Advertising Sales Channel Prediction

# Problem Statement:

When a company enters a market, the distribution strategy and channel it uses are keys to its success in the market, as well as market know-how and customer knowledge and understanding. Because an effective distribution strategy under efficient supply-chain management opens doors for attaining competitive advantage and strong brand equity in the market, it is a component of the marketing mix that cannot be ignored .

The distribution strategy and the channel design have to be right the first time. The case study of Sales channel includes the detailed study of TV, radio and newspaper channel.

The goal is to predict the total sales generated from all the sales channel. This can be achieved by building a machine learning model.

# Machine learning life cycle:

1)Data Gathering- We need to get the data from business.

2)Data Cleaning-In this step we might need to remove duplicates, fill in missing data.

3)Exploratory Data Analysis-it involves 1)checking description about different columns like mean, median, standard deviation, data types etc.2)Univariate analysis,3)Bivariate analysis,4)Multivariater analysis

4)Feature Engineering and selection-It involves removing skewness,removing outliers,converting object data type numerical,using sampling techniques if there is any data imbalance,feature scaling

5)Split data into train and test-this step divides entire data set into two parts one part is called train(70%-80% ) for training the model and other called test(20%-30%) for validating the model.

6)Model Building-In this we step we select a particular model(s) which can best explain label with the help of featuresand train it with the training data.

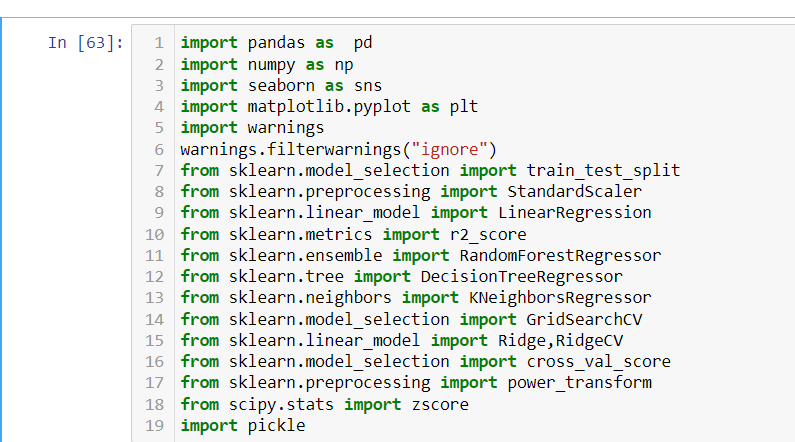
8)Model evaluation-in this step we evaluate the model with the help of various metrics.

9)By checking cross validation score we can select the best model and try to improve its accuracy using hyperparameter tuning .

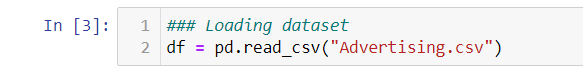
10)Deployement :Finally the best model is to be deployed.

# Step by step process for building the ML model:

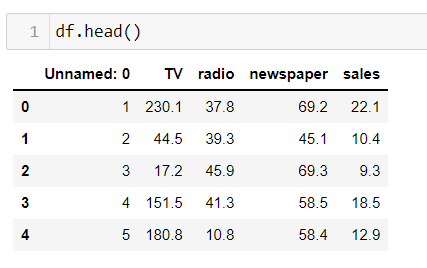
1)First let us import all the necessary libraries in the jupyter neotebook.



2) import the dataset using pandas.



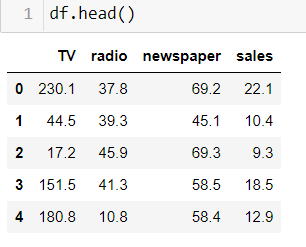
3)Checking sample data.



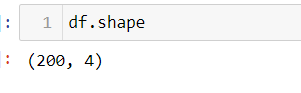
4) As we can see ‘Unnamed: 0 ‘column is nothing but an index or serial number, and it cannot help to predict sales. So, we can drop-off this column using “drop” method in pandas.



