input variables called - predictors ,independent variables,features,variables

output variables- responses,dependent variables

Linear Regression is a statistical procedure that determines the

equation for the straight line that best fits a specific set of data.

The objective of Linear Regression is to find a line that minimizes the

prediction error of all the data points.

find the predicted values y but centroid using the mean value of x and y

that gives coeffecient values

find the predicted y

calculate RSS(mean absolute error)-residual sum of square that is (y-y(predict))

then find the intercept and slope using the forumla

intercept=(x-x(mean))\*(y-y(mean))/(x-x(mean))^2

slope=y=mx+c-> c=y(mean)-mx(mean)

Mean Squared Error represents the average of the squared difference between

the original and predicted values in the data set. It measures the variance

of the residuals. square of rss or mae

Root Mean Squared Error is the square root of Mean Squared error. It measures

the standard deviation of residuals.

square root of mse

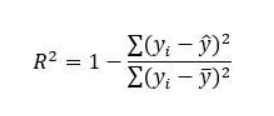
R2 measures the proportion of variance in the dependent variable that can be

explained by the independent variables in the model. In other words, it

indicates the percentage of the total variation in the dependent variable

that is accounted for by the independent variables.

fitness of the model(less than one to be good fit)



adjustent R2

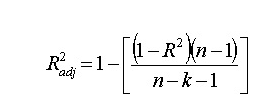
Adjusted R squared is a modified version of R square, and it is adjusted

for the number of independent variables in the model, and it will always

be less than or equal to R².In the formula below n is the number of

observations in the data and k is the number of the independent variables

in the data.



* R Squared & Adjusted R Squared are used for explaining how well the independent variables in the linear regression model explains the variability in the dependent variable. R Squared value always increases with the addition of the independent variables which might lead to the addition of the redundant variables in our model. However, the adjusted R-squared solves this problem.
* Adjusted R squared takes into account the number of predictor variables, and it is used to determine the number of independent variables in our model. The value of Adjusted R squared decreases if the increase in the R square by the additional variable isn’t significant enough.
* For comparing the accuracy among different linear regression models, RMSE is a better choice than R Squared.

Both RMSE and R- Squared quantifies how well a linear regression model fits a dataset. The RMSE tells how well a regression model can predict the value of a response variable in absolute terms while R- Squared tells how well the predictor variables can explain the variation in the response variable

Linear Regression assumptions

Linearity: The relationship between the dependent variable and the independent variables is linear. This means that a straight line can adequately represent the relationship between the variables.

Independence: The observations are independent of each other. This means that the value of one observation does not depend on the value of another observation.

Homoscedasticity: The variance of the errors is constant across all levels of the independent variables. This means that the spread of the errors should be the same for all values of the independent variables.

Normality: The errors are normally distributed. This means that the distribution of the errors should be symmetric and bell-shaped.

No multicollinearity: The independent variables are not highly correlated with each other. This means that there is no perfect linear relationship between the independent variables.