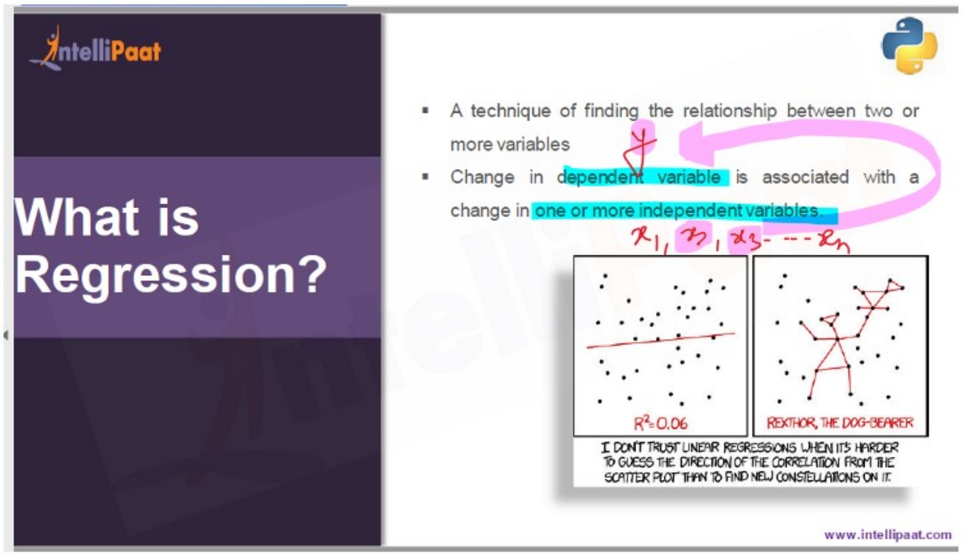
Linear Regression

15 December 2023 07:52

model is linear → linear relationship between Y & x_i's

↳ continuous / ~~categorical~~

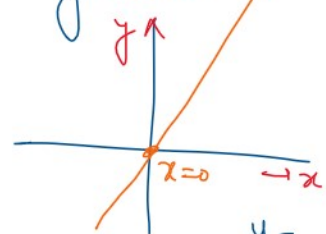
Supervised Learning Algorithm



a function in 'x'

