



ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 8.4
IJAR 2021; 7(8): 273-275
www.allresearchjournal.com
Received: 11-06-2021
Accepted: 13-07-2021

Siddhant Patil

Department of Computer
Science Engineering, All India
Shri Shivaji Memorial Society's
College of Engineering,
Savitribai Phule Pune
University, Pune,
Maharashtra, India

Shruti Patil

Department of Computer
Science Engineering, All India
Shri Shivaji Memorial Society's
College of Engineering,
Savitribai Phule Pune
University, Pune,
Maharashtra, India

Corresponding Author:

Siddhant Patil

Department of Computer
Science Engineering, All India
Shri Shivaji Memorial Society's
College of Engineering,
Savitribai Phule Pune
University, Pune,
Maharashtra, India

Linear with polynomial regression: Overview

Siddhant Patil and Shruti Patil

DOI: <https://doi.org/10.22271/allresearch.2021.v7.i8d.8876>

Abstract

In the current world, there is a need to analyze and define the relations from the data and predict outcomes for profits. Regression is a Machine Learning technique that involves finding correlations between variables and predicts a continuous output. It helps us to understand how the value of the dependent variable (target) is changing corresponding to an independent variable (predictor). The aim of this paper is to discuss linear regression with its types, polynomial regression, and the relationship between Linear Regression and Polynomial Regression and how they are interrelated. These techniques are widely used to find the trends in data and forecast some outcomes. The aforementioned techniques are explained and analyzed based on the factors like the size of the dataset, type of the data set, quality, efficiency, consistency, accuracy, variables, and performance. These methods create a visual graph that can be used for predicting various past and future outcomes. The intent of discussing the relationship between the techniques is to assist new researchers and beginners to understand how they function, so they can come up with new approaches and innovations for improvement.

Keywords: regression, linear regression, polynomial regression, simple regression, multiple regressions

Introduction

Regression is an example of supervised learning, in which the machines are trained ahead of time using the labeled training data sets, and on the basics of this training, the machine predicts the output. 'Labeled' data means that some of the data is already labeled with the correct/right output. Regression can be considered as one of the core concepts which has over time enticed a lot of attention from Machine learning and researchers due to its wide range of applications from forecasting trends, forecasting events, and so on. The main intent of the regression algorithm is to explain the relation of one variable (called dependent or target) changing corresponding to one or more other variables (called independent or predictors). Regression is one of the most favored algorithms in various fields which include business and retail, agricultural scientists, medical researchers, data scientists, the study of sediments and, so on.

There are various types of regression approaches, including linear regression, polynomial regression, logistic regression, lasso regression, support vector machine (SVM), ridge regression, each of which follows different algorithms and mechanisms. Various researchers are still developing new approaches for algorithms. Briefly, we have discussed the somewhat resemblance of Linear regression with Polynomial regression. Both models are effectively used for various purposes and scientific researches and many researchers are still developing various approaches for more precision of these algorithms and methods. In this paper, we will see the relationship between linear and polynomial regression and how it affects the models.

Regression

Regression can be said as a relation between two (or more) entities depending upon the type of data set and the end output we desire to achieve. As we are dealing with Supervised Learning the accuracy and the structure of the data set play a vital role in forming these regression models. The data is pre-processed to make it suitable for the model and which, in turn also increases the efficacy and accuracy of the model. Based on the characteristics of variables and the data set, a graph is plotted on the coordinates to examine the behaviour of

the dataset. The numbers of data points that are plotted are defined by analyzing and performing operations on datasets. The more structured and systematic model we develop, the more accuracy and efficacy of the coming outcomes will increase. Regression can be used for various purposes and approaches. In agriculture, it can be used to predict the methods and their productivity over past years to gain profit in the future, and many such real-life examples. The term 'Regression' as we know it today was developed by Francis Galton in the 19th century during his research on 'heights of descendants of the tall ancestors regress with time'; however, the method was recognized and performed in the earlier 18th century unknowingly.