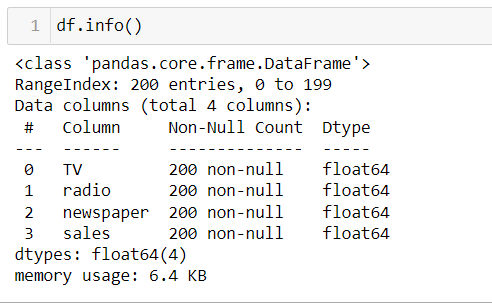
5) Checking sample data again. Now we see that ‘Unnamed: 0 ‘ column is removed.



6)Checking shape of the dataset

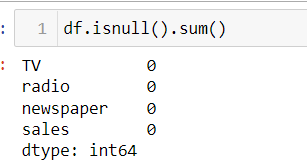


7)Getting info about the dataset.



We see that there are no null values present in any of the column and all the columns are having Dtype float64.So we don’t need any encoding here as there are no object data type.

8) Further confirming the absence of null values.

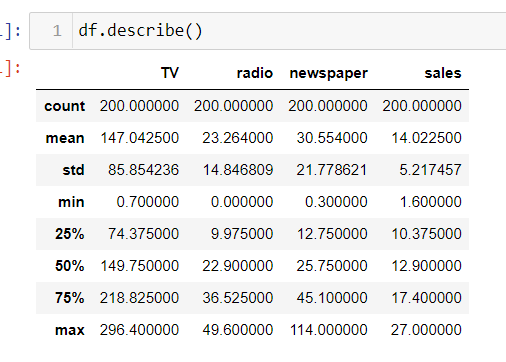


Now we confirmed that there no null values present in the columns.

# Exploratory Data Analysis

# Univariate Analysis:

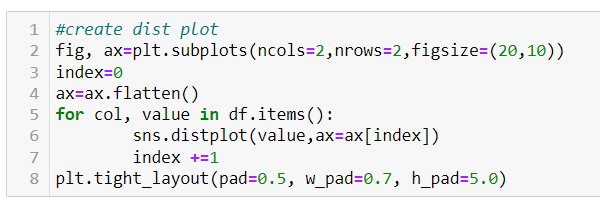
1)Let us check mean, median, standard deviation, minimum, maximum of each column using “describe” method in pandas.