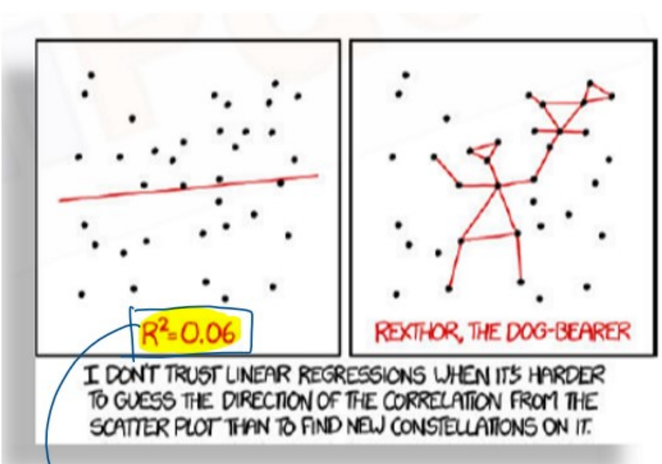
$y = 2x$

$y = mx.$



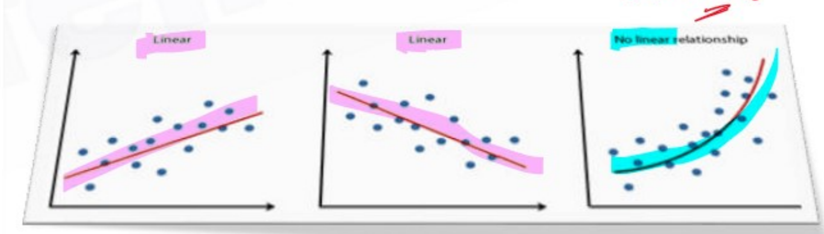
$y = 2x$

y	x
0	0
-2	-1
2	+1



6%

Non-linear

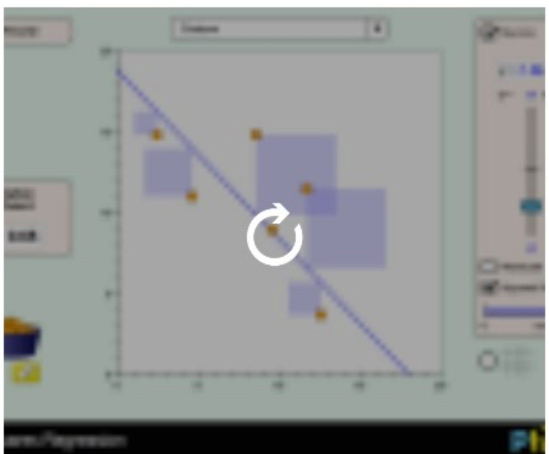


Function

1. Intuitive Approach
↓
2. Mathematical Approach
↓
3. Python Hands-on

Mathematical Approach →
- ① Calculus → [Function ↓ Limits → Continuity → Differentiability; Maxima & Minima (Global Optimization)]
- ② Linear Algebra
- ③ (Probability and Random Variables + Probability distribution — Normal distribution)
- ④ statistics

[Least-Squares Regression](#)



mathematical equation

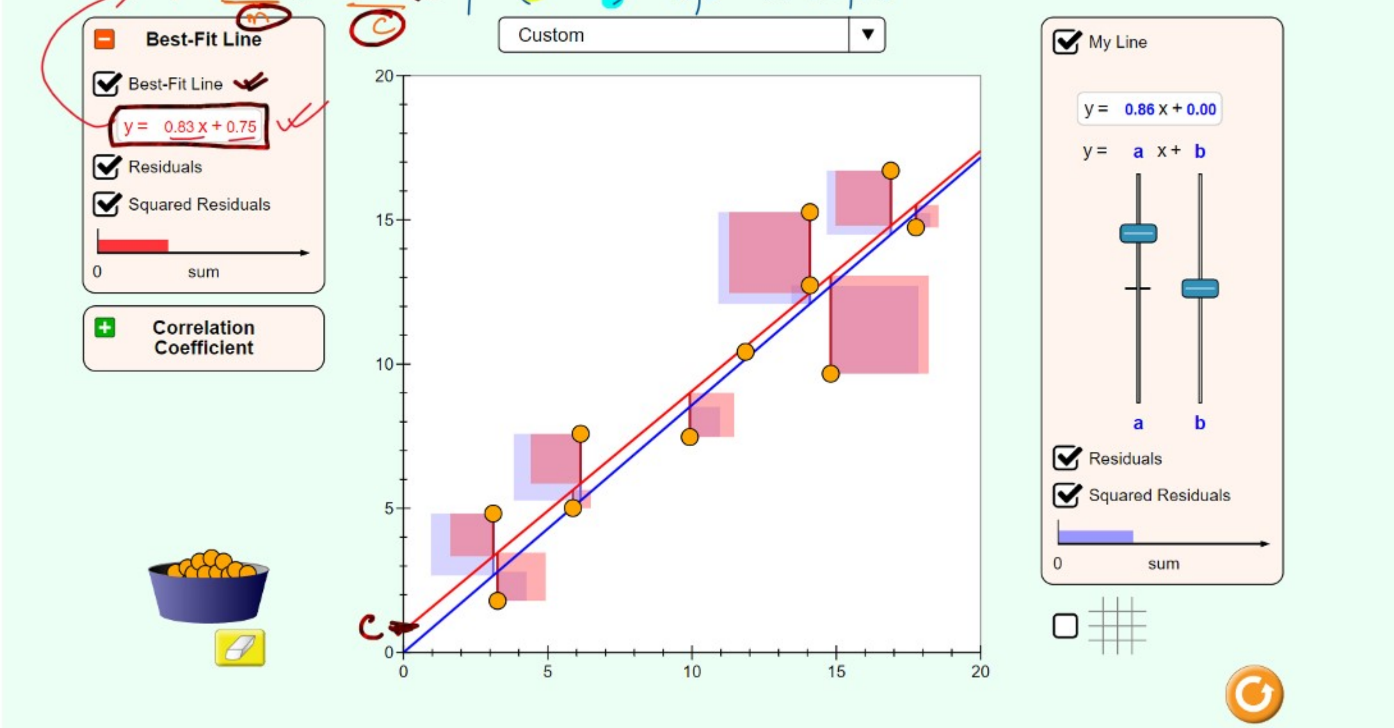
$$Y = 0.83X + 0.75$$

$$Y = \hat{\beta}_1 x + \hat{\beta}_0$$

$\hat{\beta}_1$ → slope; $\hat{\beta}_0$ → intercept

$Y = 0.83X + 0.75 \Leftrightarrow Y = (mx + c)$: slope-intercept line.

