Linear regression:

$y = a_0 + a_1 x_1$

$\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_{1i} + \varepsilon_i$

where i is the observation of the ith term.

This can be expressed in terms of

$y = \Theta^T x + \varepsilon$

where $\Theta^T = \begin{bmatrix} \beta_1 \\ \beta_0 \end{bmatrix}$ → replacing a_0, a_1.

and $x = [\, x_1 , x_0 \,]$

where $x_0 = 1$.

We are iven y_i, and we know Θ. Assuming that the term y_i are i.i.d, theat means ε_i is also i.i.d.

Since our dataset is huge, it is reasonable to assume ε_i is normally distributed, hence, This implies that y_i is also normally distributed. according to the MGF.

$$E(\varepsilon_i) = 0$$

$$\varepsilon \sim N(0, \sigma^2).$$

$$y_i \sim N(\theta^T x_i, \sigma^2)$$

moving the gaussian curve to the right.

$$P(y_i \mid \theta^T x_i ; \theta)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}\right).$$

according to a normal distribution.

What we want to do is to find the θ, which is on our estimators

$$L(\bar{\theta}) = \prod_{i=1}^{n} \text{PDF of } y_i$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}\right).$$

taking the log likelihood,

$$l(\theta) = \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} + \sum_{i=1}^{n} -\frac{(y_i - \theta^T x_i)^2}{2\sigma^2}.$$

Since we want to maximize $l(\theta)$,

we have to minimize $\sum_{i=1}^{n} \frac{(y_i - \theta^T x_i)^2}{2\sigma^2}$.

$\frac{1}{2\sigma^2}$ is a constant,

Thus giving us $\sum_{i=1}^{n} (y_i - \theta^T x_i)^2$.

Here, we assume we want our data to be centered, hence $x_i = (x_i - \bar{x})$ rewritten.

$$\sum_{i=1}^{n} (y_i - b_0 - b_1(x_i - \bar{x}))^2 = e.$$

$$\frac{de}{db_0} = \sum_{i=1}^{n} 2[y_i - b_0 - b_1(x_i - \bar{x})](-1)$$

$$= \sum_{i=1}^{n} \frac{b_0 + b_1(x_i - \bar{x}) - y_i}{2}$$

$$= 0. \therefore -\sum_{i=1}^{n} y_i + nb_0 + b_1 \sum_{i=1}^{n} (x_i - \bar{x})$$

$$n b_0 = \sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} (x_i - \bar{x})$$

$$b_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - b_1 \underbrace{\sum_{i=1}^{n} \frac{(x_i - \bar{x})}{n}.}_{\to 0.}$$

$$b_0 = \bar{y}.$$

For a centered b_0, $b_0 = \bar{y}$.
but how did we get b_0?
Let b_0^* be the non-centered.

$$b_0^* = b_0 - b_1\bar{x}$$

so our non-centered
$b_0^* = \bar{y} - b_1\bar{x}$.

so the original b_0^* depends on b_1
then.

Let us discover about b_1 then.

$$\sum_{i=1}^{n} (y_i - b_0 - b_1(x_i - \bar{x}))^2.$$

$$\frac{de}{db_1} = \sum_{i=1}^{n} 2(y_i - b_0 - b_1(x_i - \bar{x}))(x_i - \bar{x})$$

$$= 0.$$

$$(x_i - \bar{x})$$

$$\sum_{i=1}^{n} y_i(x_i - \bar{x}) - n b_0 - \sum_{i=1}^{n} b_1 (x_i - \bar{x})^2$$

$$= 0.$$

$$b_1 = \frac{\sum_{i=1}^{n} y_i (x_i - \bar{x}) \cancel{- n b_0}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

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

$$\underline{b_1} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{S_{xx}^2}.$$

so now we have

b_0 & b_1.