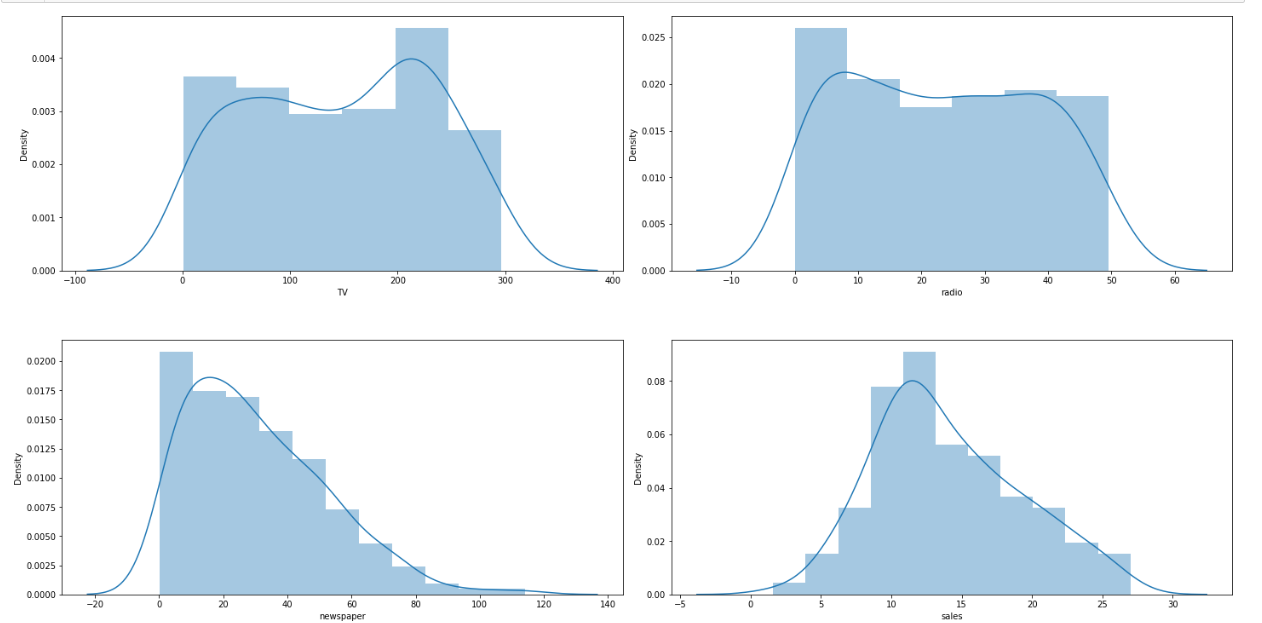


We see that the mean and median(50 percentile) are both nearly same ,which means the distribution of data is approximately symmetric or there might be very less skewness present in the data.

Since there is high difference between 75 percentile and maximum value for “newspaper” column there might be outliers present in this column.

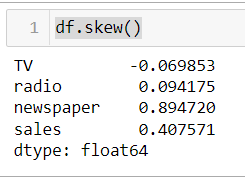
2) Checking distribution of each column using seaborn distplot.





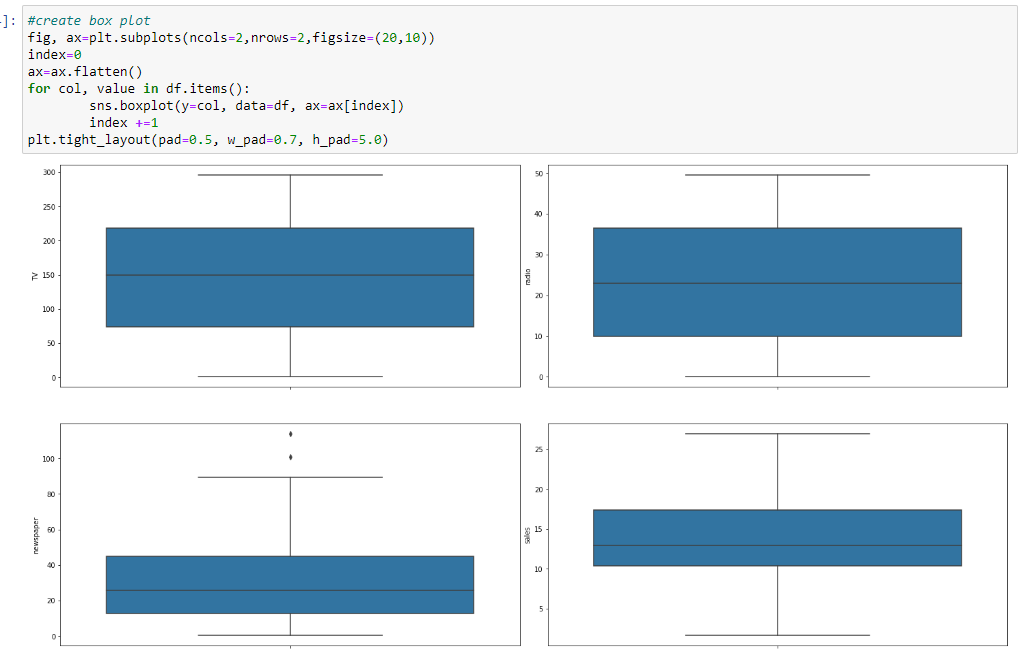
We see that newspaper distribution is rightly skewed.

3) Checking skewness using pandas library.