m, c

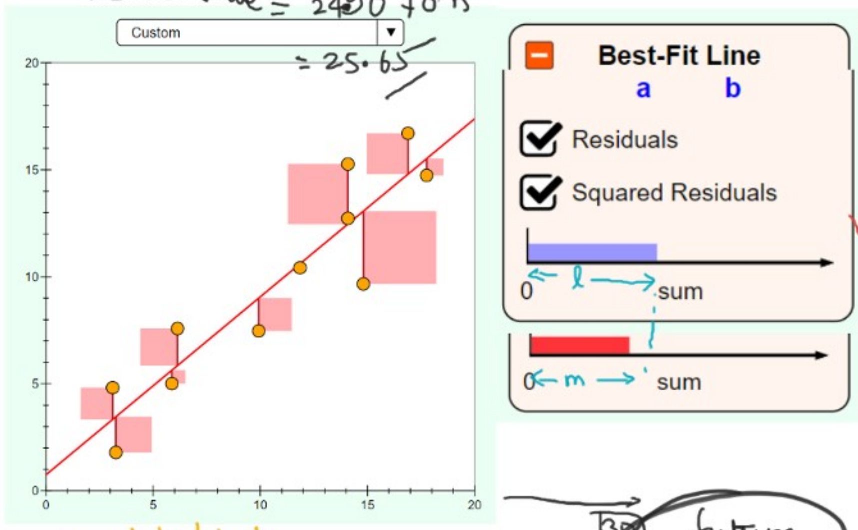


$y = 0.83x + 0.75$

$\hat{y}$ Predicted value $= 0.83 \times 30 + 0.75$ at $x = 30$

$= 24.90 + 0.75$

$= 25.65$



l: intuitive best fit line

m: LR model best fit line

a) $l > m$

b) $l < m$

c) $l = m$

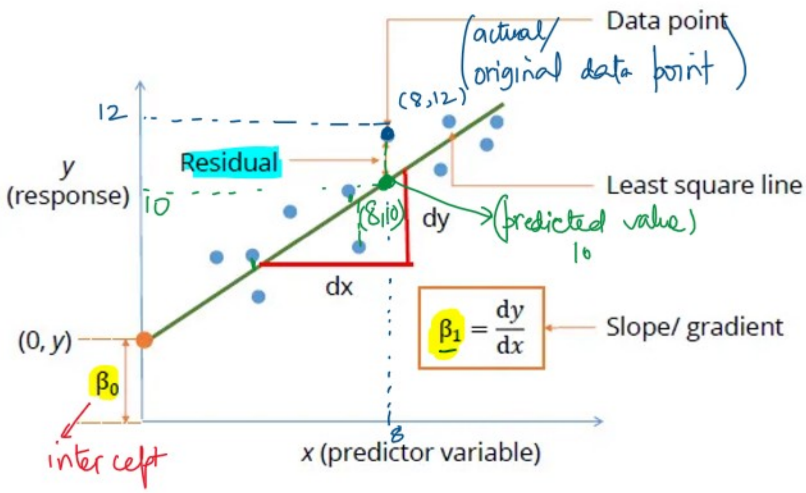
d) Not sure

● : data points

— : Linear Regression Line

● data point

— (Linear Regression / Best fit / least square line)



Residual

$y = c + mx + u$

$$\boxed{y = \beta_0 + \beta_1 x + u}$$

slope/ gradient

Actual value

Predicted value

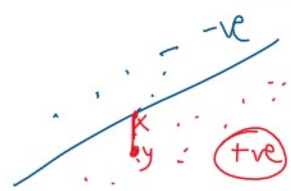
$\hat{\beta_0}$, $\hat{\beta_1}$: derive it as well.

