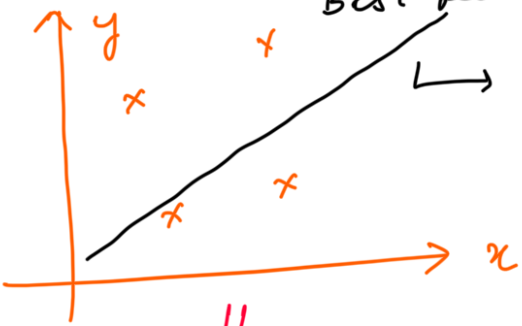
Linear Regression using OLS {ordinary Least square}

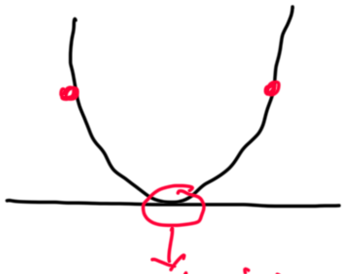


Best fit line

↳ optimization

⇓

loss function



Global minima

⇓

OLS — can I apply some formula and calculate β_0 (Intercept) and β_1 (coefficient)

Aim of OLS is to reduce the errors between the true points and predicted points

Ordinary Least Square

Error ⇑ [MSE]

$$S(\beta_0, \beta_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

⇓

find β_0 and β_1

$$\frac{\partial S}{\partial \beta_0}(\beta_0, \beta_1) \frac{2}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_1)^{2-1}$$

⇓

with respect to β_0

$$= (0 - 1 - 0)$$

$$= -\frac{2}{}\sum^{n}(y_i - \beta_0 - \beta_1 x_1) = 0 \quad \text{—(1)}$$

$$n \qquad i = 1$$

$$\frac{\partial s}{\partial \beta_1}(\beta_0, \beta_1) = \frac{2}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_1)$$

$$\Downarrow$$

with respect to β_1

$$(-x_1)$$

$$= \frac{2}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_1)(-x) = 0$$

$$= \frac{-2}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_1)(x_1) = 0 \longrightarrow ②$$

$Eq \rightarrow ①$

$$-\frac{2}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_1) = 0$$

$$\Downarrow$$

$$-\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_1) = 0$$

$$-\sum_{i=1}^{n} y_i + n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i = 0$$

$$n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\beta_0 = \sum^{n} y_i - \beta_1 \sum^{n} x_i$$

$$\beta_0 = \frac{\sum_{i=1} \ldots}{n} \quad \frac{\ldots_{i=1}}{n}$$

$$\boxed{\beta_0 = \bar{y} - \beta_1 \bar{x}} \Rightarrow \text{Intercept}$$

Eq — 2

$$-\frac{2}{n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 \right) \left(x_i \right) = 0$$

$$\Downarrow$$

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 \right) \left(x_i \right) = 0$$

$$\Downarrow$$

$$\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} (x_i)^2 = 0$$

$$\Downarrow$$

$$\sum_{i=1}^{n} \left(x_i y_i - \beta_0 x_i - \beta_1 x_i^2 \right) = 0$$

$$\Downarrow$$

Replace $\beta_0 = \bar{y} - \beta_1 \bar{x}$

$$\sum_{i=1}^{n} \left(x_i y_i - \left(\bar{y} - \beta_1 \bar{x} \right) x_i - \beta_1 x_i^2 \right) = 0$$

$$\Downarrow$$

$$\sum_{i=1}^{n} \left(x_i y_i - x_i \bar{y} + \beta_1 \bar{x} x_i - \beta_1 x_i^2 \right) = 0$$

$$\Downarrow$$

$$\sum_{i=1}^{n} \left(y_i - \bar{y} + \beta_1 \bar{x} - \beta_1 x_i^2 \right) x_i = 0$$

$$\Downarrow$$

$$\sum_{i=1}^{n} \left((y_i - \bar{y}) + \beta_1 (\bar{x} - x_i) \right) = 0$$

$$\Downarrow$$

$$\sum_{i=1}^{n} (y_1 - \bar{y}) + \sum_{i=1}^{n} \beta_1 (\bar{x} - x_i) = 0$$

$$\Downarrow$$

$$\sum_{i=1}^{n} \beta_1 (\bar{x} - x_i) = -\sum_{i=1}^{n} (y_i - \bar{y})$$

$$\Downarrow$$

$$\beta_1 = \frac{-\sum_{i=1}^{n} (y_i - \bar{y})}{\sum_{i=1}^{n} (\bar{x} - x_i)}$$

$$\Downarrow$$

$$\beta_1 = \frac{+\sum_{i=1}^{n} (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})}$$

$$+\sum_{i=1} (x_i - \bar{x})$$

coefficient $\Longrightarrow$ $\boxed{\beta_1 = \dfrac{\sum_{i=1}^{n} (y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})}}$

OLS ≃ Linear Regression (sk learn)