**Experiment No. 7**

**Title: Python Linear Regression using Sklearn**

Batch:B2 RollNo:1914078 ExperimentNo.:7

### Aim: Building Linear Regression model and its evaluation using sklearn

Resources needed:Python IDE

### Theory:

 scikit-learn is an open source Python library that implements a range of machine learning, pre-processing, cross-validation and visualization algorithms using a unified interface.

**Important features of scikit-learn:**

* Simple and efficient tools for data mining and data analysis. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means, etc.
* Accessible to everybody and reusable in various contexts.
* Built on the top of NumPy, SciPy, and matplotlib.
* Open source, commercially usable – BSD license.

**Splitting the dataset**

* Split the dataser into two pieces a training set and a testing set.
* Train the model on the training set.
* Test the model on the testing set, and evaluate how well our model did.

# splitting X and y into training and testing sets

**fromsklearn.model\_selection importtrain\_test\_split**

X\_train, X\_test, y\_train, y\_test =train\_test\_split(X, y, test\_size=0.4, random\_state=1)

print(X\_train.shape)

print(X\_test.shape)

**Training the model**

 Scikit-learn provides a wide range of machine learning algorithms which have a unified/consistent interface for fitting, predicting accuracy, etc.

# training the model on training set

**fromsklearn.linear\_modelimportLinearRegression**

reg=LinearRegression().fit(X\_train, y\_train)

# making predictions on the testing set

y\_pred =reg.predict(X\_test)

# comparing actual response values (y\_test) with predicted response values (y\_pred)

**fromsklearn.metrics import mean\_absolute\_error,mean\_squared\_error,r2\_score**

print('mean absolute error',mean\_absolute\_error(y\_test,y\_pred))

### Activities:

### Download data set appropriate for building Linear Regression model.

### Slice the predictors and target into variables X and y respectively

### Perform train test split on the data.

### Build Linear Regression model.

### Find the Mean Absolute Error, Root Mean Squared Error and R2 Score.

### Result: (script and output)

### 

### Dataset

### Finding correlation between Price and single variables

### 

### 

### Regression model using vs Price

### 

### 

### 

### 

### 

### 

### 

### 

### 

### Questions:

### 1. What is multicollineatry?

### Multicollinearity occurs when the model includes multiple variables that are not only correlated to the target variable but also with each other.

### 2. How does multicollinearity affect the Linear Regession?

### If the variables are correlated amongst each other, if one of them changes, the model results fluctuate significantly.

### Outcomes: Illustrate python libraries for machine learning and image processing

**Conclusion:** Built a Regression Model using sklearn to predict target variable in my dataset.

**References:**

1. 1.Yuxi Liu ; Python Machine Learning By Example: The easiest way to get into machine learning 1st Edition, Kindle Edition , Packt publishing Ltd , 1st edition 2017.
2. Martin C. Brown, The Complete Reference Paperback, Osborne, 2nd edition 2001
3. Frank Millstein, Data Analytics with Python: Data Analytics In Python Using Pandas, Copyright at Frank Millstein, 1st edition 2018
4. <https://www.geeksforgeeks.org/learning-model-building-scikit-learn-python-machine-learning-library/>
5. <http://scikit-learn.org/stable/documentation.html>