DSBDAL

Assignment No :- 04

Title :- Data Analytics I

objectives :- student should be able to data analysis using linear regression using python for any open source dataset.

Aim:- create linear regression model using python IR to predict home prices using Boston housing Dataset. The Boston Housing dataset contains information about various houses in data through the parameters.

Requirements:-
1) Basic of python programming.
2) concept of Regression.

Theory:- Boston Housing With linear regression With this data our objective is create a model using linear regression to predict the houses price.

The data contains the following columns:-

1) crim :- per capita crime rate by own

2) 'zn' :- properties of residential and zoned for lots over 25000 sq.ft.

3) 'indus' :- proportion of non-retail business acres.

4) 'chas' :- Charles river dumy variable (=1 if tract bounds river ; 0 otherwise).

5) 'nov' : nitrogen oxides concentration.

6) 'rm' :- average number of rooms per dwelling.

7) 'Age' :- proportion of owner-occupied units built prior to 1940.

8) 'dis' :- Weighted mean of distances to five Boston empty centres.

9) 'tax' :- full-value property-tax rate per $10,000.

10) 'rad' :- index of accessibility to radial highways.

11) 'ptratio' :- pupil-teacher ratio by town.

12) 'black' :- $1000(Bk-0.63)^2$ where Bk is the proportion of blacks by town.

13) 'Istat' :- lower status of the population.

14) 'medv' :- median value of owner-occupied homes in $$ 1000s.

* Conclusion:- Thus we learn about how to analysis data using linear regression using python.

* CSV file/ Dataset - Boston - dataset :-

Required libraries -

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.datasets import load_boston.
from sklearn.model_selection import train_test_split.
from sklearn.linear_model import linear regression.
from sklearn.metrics import mean_square_error_mean_absolute_error
from sklearn.preprocessing import standard scaler import Hamings.
```

function used: -

```
boston = load_boston()
df.head()
df.shape
df.info()
df.describe()
df.isnull().sum()
plt.figure()
sns.heatmap()
plt.show()
sns.pairplot()
scaler = standardscaler()
model = linearRegression()
sns.regplot()
```

Q.① What is mean by Regression ?

→ Relationship between dependent & independent variable is called as regression.

Q.2) Explain linear Regression ?

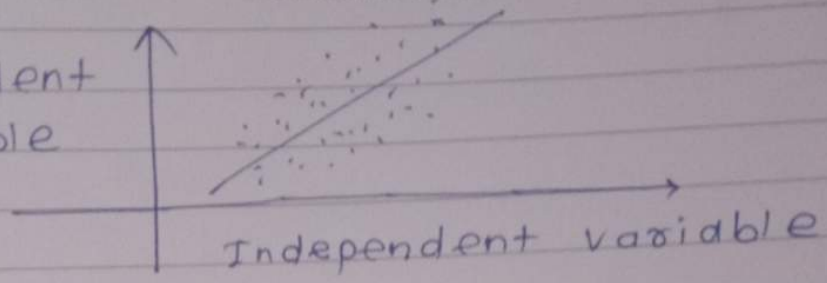
→ linear regression is a type of stastical analysis used to predict the relationship between two variables.

Q.3) Explain multiple linear Regression.

→ multiple independent variables available.

Q.4) Explain simple linear Regression.

→ In that one independent variable one dependent variable.



Q.5) Explain logistic Regression.

→ - linear regression is unbounded, so logstics regression come in picture.
- logstics regression is used when the dependent variable is categorical.

Q.6) Explain boston Housing Dataset.

→ 1) How many features in dataset (column).
2) How many rows or records in the dataset.
3) How to add column in dataset.

Q.7) Applications for logstics regression.

→ 1) Health care
2) Demand forecasting
3) Sports outcomes
4) fitness.

Q.8) Explain Polynomial regression.

→ In the linear regression eqvation power of the independent variables Has 1.

- In polynomial regression the power of independent variable is greater than 1.