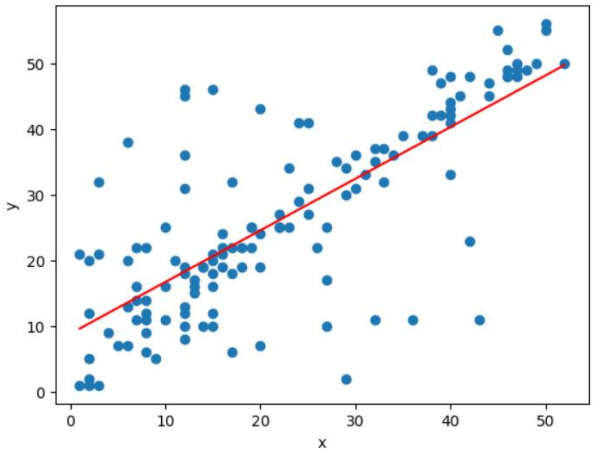
Linear regression

Introduction

Linear regression is a technique for predicting new data values from known data values. Which is a linear relationship between variables. If there is only one variable of interest, we call it simple linear regression. But if there is more than one variable of interest, we call it multiple linear regression. If we have a set of data and want to find a trendline by linear regression from the starting point. We need to understand some mathematical principles, algorithm, and specific values. This topic will include content on learning rate (α), linear equations, Gradient Descent algorithm, Mean Squared Error, and Calculus such as finding partial derivatives.



Simple linear regression equation

$$Y_i = \beta_0 + \beta_1 X_i$$

Y_i : Dependent variable

β_0 : Intercept

β_1 : Coefficient

X_i : Independent variable

Trendline in linear regression

A trendline is a line that best fits points in a data series. It can tell whether something increases or decreases at a constant rate.

Linear equations

- Standard form

$Ax + By = C$

- Slope-Intercept form

$y = mx + b$

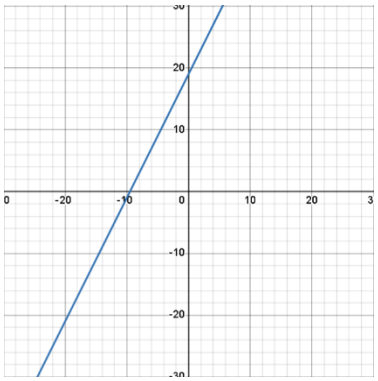
- Point-Slope form

$y\text{-}y_1 = m(x\text{-}x_1)$

m: Slope
b: y-intercept
(x_1, y_1): Point

Example

$$y = 2x + 19$$



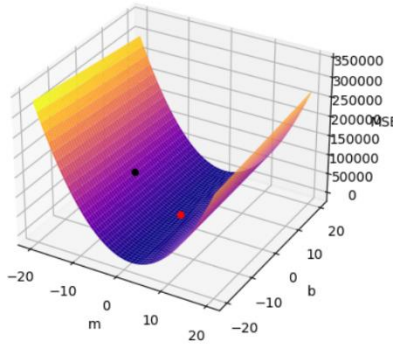
Mean Squared Error (MSE)

The Mean Squared Error (MSE) is a method of measuring the error in a statistical model. This can be found by subtracting the actual value and the predicted value. Then square it and divide it by the total number. If there is no error this value will be zero. And if there are large errors this value will also increase.

$$MSE = \frac{1}{n}\sum_{i=0}^{n}(y_i\text{-}\hat{y}_i)^2$$

Gradient Descent (Linear regression)

The gradient descent algorithm is a technique for optimizing various parameters and minimizing the cost function. It has two important parts: learning rate (α) and derivative.



This figure shows a three-dimensional plot of m, b, and mean square error. The one black dot is the random point to start the operation. The red dot is derived from an algorithm and is currently not the best value. The question is how do we find the best point? The answer is that we need to use mathematics to find the direction of the point's movement.

Learn about Gradient Descent Algorithm: https://github.com/ratikaewkam/GradientDescent

Learning rate

The learning rate is a parameter that controls how quickly the value changes. If the learning rate is large, the values will change quickly, and there may be large errors. If the learning rate is small, the values change slowly and there may be few errors.

Partial derivatives

Each partial derivative is important for improving the direction leading to a smaller cost function. (You can learn about partial derivatives from the previous lesson.)

From MSE equation

$$\text{MSE} = \frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2$$

$$m = m - \alpha\left(\frac{-2}{n}\sum_{i=0}^{n}(y_i-(mx_i + b))x_i\right)$$

$$b = b - \alpha\left(\frac{-2}{n}\sum_{i=0}^{n}(y_i-(mx_i + b))\right)$$

By: Rati Kaewkam