Linear Regression

Equation:

slope of line

input feature

$$\hat{y} = \omega x + b \quad \text{(best fit line)}$$

predicted value

intercept

Example:

x (Hour)	y (Score)
1	2
2	3
3	5

Calculation Means:

$$\bar{x} = \frac{1+2+3}{3} = 2$$

$$\bar{y} = \frac{2+3+5}{3} = 3.33$$

Calculate ω (Slop)

$$\omega = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	2	-1	-1.33	1.33	1
2	3	0	-0.33	0	0
3	5	1	1.67	1.67	1
				sum = 3.0	sum = 2

$$\omega = \frac{3.0}{2} = 1.5$$

Now intercept (b)

$$b = \bar{y} - \omega\bar{x} = 3.33 - (1.5)(2)$$
$$= 3.33 - 3 = 0.33$$

Final Regression Line:

$$\hat{y} = 1.5x + 0.33$$

Now. if the student studies 4 hours:

$$\hat{y} = 1.5 \cdot 4 + 0.33$$
$$= 6.33$$

Linear Regression using LSM

Least Squares Methods is a mathematical technique used to find the best fitting line (or curve) through a set of data points.

Example:

x	y
1	2
2	3
4	5

we know,

$$y = wx + b$$

Compute the required sums

x	y	$x \cdot y$	x^2
1	2	2	1
2	3	6	4
4	5	20	16

$\Sigma x = 7$

$\Sigma y = 10$

$\Sigma xy = 28$

$\Sigma x^2 = 21$

$n = 3$

Now.

$$W = \frac{n \Sigma (xy) - \Sigma x \Sigma y}{n \Sigma (x^2) - (\Sigma x)^2}$$

$$= \frac{3 \cdot 28 - 7 \cdot 10}{3 \cdot 21 - 7^2}$$

$$= \frac{84 - 70}{63 - 49} = \frac{14}{14} = 1$$

$$b = \frac{\Sigma y - w \cdot \Sigma x}{n}$$

$$= \frac{10 - 1.7}{3} = 3/3 = 1$$

Now,

$$\hat{y} = 1.x + 1$$

Cost Function (MSE)

The mean squared Error (MSE) is the most commonly used cost function in linear Regression. It measures the average squared difference between the actual values and predicted value.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\underbrace{y_i}_{\text{Actual value}} - \underbrace{\hat{y}_i}_{\text{predicted value}})^2$$

Lets say,

$$\hat{y} = 1.5x + 0.33$$

Prediction:

$x = 1$; $\hat{y} = 1.83$

$x = 2$; $\hat{y} = 3.33$

$x = 3$; $\hat{y} = 4.83$

Actual y : [2, 3, 5]

$$\bar{y} = \frac{2+3+5}{3} = 3.33$$

$$MSE = \frac{(2-1.83)^2 + (3-3.33)^2 + (5-4.83)^2}{3}$$

$$= \frac{0.1667}{3} = 0.0556$$

R^2

R^2 tells you how well your Regression line fits the data.

$R^2 = 1$ perfect fit

$R^2 = 0$ no better than guessing the average

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

$$= 1 - \frac{\text{error between actual value and predicted}}{\text{error between actual values and their mean}}$$

from MSE

$$= 1 - \frac{0.1667}{(2-3.33)^2 + (3-3.33)^2 + (5-3.33)^2}$$

$$= 1 - \frac{0.1667}{4.6667} = 0.9643 = 96.43\%$$

Very Good fit