Linear Regression

Linear regression is a form of regression in which it shows a linear relationship, which in turn it finds the correlation between the targets or dependent variable in accordance to the predictor of the independent variable (s) as said in [3]. This method works best when the data set usually given is a linear data set which in turn forms a linear relationship on the plot. There can be various relations determined by the manifestation of the graph– Positive, Negative and, No relation.

Steps for Linear Regression are

- Step 1:** Read the data and make sure it meets the expectations.
- Step 2:** Perform the Linear regression analysis to evaluate the relationship.
- Step 3:** Check for any error/miscalculations and visualize the data on a graph plot.

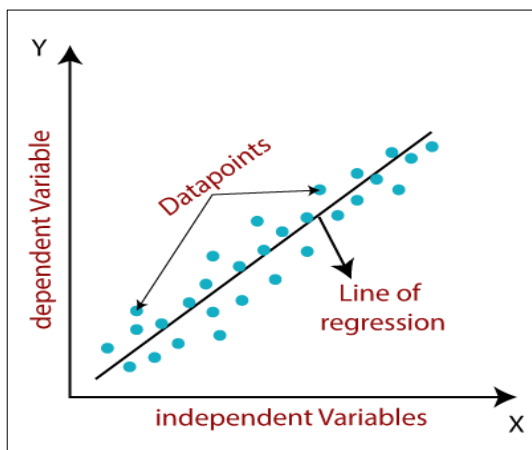


Fig 1: Linear regression

Linear regression can be further sub divided as

A. Simple Linear Regression

Simple regression is an approach of linear regression where the target value is determined using 'only one' predictor value. The target value must be a continuous or real value; however, the independent values can be categorical or continuous. The equation for simple regression:

$$y = a_0 + a_1x + \epsilon$$

Equation for Simple Regression

Where, a_0 =intercept of line, a_1 = Slope of Line, ϵ = error/ miscalculation.

x, y = predictor and target respectively.

B. Multiple Linear Regressions

This form of linear regression is performed when there are 'more than one' predictor values which are having cause-effect relation [2] with the target value as in. It is an annexure to the simple regression. In this method, there are small to no chances of Multicollinearity between the predictors, which means there is no inter-correlation between the predictors' values. The equation for multiple regressions is as follows:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

Equation for Multi Regression

Where, y = Output/Response variable
 $b_0, b_1, b_2, b_3, \dots, b_n$ = Coefficients of the model.
 $x_1, x_2, x_3, x_4, \dots$ = predictor value.

Error Calculation Steps

Linear regression often uses Mean Square Error (MSE), to calculate the errors or miscalculations.

1. Measure the distance of the observed y value on the regression line from the predicted y value at every x value.
2. Square each of this distance.
3. Then calculate the mean of these squared distances.
4. The error should be minimum for maximum accuracy and efficiency.

Polynomial Regression

Polynomial regression can be defined as a special case of linear regression; it is used where the linear regression model has low accuracy/efficiency and more errors or miscalculations. In this method, we tend to fit the regression line using a polynomial equation which fits the data in the best curvilinear fashion possible to get the maximum accuracy with avoiding the effect of over-fit or under-fit. The dataset passed through this model is mostly in a non-linear fashion as linear regressions do not work accurately here.

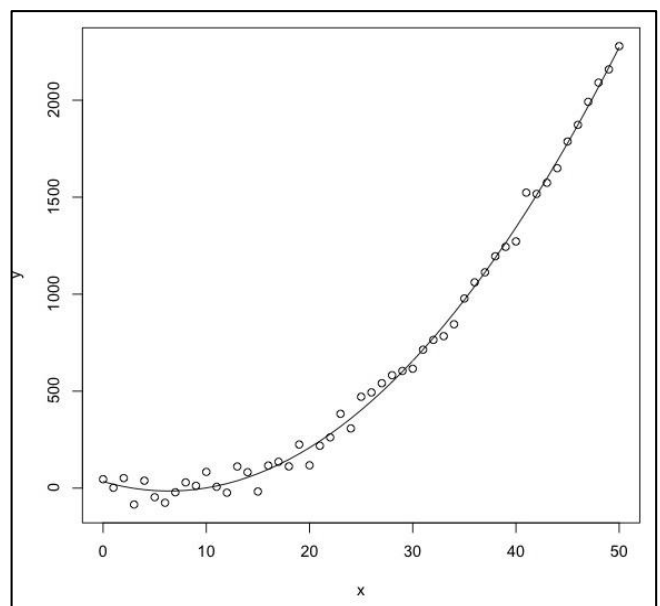


Fig 2: Polynomial Regression

Comparing the equations

Linear regression

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 + \dots + b_nx^n \dots (1)$$