$y = 2 + 0.5x$

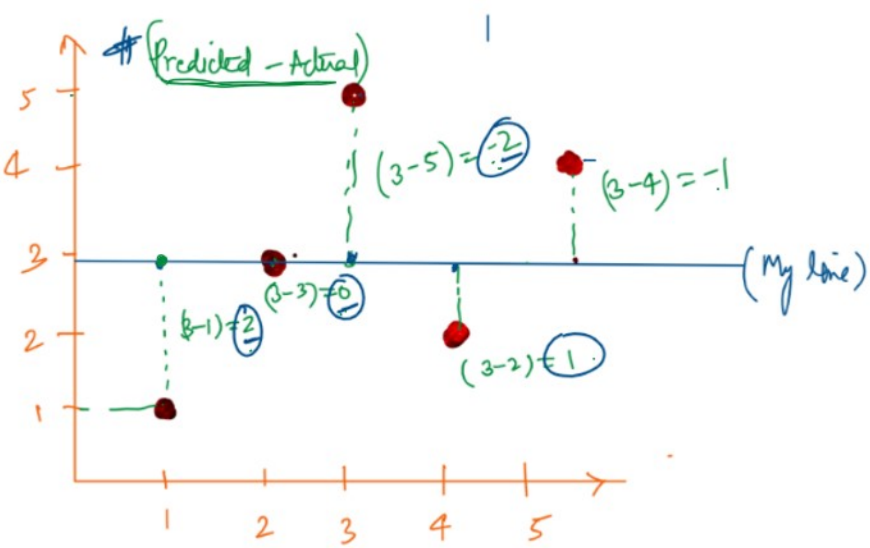
(Predicted − actual)

(10 − 12)

= −2



+ve residues cancel −ve residues

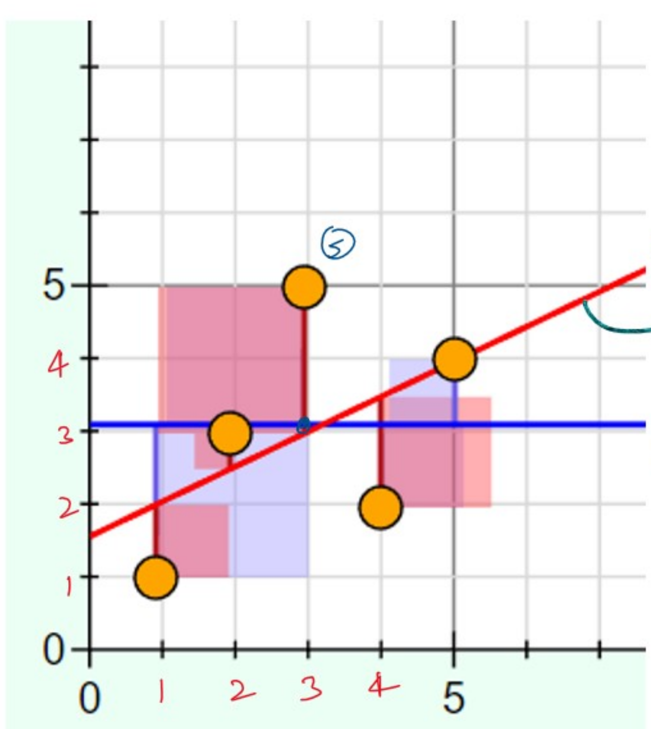


① Sum of residues = $2 + 0 - 2 + 1 - 1 = 0$

(zero sum of residues)

⇓

(zero error in the model)



Best fit line → (7.8) ordinary least square

↓

(minimum)

(my line) (10)

Observation: Residues nullify each other and may not be truly representing the residues.

Observation: Residues nullify each other and may not be truly representing the residues.