We see that except for newspaper all the other columns having skewness between -0.5 and +0.5, which is acceptable range for skewness.

4)Checking outliers using box plot with the help of seaborn library.



We see that there are very few or almost no outliers in the data

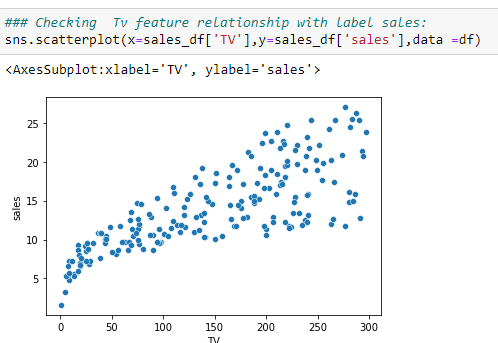
To remove or not to remove outliers depends upon that particular dataset.Outliers can be due to human errors while collecting or measuring the data or they can be due to variability in the data.

If it is due to variability in the data they should not be removed.Otherwise our model wouldnt be able to predict outcome properly.But if the outliers are due to human error,it is considered as noise and must be treated before training the model.Otherwise it will cause overfitting, means we will have good training score but poor testing score.So we need to remove outliers in that case.

# Bi -Variate analysis:

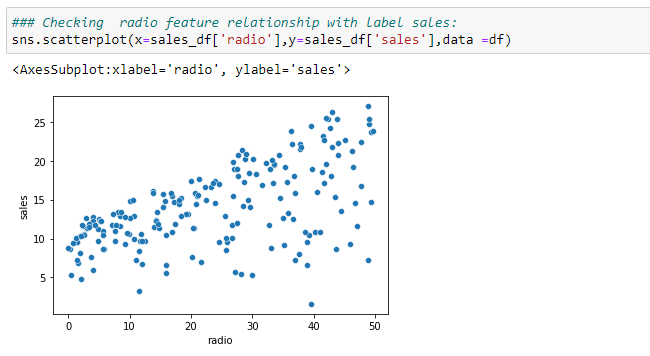
Bivariate analysis gives relationship between two variables.It can be used to check relationship between different varibles in our dataset.

1)Scaterplot is a function in seaborn that can be used to find relationship with two varibles.

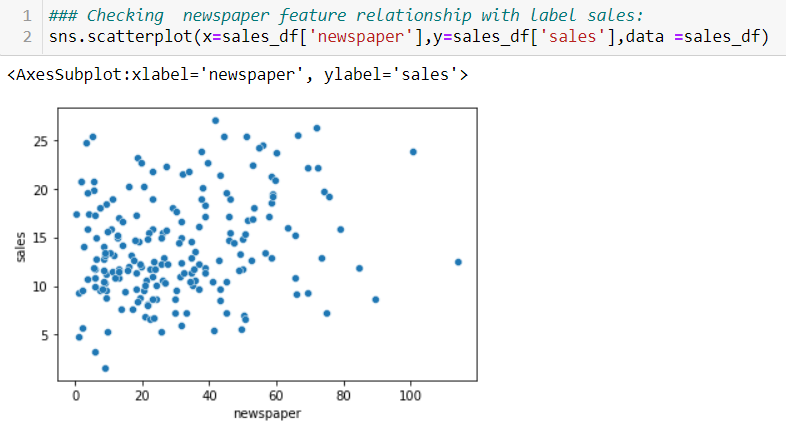


This is the plot between sales and tv.We see that as we invest more in TV promotions there is increase in sales.So there is a positive relationship between sales and tv.

2)This is the plot between sales and radio.We see that as we invest more in radio promotions there is increase in sales.So there is a positive relationship between sales and tv.



3) This is the plot between sales and newspaper. There is no clear relationship between the both.



# Checking Correlation:

Checking correaltion is very important step in the EDA.Correlation gives both magnitude and direction of relationship between different features.And if there is very high corelation(means the two variables are highly corelated) between two features then we have a problem known as ‘multicollinearity’.

