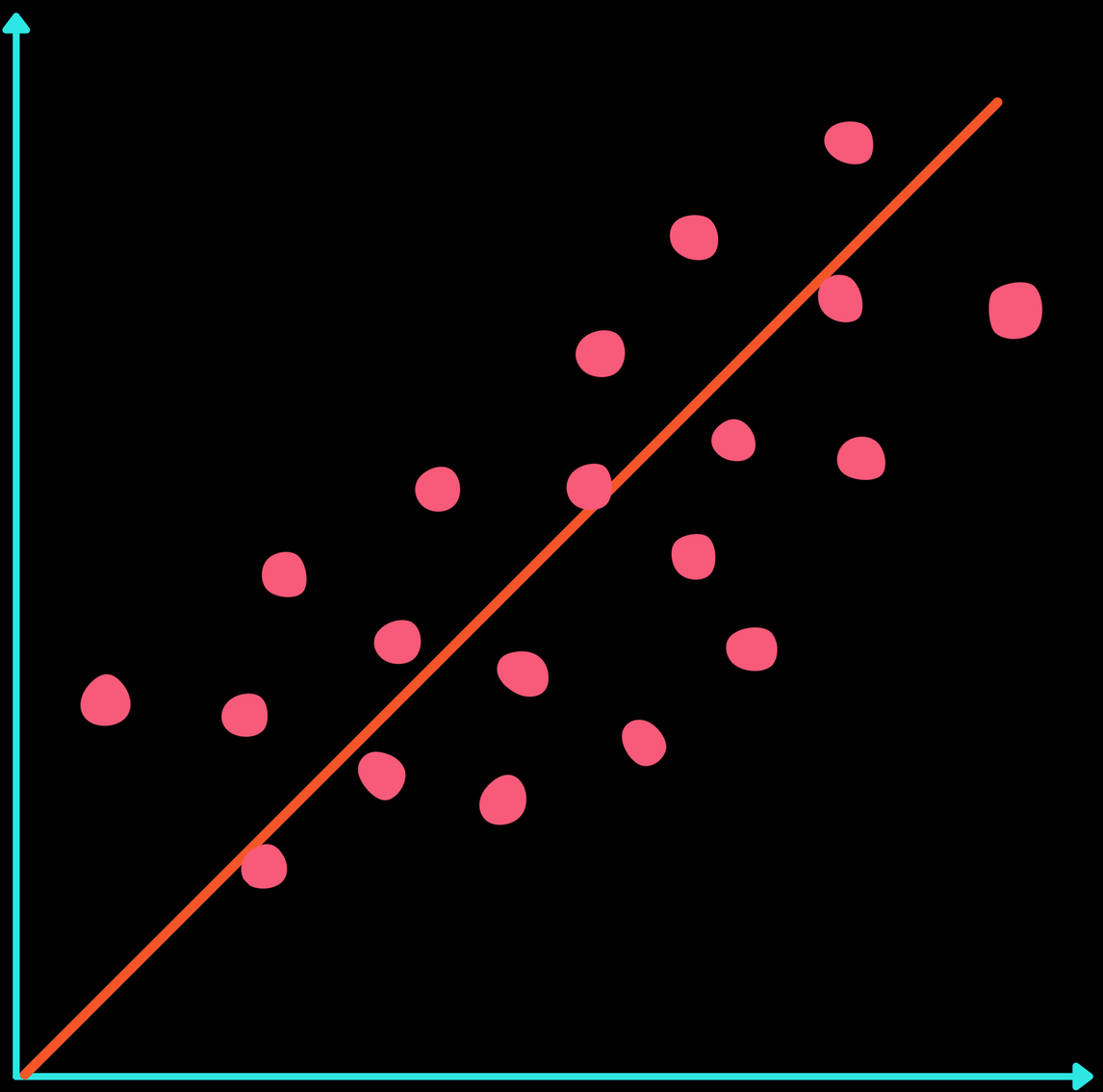
01

SIMPLE INTRO TO LINEAR REGRESSION



PART - 1

learn.machinelearning

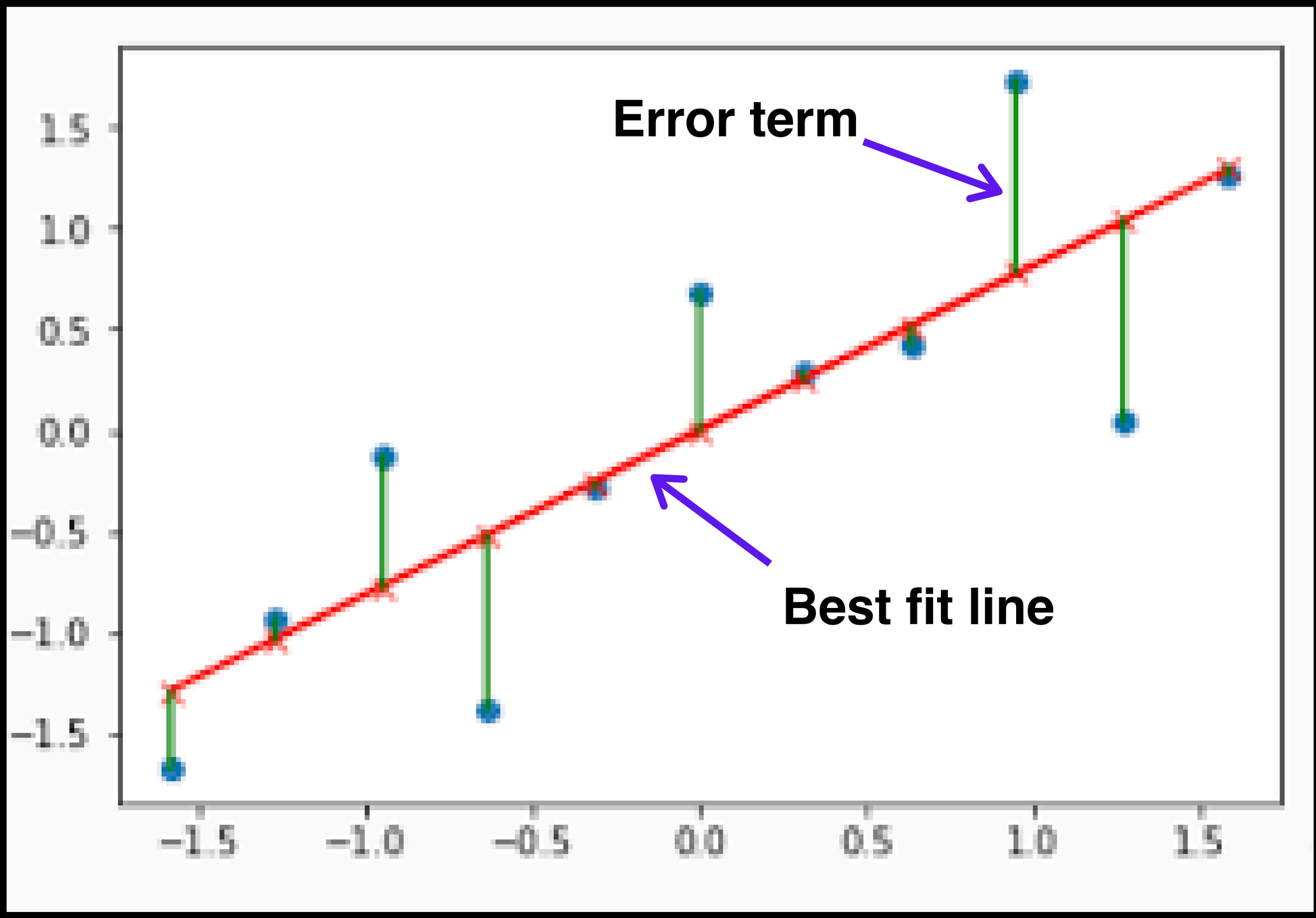
What is Linear Regression?

This algorithm is used to find the relationship between 2 continuous variables (one independent variable and one dependent variable)

Linear regression is a linear model, e.g. a model that assumes a linear relationship between the input variables (x) and the single output variable (y). More specifically, y can be calculated from a linear combination of the input variables (x).

