



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 3, March 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 8.379**



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

# Linear Regression: A Practical Implementation in Python

Prof.Sarita Patil, Prof. Nitin Shirirao

Assistant Professor, Siddhant Institute of Computer Application, Savitribai Phule Pune University, Pune, India

Director, Siddhant Institute of Computer Application, Savitribai Phule Pune University, Pune, India

**ABSTRACT:** Linear Regression is a supervised machine learning algorithm. It tries to find out the best linear relationship that describes the data you have. It has two assumptions. First is that there exist a linear relationship between a dependent variable and independent variable. The second is a dependent variable is continuous value i.e. real numbers.

**KEYWORDS:** ML, SL, Linear regression, Analysis.

## I. INTRODUCTION

Machine learning is a subfield of artificial intelligence that focuses on the design of system that can learn and make decisions and make prediction based on experience or previous data which we have provided the system. Machine learning enables computer to act and make data driven decision rather than being explicitly programmed to carry out a certain task. These programs are designed to learn and improve over time when exposed to new data.

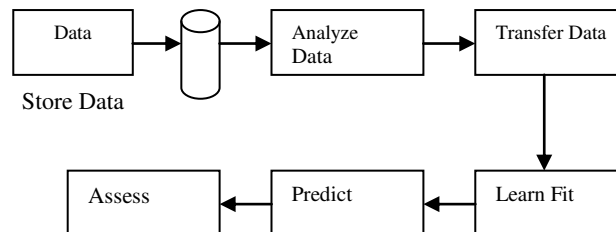


Fig. 1.1 Machine Learning

### Types of Machine Learning:

There are three types of machine learning are as:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

Supervised learning:

Supervised learning model provides labeled data (input and output data) to the AI system, and on the basis of that data, model predict the output.

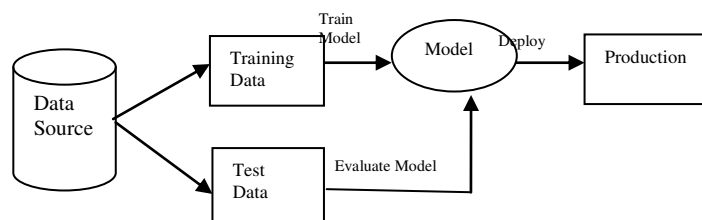


Fig.1.2. Supervised Machine Learning workflow

There are two types of supervised learning techniques are as:

1. Classification:-Classification is about predicting a class or discrete values.  
E.g.: Male or Female
2. Regression: - Regression is about predicting a quantity or continuous values.  
E.g.: Salary, Age

Classification:

1. Random Forest
2. Decision Tree
3. Logistic Regression
4. Support Vector Machine

Regression:

1. Linear Regression
2. Regression Tree
3. Non-Linear Regression Tree
4. Bayesian Linear Regression
5. Polynomial Regression

In this paper we have covered linear regression algorithm based on supervised learning algorithm.

### Regression Analysis:

Regression Analysis is the measure of the average relationship between two or more variable in terms of the original units of the data. Linear Regression is a type of Regression analysis.

## II. LINEAR REGRESSION

Linear regression algorithm is a machine learning algorithm based on supervised algorithm. Linear regression algorithm is focus on the relationship between a dependent variable and one or more independent variable. Linear regression shows the linear relationship, it finds how the mean value of one variable is changing according to the like values of the other variable. Linear regression provides a sloped straight line representing relationship between variables.

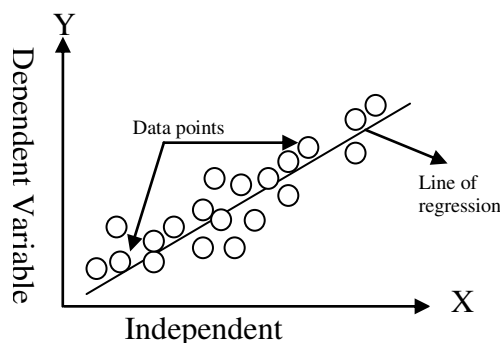


FIG. 1.3. LINEAR REGRESSION GRAPH



Mathematical linear regressions are as follows:

$$Y = a + bX \quad (1)$$

Here,

Y = Dependent Variable

X = Independent Variable

b = Slope of the line

a = y-intercept

X and Y variables are training datasets of linear regression.

$$a = \frac{\bar{y}(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

### III. TYPES OF LINEAR REGRESSION

There are two types of linear regression algorithm model are as follows:

#### 1. Simple Linear Regression:

Simple linear regression model we have two way relationship and this two way relationship is consist of x and y. Here y generally consists of as the dependent variable and x is considered as the independent variable.