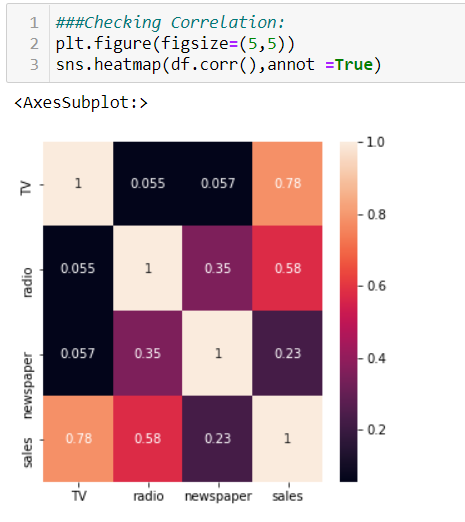
### Why multicollinearity is bad:

Because this will fail with the basic assumption for linear regression which is-“all the input varibles are independent to each other”. Thus regression model will perform very poor incase of multicollinerity model.

### How to remove multicollinearity problem:

We can remove multicollinearity problem using different ways like using correlation matrix and removing features with very high(0.9 above) correlation coefficient,by using VIF method where we remove features with vif score greater than 10,or by using lasso or ridge regression(they automaticall shrink or remove the coefficients of features which are corelated).

Using “corr” method in pandas we can find correltion matrix



Here we see that input features are not very highly correlated with each other. So we don’t have here multicollinearity problem. As from above heatmap, we can also see that TV and radio correlation with sales is more than 50% whereas newspaper has very weak relationship with sales.

## Feature Engineering

### Removing Skewness

As we have seen above the newspaper column is skewed right. Linear regression and logistic regression algorithms assume that the that the features are normally distributed or they follow “Gaussian-distribution”.By measuring skewness in the model one can determine wheter the particular feature is normally distributed or not. If a particular feature is not normally distributed we have got various techniques to convert the data into normal distribution.

Transformation techniques to remove skewness:

1)Logarithmic transformation.

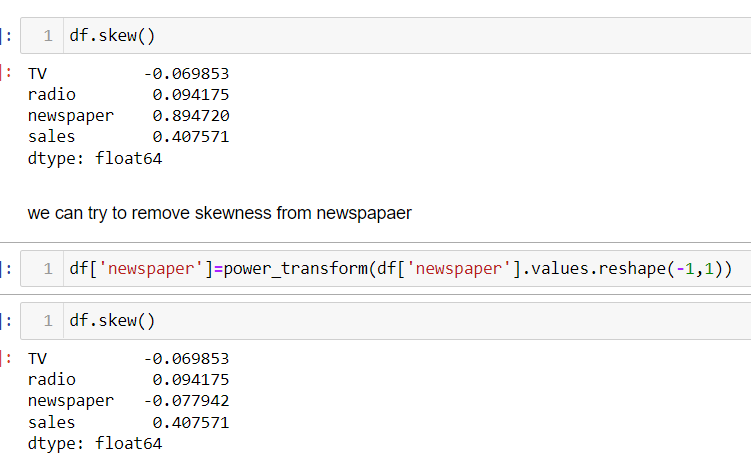
2)Reciprocal transformation.

3)Square root transformation.

4)Exponential transformation.

5)Boxcox transformation.

The above mentioned transformation techniques can be used to make the feature normally distribute.In pandas we have a library called “power\_transform” which uses above mentioned techniques for transformation.

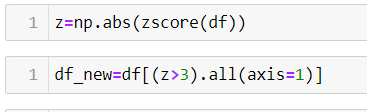


We see that newspaper column’s skewness is in proper range.

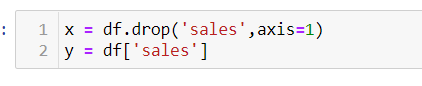
# Removing outliers:

As explained above if outliers are due to human errors they must be removed,the most commonly used techniques to remove outliers are 1)Z-score technique 2)IQR(inter quartile range) technique.

