A Best-fit Curve

Sample Data: $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$

Assume a linear model, $\hat{y}_i = mx_i + b$ (unknown parameters)

The error (residual) $= y_i - \hat{y}_i$

$$= y_i - (mx_i + b)$$

$$= y_i - mx_i - b$$

sum of all squared error:

$$S(m,b) = \sum_{i=1}^{n} (y_i - mx_i - b)^2$$

we want to minimize this error →

Goal: find m & b for which $S(m,b)$ minimum

so, we need partial derivatives.

$$\frac{\partial S}{\partial b} = \sum_{i=1}^{n} -2(y_i - mx_i - b)$$

for minimum, $\frac{\partial S}{\partial b} = 0$

$$\Rightarrow \sum_{i=1}^{n} y_i - mx_i - b = 0 \quad \text{--- (i)}$$

similarly:

$$\frac{\partial S}{\partial m} = \sum_{i=1}^{n} 2(y_i - mx_i - b)(-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i(y_i - mx_i - b) = 0 \quad \text{— (ii)}$$

solve the system:

(i)
$$\sum_{i=1}^{n} y_i - mx_i - b = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i - m\sum_{i=1}^{n} x_i - nb = 0$$

$$\Rightarrow b = \cancel{\frac{1}{n}}\left(\frac{\sum_{i=1}^{n} y_i}{n} - m\frac{\sum_{i=1}^{n} x_i}{n}\right)$$

$$\Rightarrow \boxed{b = \bar{y} - m\bar{x}}$$

average as $\bar{y} = \frac{\sum y}{n}$

(ii)
$$\sum_{i=1}^{n} x_i(y_i - mx_i - b) = 0$$

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

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

$$\Rightarrow \sum x_i(y_i - \bar{y}) - m\sum x_i(x_i - \bar{x}) = 0$$

Proof of $\sum x_i - \bar{x} = 0$

$$\bar{x} = \frac{1}{n}\sum x_i$$

$$\Rightarrow \sum x_i = n\bar{x}$$

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

$$\Rightarrow n\bar{x} - n\bar{x}$$

$$= 0$$

we have:

$$\sum x_i(x_i - \bar{x}) = \sum (x_i - \bar{x} + \bar{x})(x_i - \bar{x})$$

$$= \sum (x_i - \bar{x})^2 + \sum \bar{x}(x_i - \bar{x})$$

$$= \sum (x_i - \bar{x})^2 + \bar{x}\cancel{\sum (x_i - \bar{x})}^{\to 0}$$

$$= \sum (x_i - \bar{x})^2$$

similarly,

$$\sum x_i(y_i - \bar{y}) = \sum (x_i - \bar{x} + \bar{x})(y_i - \bar{y})$$

$$= \sum (x_i - \bar{x})(y_i - \bar{y}) + \bar{x}\sum (y_i - \bar{y})^{\to 0}$$

$$= \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

$\therefore$ the eqn become $\longrightarrow$

$$\sum_i (x_i - \bar{x})(y_i - \bar{y}) - m \sum (x_i - \bar{x})^2 = 0$$

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

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