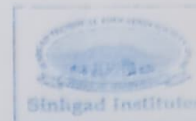Title of the Assignment: Create a Linear Regression model
using python/R to predict home prices using Boston
Housing dataset.

Objective of the Assignment: students should be able to
data analysis using linear regression using python
for any open source dataset.

Prerequisitel-
1. Basic of python programming.
2. Concept of Regression.

Concepts for Theory:-
1. Linear Regression: Univariate and Multivariate
2. Least square method for Linear Regression
3. Measuring Performance of Linear Regression
4. Example of Linear Regression
5. Training data set and Testing dataset.

1. Linear Regression:- It is a machine learning algorithm based
on supervised learning. It targets prediction values on
the basis of independent variables.

• It is prefeared to find out the relationship between for
using and variable.

• A linear relationship between a dependent variable(x) is
continous while independent variable (r) relationship may
be continous or discrete.

• Linear regression is popular because the cost function
is mean squared Error (MSE) which is equal to the
average squared difference between an observation actual
and predicted values.

Multivariable Regression: It concern the study of two or more predictor variables usually a transformation of the original features into polynomial features. from a given degree is preferred Linear Regression is applied anib.

2. Least square method for Linear Regression.
- Linear Regression involves establishing linear relationships between dependent independent variables.
- A simple linear model is the one which involves only one dependent and more one independent variable.
- ~~variable~~, However, for a simple univariable linear model, it can be denoted by the regression equation
$$\hat{Y} = B_0 + B_r$$

- This linear requation represents a line also known as the regression line The least square estimation technique is one of the basic techniques used to guess the values of the parameters and based on sample set.

3. Measurning Performance of Linear Regression
Mean square Error (MSE)
The mean squared error (MSE) represents the error of the estimator or predictive model created based on the given set of observation in the sample. Two or more regression models created using a given sample data can be compared based on their MSE. The lesser the MSE, the better the regression model.

Interpretation 2

Even if x=0 value of independent variable, it is expected that value of y is 26.768

5. Training data set and Testing data set.
- Machine Learning algorithm has two phases.
  1. Training and 2. Testing

- The input of the training phase is training data which is passed to any machine learning algorithm and machine learning model is generated as output of training phase.

[a] Training phase
- Training phase dataset is provided as input to this phase
- Training dataset is dataset having attributes and class labels and used for training machine learning to prepare models.
- In a simple way the actual output of training data and practical output of the models does not match the training error. it is said to have occured.

[b] Testing phase
- Testing dataset is provided as input to this phase
- Test dataset is dataset for which class labels is unknown. it is tested using models.
- A test dataset used for assessment of the finally chosen module.
- In a simple way the actual output of the testing data and provided output of the modules does not match the testing error Eour is said to have occured.

$$MSE = \frac{1}{n} \underbrace{\sum (y - \bar{y})^2}$$

the square of the difference
between actual and predicted.

An MSE of zero(o) represents the fact that the predictor
is perfect predictor. RMSE
Root mean squared Error method that basically calculates
the least-squared error and takes a root of the
summed values.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{1}{n} (\bar{y}_i - y_i)^2}$$

A value of R-squared closer to 1 would mean that the
regression model covers most part of the variance of the
values of the response variable and canbe termed or a good.

4. Example of Linear regression.
consider following data for 5 Students.
Each x_i (i=1 to 5) represents the score of ith student in
standard x and corresponding y_i (i=1 to 5) represents the
score of ith student XII.

i) Linear regression equation best predicts standard XIIth score
ii) interpretation for the equation of linear regression.

ii) interpretation of the regression Line.
interpretation !
For an increase in value of x by 0.544 units
these is an increase in value of by one unit.

c] Generalization

- Generalization is the prediction of the future based on the past system
- It needs to generalize beyound the training data to some future data that it might not have seen yet.
- The ultimate aim of the machine learning models is to minimize the generalization error.
- The general, the dataset is divided into two Partition having training and test Sets

- Conclusion:-
    In this way we have done data analysis using linear regression. Dataset and predict the price of houses using the features of the Boston dataset.