Here I’m showing how to use Z-score method to remove Outliers. According empirical formula rule 99.7% data falls within 3 standard deviations. Using this we can eliminate outliers(if they are there) as follows



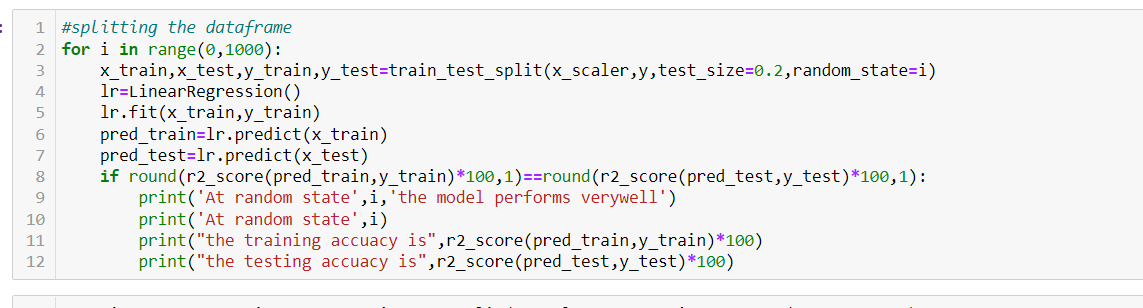
# Seperating features and label:



# Train test splitting:

This is very important step in machine learning life cycle. In this step we split dataset into two parts.One part (70%-80%) is called train set and the other (20%-30%) is called test set.It consists random sampling without replacement,means the splitting is done randaomly.It is compulsory to split the data otherwise training and testing on the same data set will lead to overfitting problem.It cant perform good for unseen data.

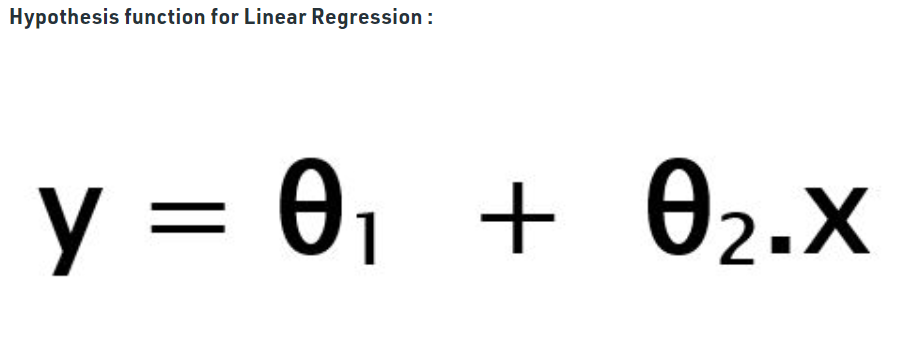
With help of scikit learn libraries we can perform train test split as follows



# Model training :

# Linear Regression:

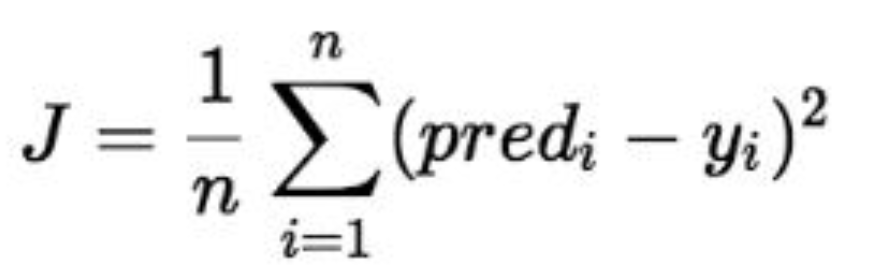
Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.