In the form of the equation in the simple linear regression model is:

$$Y = b_0 + b_1X + \mu \quad (2)$$

Here,

X = Independent Variable

Y = Dependent Variable

$b_0$  = Autonomous consumption

$b_1$  = Proportion

$\mu$  = Error Term

#### 2. Multiple Linear Regressions:

Multiple linear regression model is a dependent variable is depends on more than one independent variable. Multiple linear regression model equation becomes:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + \epsilon \quad (3)$$

Here,

k independent variables:  $x_1, x_2, x_3, \dots, x_k$

Regression coefficients:  $b_0, b_1, \dots, b_k$

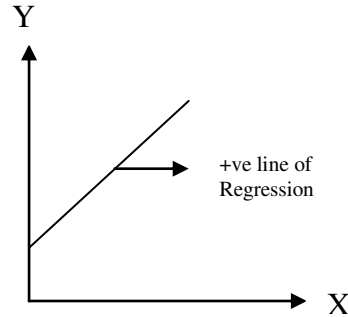
$\epsilon$ : Random error

#### Linear Regression line:

A regression line can show two types of relationship:

- **Positive Linear Relationship:**

In positive linear is a dependent variable is increases in Y axis and independent variable are increases in X axis.

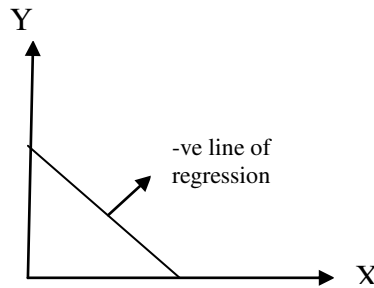


The positive line equation will be:

$$Y = a_0 + a_1X$$

• **Negative Linear Relation:**

In negative linear is a dependent variable is decreases in Y axis and independent variable are increases in X axis.



The negative line equation will be:

$$Y = -a_0 + a_1X$$

**IV. MATHEMATICAL EXAMPLE OF LINEAR REGRESSION:**

**Q.** TO SOLVE THE LINEAR REGRESSION EQUATION OF THE GIVEN DATA SETS.

X	3	2	4	6
Y	3	6	5	10

SEP 1: TO FIND THE VALUE OF  $\sum X$ ,  $\sum Y$ ,  $\sum X^2$  AND  $\sum XY$

CONSTRUCT THE TABLE AND FIND THE VALUE

X	Y	X <sup>2</sup>	XY
3	3	9	9
2	6	4	12
4	5	16	20
6	10	36	60
$\sum x=15$	$\sum y=24$	$\sum x^2=65$	$\sum xy=101$



**Step 2:**

To find the value of a

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

Put the value in the equation

$$a = \frac{24*65 - 15*101}{4*65 - 65}$$

$$a = -\frac{45}{195}$$

$$a = 0.23$$

Put the value in the equation

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{4*101 - 15*24}{4*65 - (15)^2}$$

$$b = \frac{44}{35}$$

$$b = 1.25$$

**Step 3:**

Hence we got the value of a= 0.23 and b= 1.25

Now put the value of a and b in the equation (1)

$$Y = 0.23 + 1.25X$$

**V. IMPLEMENTATION OF LINEAR REGRESSION USING PYTHON**

Step 1: Importing the dataset

Step 2: Data pre-processing

Step 3: Splitting the test and train sets

Step 4: Fitting the linear regression model to the training set

Step 5: Predicting test results

Step 6: Visualizing the test results

**Step 1: Import the libraries and datasets**

First we will import the dataset using numpy, pandas and matplotlib libraries

import numpy as np



```
import pandas as pd
import matplotlib.pyplot as plt
dataset = pd.read_csv('Salary_Data.csv')
dataset.head()
```

The dataset.head() shows the columns of dataset. The output of the above program is:

Out[1]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

### Step 2: Data Preprocessing:

Now we have imported dataset, we will perform data preprocessing

```
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values
```

The x is indicating independent variable of array and y is indicating dependent variable of vector. The dependent variable is must be vector and independent variable is must array.

