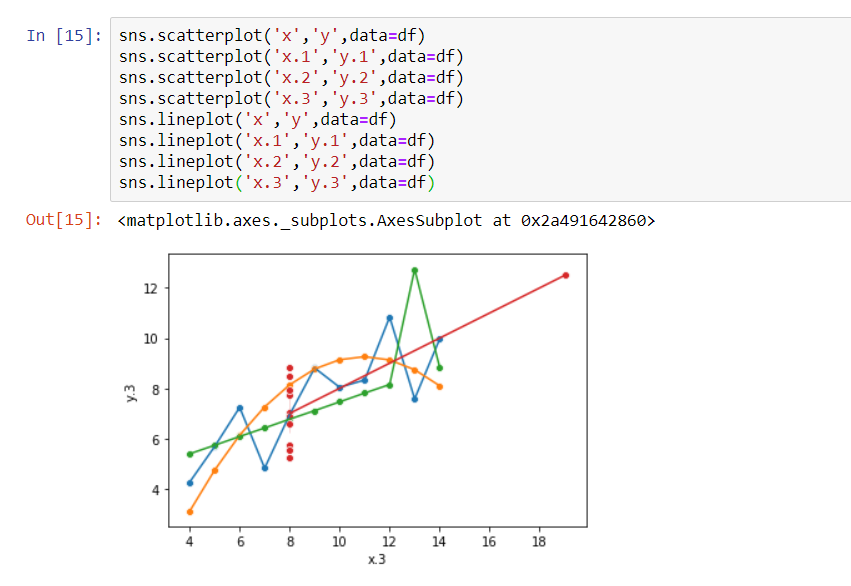
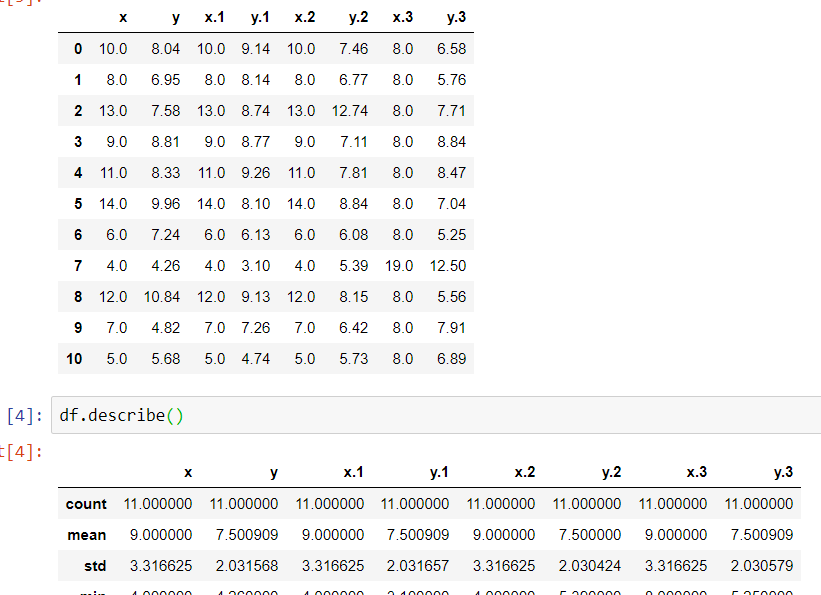
It is usually denoted by r2.

r2 =0 signifies dependent variable can’t be predicted using independent variables as there is not relation between them

r2 =1 variance of dependent variables is completed explained by independent variables as there is a perfect correlation between them.

4.Explain the Anscombe’s quartet in detail.

Ans: Anscombe quartet consists of four data sets which are statistically same but on plotting the graph they are way different. This was constructed by Francis Anscombe. This tells the importance of data visualization in model development.



In the above plots you can see how the visualization are different even though statistical mean and standard deviation of all are same as shown in data frame (df.describe()).

5.What is Pearson’s R?

Ans: Pearson’s R is one type of correlation coefficient used in linear regression, which gives the measure of strength and direction of linear relation between two variables. It is also called as PPMC (Pearson product moment correlation). It also ranges from -1 to 1 and denoted by R or rho ().

Its mathematical formula

R=

Where n is number of data points.

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

Ans: Scaling is a method to normalize/standardize the independent variable within a range.

Scaling is performed to standardize the highly varying magnitude, since few algorithms are sensitive to magnitudes and to help algorithm to converge, with a smaller number of iterations.

Normalized and standardized scaling are two types of scaling.

|  |  |
| --- | --- |
| Normalized scaling | Standardized scaling |
| Formula = | where |
| It rescales the data into a range [0,1] | It rescales data to have mean=0 and standard deviation of 1 |
| Outliers are lost in this approach | Not all outliers are lost |
| It is used when there are not outliers in the data or when outliers are not considered to build the model | It is used when model build is dependent on outliers |
| MinMaxScaler() object is used to scale | StandardScaler() object is used to scale |
| It is also called as minmax scaling | It is also called z score standardization |
| For binary categorical variables values are same before and after min max scaling | For binary categorical variables values vary after standardized scaling |

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

Ans: VIF= formula of VIF , where is coefficient of determination

VIF becomes infinity when 1- =0, i.e. =1

This happens when we have perfectly correlated variables in our independent variables. It might also happen when we pass target variables as one of the predictor variables to build the model.