While training the model we are given :  
**x:** input training data (univariate – one input variable(parameter))  
**y:** labels to data (supervised learning)

When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best θ1 and θ2 values.  
**θ1:** intercept  
**θ2:** coefficient of x

The model gets best values for θ1 and θ2 .There two ways to find the best values for these coeficients.They are 1) Least squared error method-In this method the cost function is minimized to get values of θ1 and θ2



Where J is the cost function

2)Gradient descent- To update θ1 and θ2 values in order to reduce Cost function (minimizing RMSE value) and achieving the best fit line the model uses Gradient Descent. The idea is to start with random θ1 and θ2 values and then iteratively updating the values, reaching minimum cost.

Every model will have some assumptions.So before selecting a particular model we need to check whether the dataset is able to fullfill those assumptions

assumptions for Linear regression model:

1. Linear relationship between target and features.
2. Very low or no multicollinearity problem.
3. There should not be heteroscedasticity-Errors should be random for all the predicted values.There should not be any relationship between errors and predicted values.
4. Normal distribution of errors.
5. All the features are independent of each other.

1)Importing the model from the libraries of scikit learn(already these libraries are imporeted in step one)



2)Instantiating and training the model.



3)Model validation and evaluation.

1)First we can import necessary evaluation metrics for regression model.

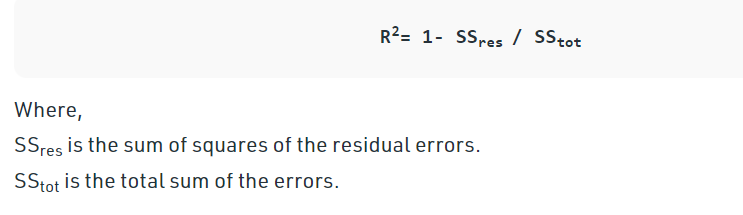


2) pred\_test gives the predicted values for the features x\_test

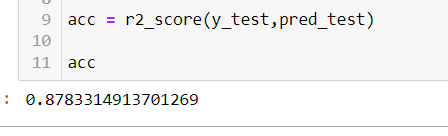


3)Then,we find r2score. r2\_score is the metric for continuous data prediction.It is called ‘Coefficient of determination’.it is gives the proportion of variation in the dependent variable that is predicted from independent variables.

Mathematical Formula:



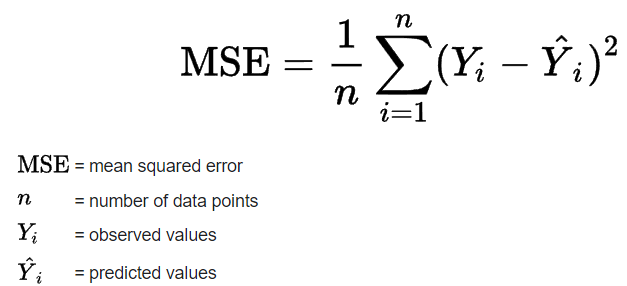
We can find r2score using scikit library which we imported.



So we got r2 score 0.878,it means 87.8% of the changeability of the dependent output attribute can be explained by the model while remaiming 12.2% of the varibility is unaccounted for .The better the r2 score the better is the model.

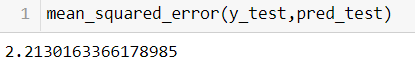
4)Using other metrics like mean absolute error,mean squared error

Mean Squared error:it is the most common regression metric .It is the average of the squared difference between the predicted and actual value. Since it is differentiable due to shape, it is easier to optimize.Its formula is



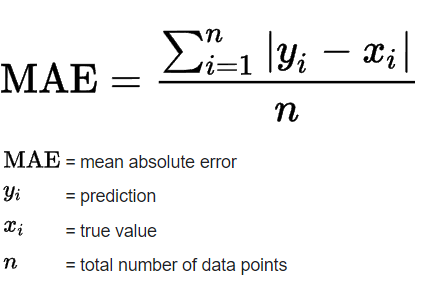
MSE penalizes large errors.

We can find MSE using scikit library which we imported.