OLS Method: ordinary least squares

Sum of squared residues

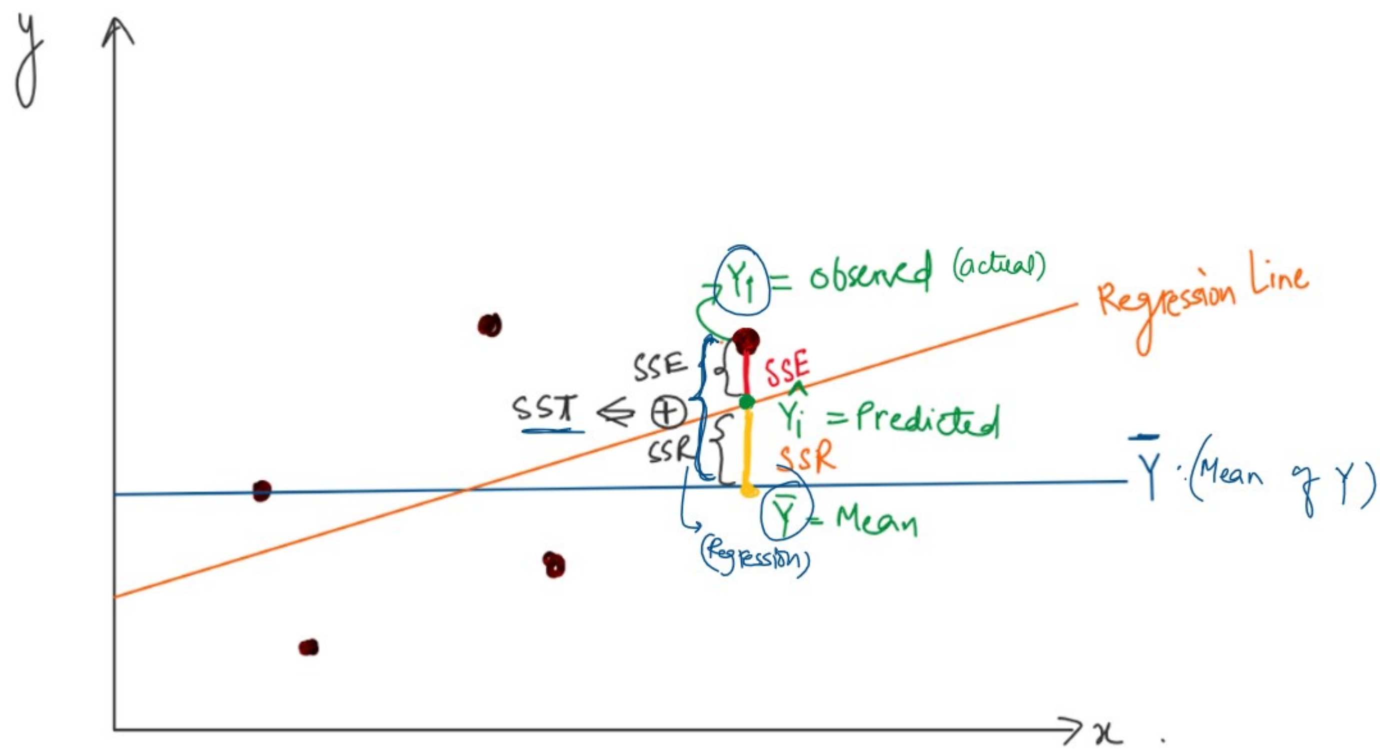
Sum of residues $= \cancel{2} + 0 \cancel{-2} \cancel{+1} \cancel{-1} = 0$.

sum of squared residues $= 2^2 + 0^2 + (-2)^2 + 1^2 + (-1)^2$

$= 4 + 0 + 4 + 1 + 1$

$= 10$.

SSE, SSR and SST



sum of squares error (residues) (SSE)

- it is the sum of the squared difference between the observed (actual) value (Y_i) and predicted value $(\hat{Y}_i)$

- it shows the unexplained variance by regression.

$$\boxed{SSE = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2}$$

sum of squares regression (SSR)

- it is the sum of the squared differences between predicted value $(\hat{Y}_i)$ and the mean of the dependent variable $(\bar{Y})$

- it is a measure that describes how well

- It is a measure that describes how well our line fits the data.

$$\boxed{SSR_{\text{(regression)}} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}$$

$\bar{Y}$ → Mean

sum of squares total (SST)

- it is the squared differences between observed dependent variable and its mean.

$$\boxed{SST / TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

- it is a measure of total variability of the dataset.