### Step 3: Splitting the datasets

Split the datasets into two sets, first is test set and second is train set.

In this dataset train our model according to the years and salary. We then test our model on the data set.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=1/3, random_state=0)
```

### Step 4 : Fitting linear regression model into the training set

In this step we have to use sklearn's linear model library and import linear regression class. In this step we have to create an object for linear regression class it is called as regressor. In this method we have to call fit method.

```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

### Step 5: Predicting Test Result

In this method we have to use predict method. In this method we make predictions for the test sets. In this method we are use only array and sparse matrix.

```
y_pred = regressor.predict(X_test)
y_pred
```



```
array ( [ 40835.10590871, 123079.39940819, 65134.55626083, 63265.36777221, 115602.64545369, 108125.8914992, 116537.23969801, 64199.96201652, 76349.68719258, 100649.1375447 ] )
```

**y-pred output**

**y\_test**

```
array( [ 37731. , 122391. , 57081. , 63218. , 116969. , 109431. , 112635. , 55794. , 83088. , 101302. ] )
```

**y-test output**

y\_test is real salary of the test set.

y\_pred is predicted salaries.

### Step 6: Visualizing the test results

#### Plot for train

```
plt.scatter(X_train, y_train, color='red')
plt.plot(X_train, regressor.predict(X_train), color='blue')
plt.title("Salary vs Experience (Training set)")
plt.xlabel("Years of experience")
plt.ylabel("Salaries")
plt.show()
```

#### Output:



#### Plot for Test

```
plt.scatter(X_test, y_test, color='red')
plt.plot(X_train, regressor.predict(X_train), color='blue')
plt.title("Salary vs Experience (Testing set)")
plt.xlabel("Years of experience")
plt.ylabel("Salaries")
```



plt.show()

## Output



## VI. CONCLUSION

Linear regression technique is focus on relationship between dependent variable and independent variable. In this article, we have solved linear regression mathematical example and implement linear regression using python using salary datasets.

## REFERENCES

1. Schneider A, Hommel G, Blettner M. Linear regression analysis: Part 14 of a series on evaluation of scientific publications. *DtschArzteblInt* 2010;107:776-82.
2. Freedman DA. *Statistical Models: Theory and Practice*. Cambridge, USA: Cambridge University Press; 2009.
3. Chan YH. *Biostatistics 201: Linear regression analysis*. Age (years). *Singapore Med J* 2004;45:55-61.
4. Panchenko D. 18.443 *Statistics for Applications, Section 14, Simple Linear Regression*. Massachusetts Institute of Technology: MIT OpenCourseWare; 2006.
5. Elazar JP. *Multiple Regression in Behavioral Research: Explanation and Prediction*. 2nd ed. New York: Holt, Rinehart and Winston; 1982
6. Yumao Li. Theory and Application of Univariate Linear Regression Method [J]. *Journal of Chifeng University (Natural Science Edition)*, 2017,33 (15): 1-2. (in Chinese)
7. Goldberger, Arthur S. (1964). *Classical Linear Regression. Econometric Theory*. New York: John.
8. YidanShen. *Simple and clear mathematical dictionary*[M]. Press: Beijing Institute of Technology-y Press,2007 (in Chinese).
9. Trevor Hastie. *The elements of Statistical Learning*: Springer,2009: P11.
10. Yumao Li. Theory and Application of Univariate Linear Regression Method [J]. *Journal of Chifeng University (Natural Science Edition)*, 2017,33 (15): 1-2. (in Chinese)



INNO  SPACE  
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

 **doi**<sup>®</sup>  
**CROSS** **ref**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  [ijircce@gmail.com](mailto:ijircce@gmail.com)



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details