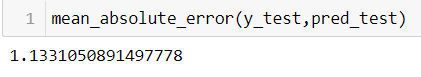


Mean Absolute error:

This is simply the average of the absolute difference between the target value and the value predicted by the model. Not preferred in cases where outliers are prominent.



We can find MAE using scikit library which we imported



Cross Validation:

Rightnow we are training our model on one partcular dataset and getting good accuracy.By this approach of using train test split we cant generalize our model.Our model moghtbe working good for this particular dataset,we cant say how it works for real-time data.There might be problem of overfitting,etc.So, Cross-validation is a technique for evaluating ML models by training several ML models on subsets of the available input data and evaluating them on the complementary subset of the data. We use cross-validation to detect overfitting, ie, failing to generalize a pattern.

Following are different methods for implementing cross validation:

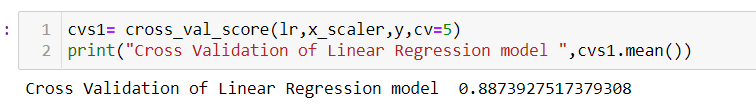
1.Hold Out method-this is the one we used above ie splitting the data into training(70%-80%) and testing(20%-30%) and then training the model using training dataset.As discussed above it is not very efficient because there can be overfitting.

2. Leave One Out Cross-Validation-In this method, we divide the data into train and test sets – but with a twist. Instead of dividing the data into 2 subsets, we select a single observation as test data, and everything else is labeled as training data and the model is trained. Now the 2nd observation is selected as test data and the model is trained on the remaining data.

3. K-Fold Cross-Validation-In this resampling technique, the whole data is divided into k sets of almost equal sizes. The first set is selected as the test set and the model is trained on the remaining k-1 sets. The test error rate is then calculated after fitting the model to the test data.

In the second iteration, the 2nd set is selected as a test set and the remaining k-1 sets are used to train the data and the error is calculated. This process continues for all the k sets.

We are implementing K-Fold Cross-Validation for our model,we can use scikit learn library to calculate cross validation scores and then take mean to calculate average score for our model.



# Regularization:

Overfitting and Underfitting are the two main problems that occur in machine learning and degrade the performance of the machine learning models.

Overfitting occurs when our machine learning model tries to cover all the data points or more than the required data points present in the given dataset. Because of this, the model starts caching noise and inaccurate values present in the dataset, and all these factors reduce the efficiency and accuracy of the model. The overfitted model has low bias and high variance.

In the case of underfitting, the model is not able to learn enough from the training data, and hence it reduces the accuracy and produces unreliable predictions. An underfitted model has high bias and low variance.

How to avoid Overfitting and Underfitting issues?

We can avoid these problems with the help of regularization. Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting and underfitting. Regularization works by adding a penalty or complexity term to the complex model.

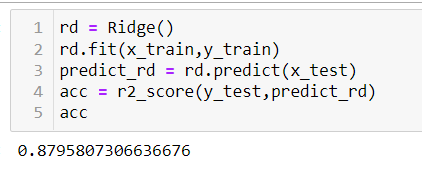
The commonly used regularization techniques are : 

1. L1 regularization or LASSO regression- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.It can shrink the coeficients of best fitting line to zero.
2. L2 regularization or Ridge regression: Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization. It can shrink the coeficients of best fitting line but it will not make them zero like Lasso regression.
3. Elasticnet:Elastic netis combination of both Lasso and ridge

Implementing Regulariztion(Ridge regression) :

First let us import scikitlearn library:





Wesee that the ridge regression is almost giving the same r2score as Linear regression.So there is no problem of underfitting in our dataset.

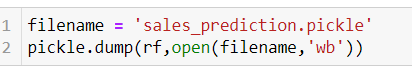
# Model Saving

In machine learning, while working with scikit learn library, we need to save the trained models in a file and restore them in order to reuse them to compare the model with other models, and to test the model on new data.

1)Lets import pickle library



2)Saving the model



Thus we implemented complete life cycle of a machine learning model.