Mathematically,

SST = SSR + SSE

$$\boxed{\Rightarrow \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2}$$

Total variability of the dataset (SST)

variability explained by the regression line (SSR)

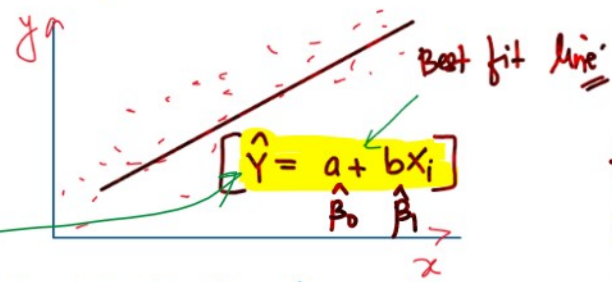
unexplained variability (SSE)

100% = 75% + 25%

75% → Accuracy

Derivation of linear regression equation

Given a set of 'n' points (x_i, y_i) on a scatter plot.



$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$
$$\hat{Y} = a + bx_i$$

find the best fit line; $\hat{Y} = a + bx_i$

such that the sum of squared errors in Y : $\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$ is minimized

actual (Observed) (Predicted)

$$Q = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} \left[Y_i - (a + bx_i)\right]^2$$

$$Q = \sum_{i=1}^{n} (Y_i - a - bx_i)^{\boxed{2}}$$

Using maxima/minima concept, let us partially differentiate with 'a' and 'b' respectively.

$$Z = x^2 y$$

$$\frac{\partial z}{\partial x} = 2xy \quad \Big| \quad \frac{\partial z}{\partial y} = x^2$$

– ('y' constant) ('x' constant)

$$\boxed{\frac{\partial Q}{\partial a} = 0}$$

$$\frac{\partial Q}{\partial a} = \sum_{i=1}^{n} 2 \left(\boxed{Y_i - a - bx_i}\right)^1 \times (0 - 1 - 0)$$

(chain rule)

$$\frac{d}{\partial a}(Y_i - a - bx_i) = (0 - 1 - 0)$$

const.

$$\frac{\partial Q}{\partial a} = -2 \sum_{i=1}^{n} (Y_i - a - bx_i)$$

Differentiation

$$y = x^2$$
$$\frac{dy}{dx} = 2x$$

$$y = x^3$$
$$\frac{dy}{dx} = 3x^2$$

$$y = c$$
$$\frac{dy}{dx} = 0$$

$$y = x^n$$
$$\frac{dy}{dx} = nx^{n-1}$$

[https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/partial-derivatives/v/partial-derivatives-introduction] https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/partial-derivatives/v/partial-derivatives-introduction

$$\frac{dy}{dx} = nx^{n-1}$$

$$\frac{\partial Q}{\partial a} = -2\left[\sum_{i=1}^{n} Y_i - \underbrace{\sum_{i=1}^{n} a}_{(na)} - b\sum_{i=1}^{n} X_i\right] = 0$$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\boxed{\Rightarrow \sum_{i=1}^{n} X_i = n\bar{X}}$$

$$\frac{\partial Q}{\partial a} = -2n\bar{Y} + 2na + 2bn\bar{X} = 0$$

Dividing by 2n

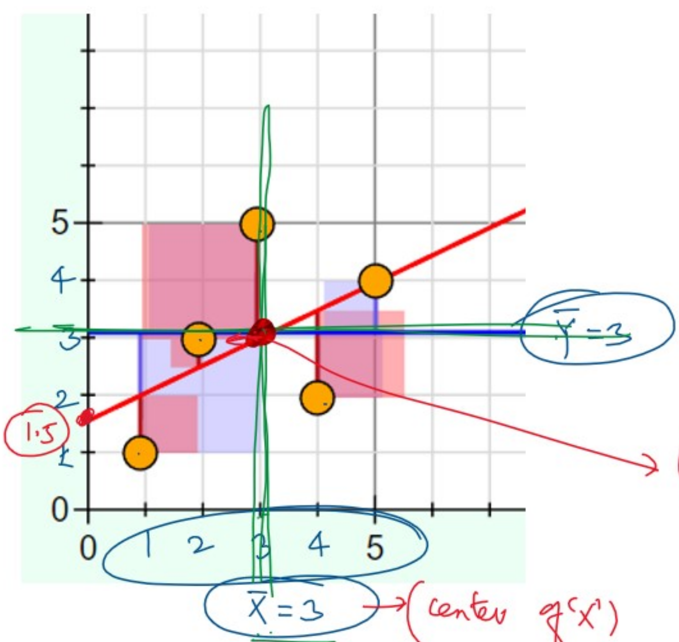
$$\Rightarrow -\bar{Y} + a + b\bar{X} = 0$$

$$\begin{bmatrix} Y \\ \vdots \\ \vdots \end{bmatrix} = a + b\begin{bmatrix} \vdots \\ \vdots \end{bmatrix}$$

$$\boxed{\Rightarrow a = \bar{Y} - b\bar{X}}$$

$a \rightarrow \hat{\beta}_0$, $b \rightarrow \hat{\beta}_1 = ??$

'a' is constant (Y- intercept) is such that the line must go through the mean of 'X' and 'Y'



$\bar{Y} = 3$ (center of Y)

(this data point- center of data cloud.)

