29-01-2022

29 January 2022 14:01

i) $\sum_{i=1}^{n}(\bar{x}^2 - x_i\bar{x}) = 0$

ii) $\sum_{i=1}^{n}(\bar{x}\bar{y} - y_i\bar{x}) = 0$

$E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ $y = bx + a$

$= \sum_{i=1}^{n}(y_i - bx_i - a)^2 = 0$

$\frac{\partial E}{\partial a} = 0$

$a = \frac{\sum_{i=1}^{n} y_i}{n} - b\frac{\sum x_i}{n}$

$= \bar{y} - b\bar{x}$

$\frac{\partial E}{\partial b} = 0$

$\Rightarrow 2 \cdot \sum^{n}(y_i - a - bx_i)(-x_i) = 0$

$$\Rightarrow 2 \cdot \sum_{i=1}^{n} (y_i - a - b x_i)(-x_i) = 0$$

$$\Rightarrow -2 \sum_{i=1}^{n} (x_i y_i - a x_i - b x_i^2) = 0$$

Put $a = \bar{y} - b\bar{x}$

$$\Rightarrow \sum_{i=1}^{n} \left(x_i y_i - (\bar{y} - b\bar{x}) x_i - b x_i^2\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (x_i y_i - \bar{y} x_i + b\bar{x} x_i - b x_i^2) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (x_i y_i - \bar{y} x_i) - b \sum_{i=1}^{n} (x_i^2 - \bar{x} x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (x_i y_i - \bar{y} x_i) = b \sum_{i=1}^{n} (x_i^2 - \bar{x} x_i)$$

$$\Rightarrow b = \frac{\sum_{i=1}^{n} (x_i y_i - \bar{y} x_i) + 0}{\sum_{i=1}^{n} (x_i^2 - \bar{x} x_i) + 0}$$

$\Rightarrow$ from (i) & (ii) substitute 0

$\Rightarrow$ given (1) & ...

$$\Rightarrow b = \frac{\sum_{i=1}^{n}(x_i y_i - \bar{y} x_i) + \sum_{i=1}^{n}(\bar{x}\bar{y} - y_i \bar{x})}{\sum_{i=1}^{n}(x_i^2 - \bar{x} x_i) + \sum_{i=1}^{n}(\bar{x}^2 - x_i \bar{x})}$$

$$\Rightarrow b = \frac{\sum_{i=1}^{n} y_i (x_i - \bar{x}) - \sum_{i=1}^{n} \bar{y}(x_i - \bar{x})}{\sum_{i=1}^{n} x_i (x_i - \bar{x}) - \sum_{i=1}^{n} \bar{x}(x_i - \bar{x})}$$

$$\Rightarrow b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})}$$

=

find b

then a

$y = bx + a$: best fit line for given training set.

$y = bx$ training set.

$$b = \frac{Cov(x, y)}{Var(X)}$$

Eg:

X	Y
1	4
2	12
3	28
4	52
5	80

$$b = \frac{y_2 - y_1}{x_2 - x_1} = 19$$

$a = -15$ $\quad y = 19x - 15$ ①

$\overline{X} = 3$ (mean)

$\overline{Y} = 35.2$ (mean)

Now, New formula

$$a = \overline{y} - b\overline{x}$$

$$b = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

X	Y	$x_i - \overline{x}$	$y_i - \overline{y}$	$(x_i - \overline{x})(y_i - \overline{y})$	$(x_i - \overline{x})^2$
1	4	-2	-31.2	62.4	4
2	12	-1	-23.2	23.2	1
3	28	0	-7.2	0	0
4	52	1	16.8	16.8	1

3	28	0	-7.2	0	
4	52	1	16.8	16.8	1
5	80	2	44.8	89.6	4
				192	10

$\therefore \quad b = 19.2$

$a = -22.4$

$$y = bx + a$$
$$= 19.2x - 22.4 \quad — (2)$$

Compare (1) & (2)

$y-\hat{y}$ of line error	$\hat{y}$ Line 1
0	$y = 4$
-11	$y = 23$
-14	$y = 42$
-9	$y = 61$
0	$y = 80$

Squared error

0
121
196
81
0

Line 2 $\hat{y}$	$y-\hat{y}$	sq. error
$y = -3.2$	7.2	51.84
$y = 16$	-4	16
$y = 35.2$	-7.2	51.84
$y = 54.4$	-2.4	5.76
$y = 73.6$	6.4	40.96
		$\Sigma = 166.4$

Compare Error

$\Sigma = 398$

Multiple Linear Regression

— Derivation