02

Regression

Regression

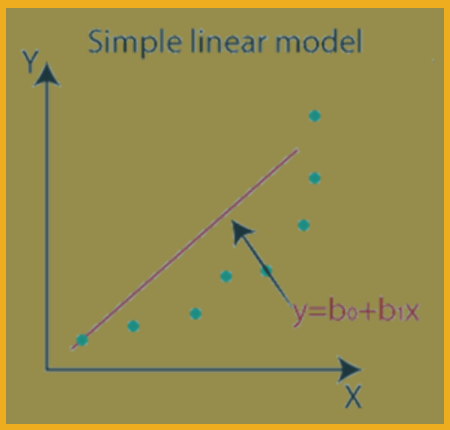
Regression analysis is a statistical tool for the investigation of relationships between variables. Usually, the investigator seeks to ascertain the causal effect of one variable upon another — the effect of a price increase upon demand, for example, or the effect of changes in the money supply upon the inflation rate.

Regression in machine learning consists of mathematical methods that allow data scientists to predict a continuous outcome (y) based on the value of one or more predictor variables (x).

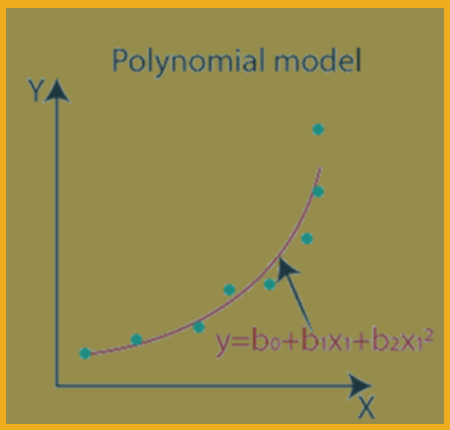
Keyword: Continuous Data

Regression

- Linear Regression



- Polynomial Regression



Linear Regression

Linear regression finds the linear relationship between the dependent variable and one or more independent variables using a best-fit straight line. The distance between each point and the line is minimized to achieve the best fit line. You might have heard the term “Logistic Regression”, but it is actually one of the classification algorithms.

Least Square Method

To help us get a better understanding let’s simulate the process of finding best fit line intuitively from:

https://phet.colorado.edu/sims/html/least-squares-regression/latest/least-squares-regression_en.html

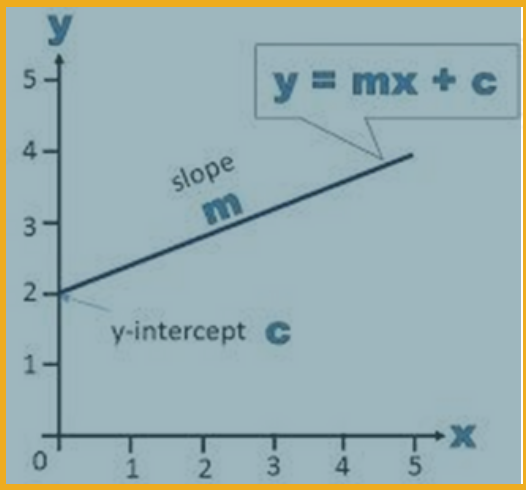
Error/loss is calculated by subtracting the actual value from the predicted one. Since the result from subtracting might be negative, we square the difference to make it a positive value. In regression we want to find the weights where the error is minimized.

Best Fit Line

Let’s see the Formula is:

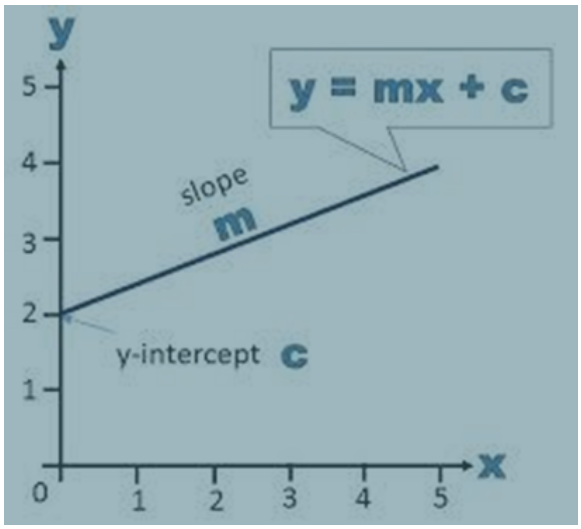
$y = mx+c$

y = dependent variable
m = the slope of the line
x = independent variable
c = y-intercept



Calculating the Weights

$$m = \frac{\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})}{\sum_{i=1}^{n} (x_i - \overline{X})^2}$$



Picture Source

- pexels.com