$\bar{X} = 3 \rightarrow$ (center of 'x')

$\boxed{a = 1.5}$

[https://www.youtube.com/watch?v=4b4MUYve_U8 : Andrew Ng - DO NOT WATCH NOW -- try it after our course is over]

Second condition for minimizing Q'is:

$$\boxed{\frac{\partial Q}{\partial b} = 0}$$

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

$$\frac{\partial Q}{\partial b} = \sum_{i=1}^{n} 2(Y_i - a - bx_i) \times (0 - 0 - x_i)$$

$$\frac{\partial Q}{\partial b} = -2 \sum_{i=1}^{n} (Y_i - a - bx_i)(x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} [Y_i x_i - a x_i - b x_i^2] = 0$$

Let us substitute 'a' from $a = \bar{Y} - b\bar{x}$

$$\Rightarrow \sum_{i=1}^{n} [Y_i x_i - (\bar{Y} - b\bar{x}) x_i - b x_i^2] = 0$$

$$\Rightarrow \sum_{i=1}^{n} [\underbrace{Y_i x_i - \bar{Y} x_i} + \underbrace{b\bar{x} x_i - b x_i^2}] = 0$$

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

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

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

$$\sum_{i=1}^{n} b = b \sum_{i=1}^{n} 1$$

$\downarrow$ $b + b + \dots$ n times (bn)

$\rightarrow$ $\underline{b}(1 + 1 + \dots + n)$, $b(n)$, $\underline{\underline{bn}}$

$$\Rightarrow b = \frac{\sum_{i=1}^{n}\left[Y_i X_i - \bar{Y} X_i\right]}{\sum_{i=1}^{n}\left[X_i^2 - \bar{X} X_i\right]}$$

sum product$(X_i\, Y_i)$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\Rightarrow b = \frac{\sum_{i=1}^{n} X_i Y_i - \bar{Y}\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2 - \bar{X}\sum_{i=1}^{n} X_i} \qquad (n\bar{X})$$

$$\Rightarrow b = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{Y}\bar{X}}{\sum_{i=1}^{n} X_i^2 - n(\bar{X})^2}$$

Intuitively by using two expressions:

$$\sum_{i=1}^{n}\left(\bar{X}^2 - X_i \bar{X}\right) = 0$$

$$= n\bar{X}^2 - \sum X_i \bar{X}$$

$$= n\bar{X}^2 - \bar{X}\cdot\sum_{i=1}^{n} X_i \rightarrow n\bar{X}$$

$$= \cancel{n\bar{X}^2} - \cancel{n\bar{X}^2}$$

$$= 0$$

$\bar{X} = K$

$\bar{X}^2 = K^2$

$\sum_{i=1}^{n} K^2 = nK^2 \rightarrow n\bar{X}^2$

$\sum_{i=1}^{n} X_i \rightarrow n\bar{X}$

lly, $\sum_{i=1}^{n}\left(\bar{Y}^2 - Y_i\bar{Y}\right) = 0$

$$b = \frac{\sum_{i=1}^{n}\left(X_i Y_i - X_i\bar{Y}\right) + \sum_{i=1}^{n}\left(\bar{X}\bar{Y} - Y_i\bar{X}\right) \rightarrow 0}{\sum\left(X_i^2 - X_i\bar{X}\right) + \sum_{i=1}^{n}\left(\bar{X}^2 - X_i\bar{X}\right) \rightarrow 0}$$

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

reverse

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

$$\boxed{b = \frac{cov(x, y)}{Var(x)}}$$

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

Covariance Formula

For Population

$$Cov(x,y) = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{N}$$

$$\rho = \frac{cov(x, y)}{\sigma_x \sigma_y}$$

For Sample

$$Cov(x,y) = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{(N - 1)}$$