Polynomial Regression

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 + \dots + b_nx^n \dots (2)$$

If we look at the equations and notice, polynomial regression equation (2) is also a form of linear regression equation (1) but only differ by the degree of variables of the equations. The linear equations always have a degree of one. The polynomial equation does not depend on variables but instead, it depends on the coefficients which are arranged in a linear fashion, therefore it can also be called a polynomial linear regression equation.

From this, we can clearly notice how polynomial regression is in use where there is a non-linear dataset where linear regression has low accuracy. The correlation of linear and polynomial regression can be described through this.

Discussion

The method of linear regression and ^[1] polynomial regression as used in is one of the examples where the relationship between linear and polynomial regression was employed successfully. Using linear regression the overview for covid-19 cases fatality and recovery scenario was observed followed by polynomial regression where the expected number of patients was predicted up till certain period of time by training and validating the model by the dataset. Using this ^[1] we got an overview of the outbreak of covid-19 in India.

Furthermore, in ^[4], this same methodology of the relationship of linear and polynomial regression was used with some extra steps for investigating the outbreak of covid-19 particularly for some states of India focusing on individual impact rates respectively.

Since this relational approach ensures to give an insight of the data through linear regression and furthermore predict the expected output through polynomial regression, we may conclude that this methodology is more insightful to overview and predict the relationships.

Advantages

- It is efficient and easier to implement and interpret the data.
- It can perform really well for labeled datasets.
- Polynomial regression can fit a wide range under the curvilinear.
- This methodology can be used cooperatively for better results and good insights.
- It can work well even with larger amounts of data sets.

Disadvantages

- These algorithms are sensitive to outliers and can produce a drastic effect on the end result.
- Data can be under fitted or over fitted depending upon the degree of the polynomial.
- If multicollinearity occurs, the errors and miscalculations may tend to increase.
- Data set must be supervised, that is labeled.

Conclusion

After going through all the details of both linear regression and polynomial regression we can conclude that, when these two methods are used we can recognize and predict the details of the necessary outcome required for our benefits.

Future work on these methods can be focused on developing the accuracy and efficacy of the results and help to obtain the results more efficiently.

References

- Poonam Chauhan, Ashok Kumar, Pooja Ja Mdagni. Regression Analysis of COVID-19 Spread in India and its Different States 2020.
- Kaya Uyanık, Gül den, Güler, Neşe. A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences* 2013;106:234-240. 10.1016/j.sbspro.2013.12.027
- Kumari, Khushbu, Yadav, Suniti. Linear regression analysis study. *Journal of the Practice of Cardiovascular Sciences* 2018;4:33. 10.4103/jpcs.jpcs_8_18. Covid-19 pandemic Analysis using
- Regression Raji P, Dee ba Lakshmi GR. medRxiv 2020.10.08.20208991;
- Prema Kumari, Aswini Kumar. A Study of Polynomial Regression towards Machine Learning- Ignited Minds *Journals* 2017, 2230-7540.
- Chatterjee C, Sarkar R. Multi- step polynomial regression method to model and forecast malaria incidence. *PLoS One* 2009;3:e4726.
- Wang B, Ding X, Wang F. Determination of polynomial degree in the regression of drug combinations, in *IEEE/CAA Journal of Automatica Sinica* 2017;4(1):41-47 Jan. doi: 10.1109/JAS.2017.7510319.
- Shalabh. A revisit to efficient forecasting in linear regression models, *Journal of Multivariate Analysis*, 2013;114:161-170. ISSN0047259X.