The equation is Y= W1*X+b (for one input variable) where,

- Y = Dependent variable
- W1 = Gradient/slope/Weight
- X = Independent variable
- b = Bias



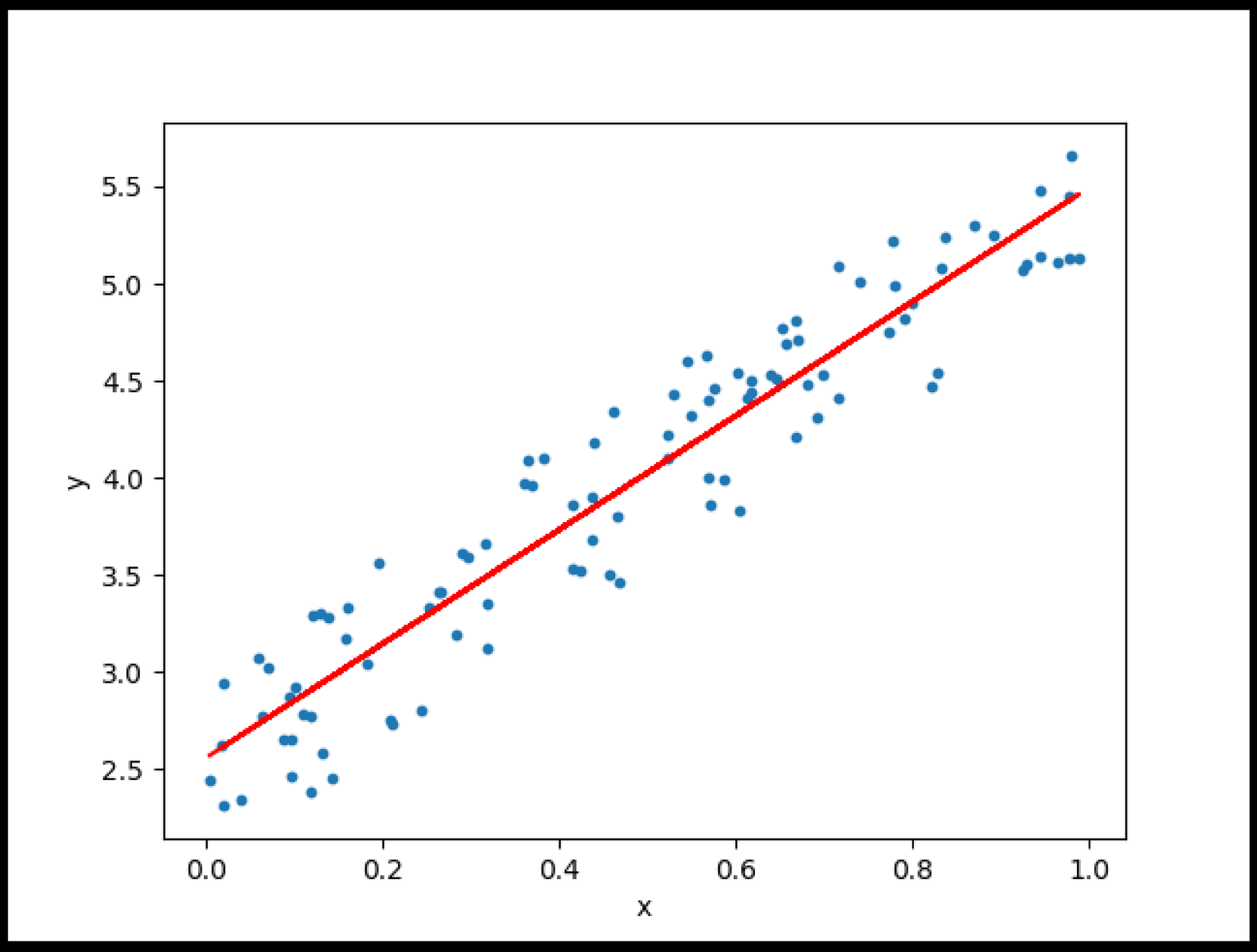
The equation Y= W1*X+b is the same as that of a straight line (Y=MX+c)

The core idea is to obtain a line that best fits the data to which we get low prediction error. Error is the distance between the point to the best fit line.

The values m and b must be chosen in a way that minimizes the error rate.

If the sum of squared errors ($\Sigma(Y - \overline{Y})^2$) is taken as a metric to evaluate the model, the goal to obtain a line that gives the low error value.