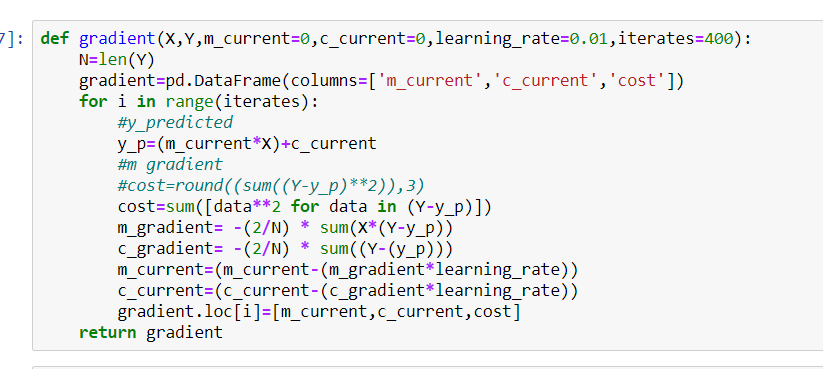
8.What is the Gauss-Markov theorem?

Ans: Gauss-Markov theorem states that, in linear regression model if error terms are noncorrelated, homoscedasticity and their mean is zero. Then Ordinary Least square (OLS) estimates gives the best linear unbiased estimate (BLUE). In other words if error terms satisfies the assumptions of linear regression, then ordinary least squares gives the best fit line for that linear model.

9. Explain the gradient descent algorithm in detail.

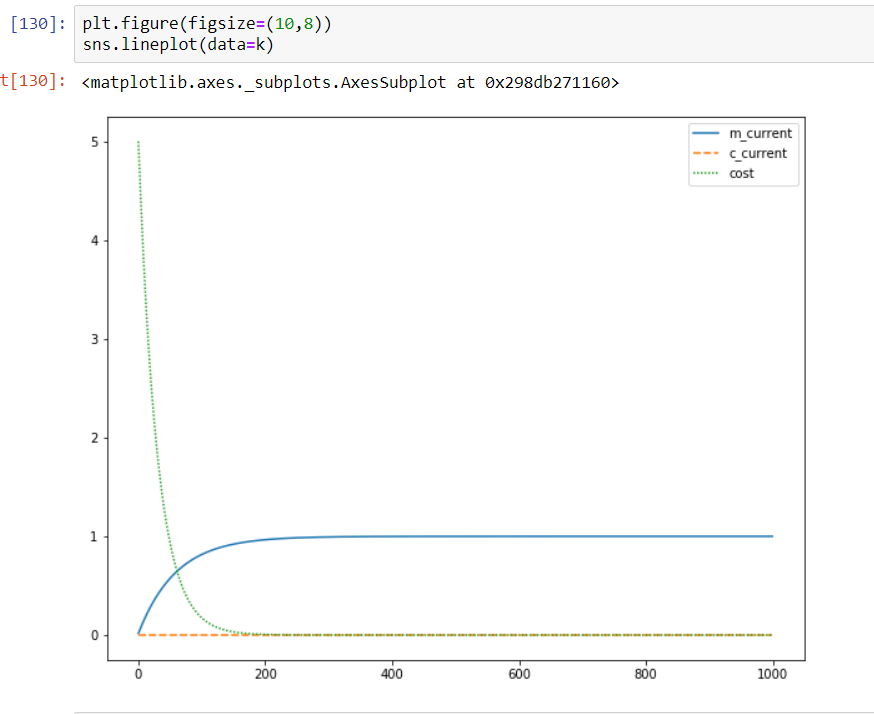
Ans: Gradient descent is first ordered iterative optimization algorithm used to minimize a function. It reaches the local minima of the given function by moving in the direction of negative gradient. Learning rate is the step rate at which algorithm is moving towards the negative gradient in each iteration. It is also known as steep descent method.

| where .



In each iteration new values of m and c are created using learning rate and

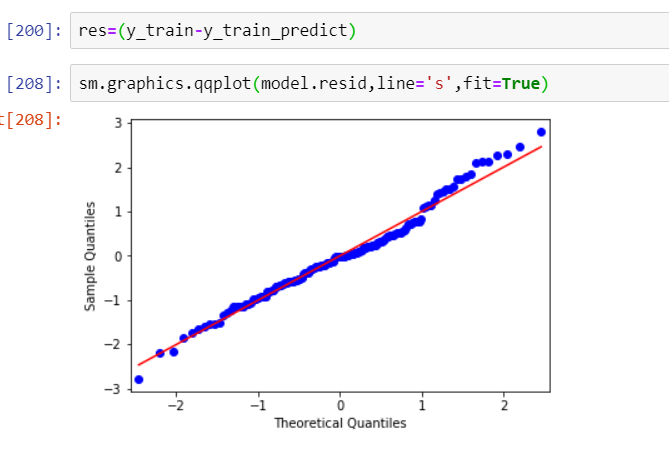
These newly created m and c values try to reduce the cost function and finally converges at optimized point.



Above graph those how cost of function as reduced and finally reached the optimum point.

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

Ans: Q-Q plot is a graphical tool used to compare two probability distributions by plotting their quantiles against each other.



If the given quantiles are from same distribution, then Q-Q plot resembles to a line y=x.

If distributions are linearly related, then Q-Q plot resembles a line y=mx+c

Q-Q plots are used in the linear regression to verify whether the residual or error terms are normally distributed or not. As it is one of the assumptions in the linear regression. Q-Q plots also tells the skewness of the residuals.