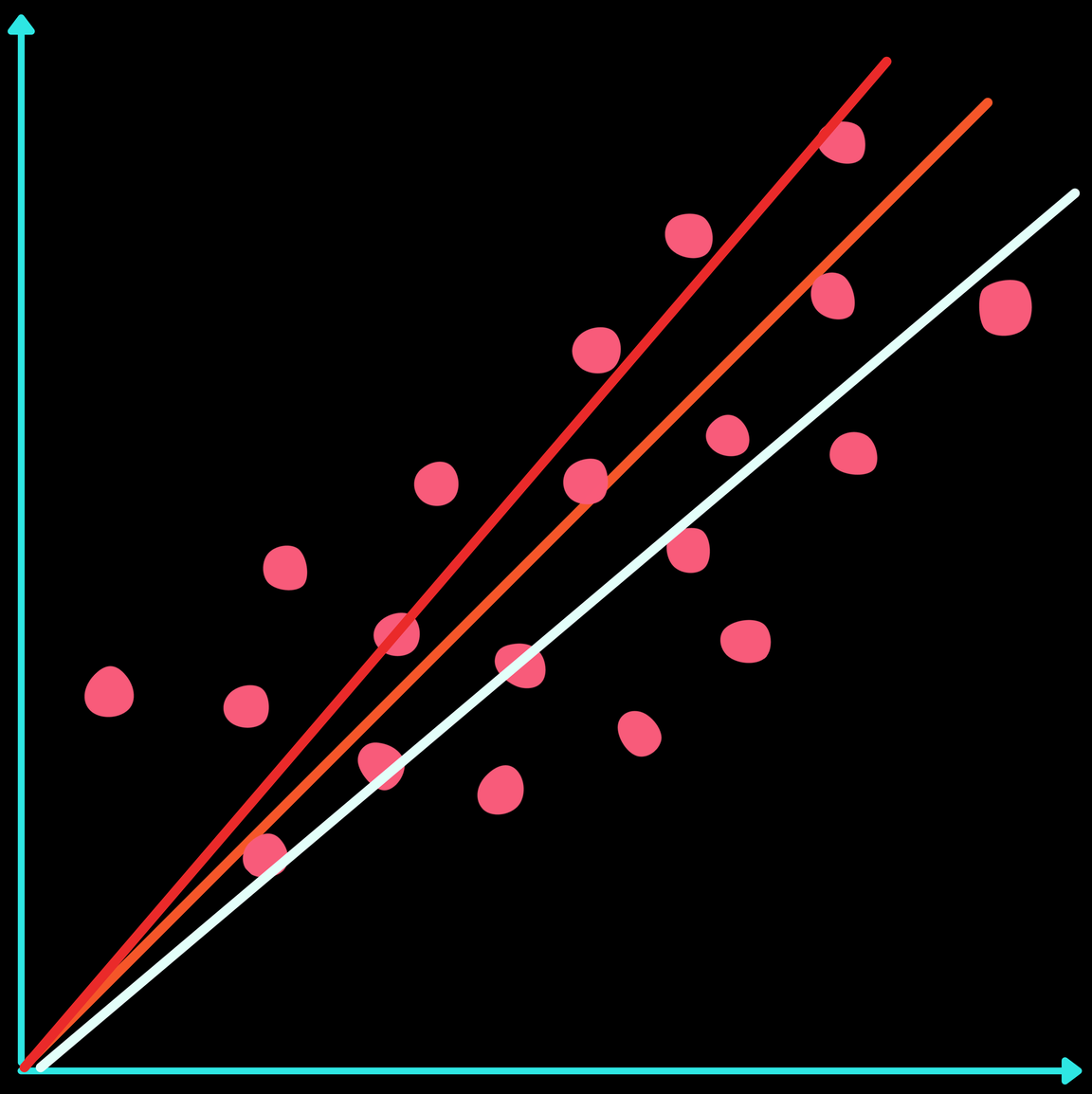
We can train this algorithm using multiple methods, we can use the ordinary least squares method to calculate m and b or we can also use gradient descent to do that.



Math behind LR

Goal is to find that red straight line (which is best fit) to the data. First lets talk about how to draw a linear line in the graph, In math we have an equation which is called linear equation $y = mX+b$

- m = slope ,
- b = Y-intercept



Math behind LR

so how can we calculate m and b values? and how do we know exact m and b values for the best fit line??

We take some random values of m and b and by taking all the X values we will find the Y values and we drew line with those Y values. We do this until we find the best fit line with low error value.

Math behind LR

How do we change m and b values for the best fit line?? • Either we can use an algorithm called Gradient Descent Or we can use direct formulas from statistics. Let's first use the statistics formulas then we can go to GD.

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

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

Math behind LR

Here $\bar{x}$ is the mean of all the values in the input X and $\bar{y}$ is the mean of all the values in the desired output Y. This is the Least Squares method.

And for prediction we just use the formula $Y = mX + b$ and use the m and b values which we got from the above formula.

Turn on post notifications

If you have any questions, please do let us know in the comments.

Like

Save