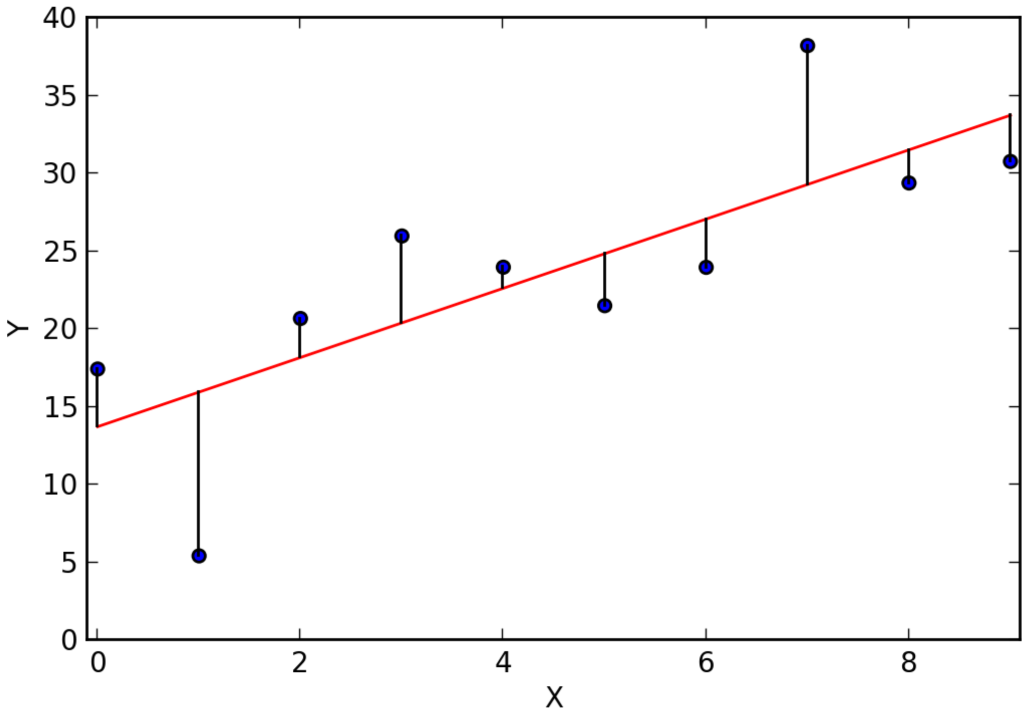
Linear Regression, Least square's method derivation



Each scatter point's coordinate is (x_i, y_i). Each black line represents error E_i.

Now we want to find the red best fit line $y = b_1 x + b_o$.

Minimize the square of the errorsE_i

$$E_{LR} = \sum_{i=1}^{n} E_i^2 = \sum (y_i - b_1 x_i - b_o)^2$$

We want to find the minimized E_{LR} with b_1 and b_o

$$\frac{\partial E_{LR}}{\partial b_1} = \sum \frac{\partial}{\partial b_1} (y_i - b_1 x_i - b_o)^2 = 0$$

$$= \sum -2x_i (y_i - b_1 x_i - b_o) = 0$$

$$= \sum -x_i (y_i - b_1 x_i - b_o) - (1)$$

$$\frac{\partial E_{LR}}{\partial b_0} = -(y_i - b_1 x_i - b_o) = 0 - (2)$$

We want to solve (1) and (2) for b_0 and b_1

Refine (1),

$$(1): \sum b_1 x_i^2 + \sum b_o x_i = \sum y_i x_i$$

$$(2): \sum b_1 x_i + \sum b_o = \sum y_i$$

Turn to matrix form:

$$\begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix} \begin{pmatrix} b_1 \\ b_o \end{pmatrix} = \begin{pmatrix} \sum y_i x_i \\ \sum y_i \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ b_o \end{pmatrix} = \begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix}^{-1} \begin{pmatrix} \sum y_i x_i \\ \sum y_i \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ b_o \end{pmatrix} = \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{pmatrix} \mathrm{n} & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \sum y_i x_i \\ \sum y_i \end{pmatrix}$$

$$b_1 = \frac{n\sum y_i x_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

Now using (2),

$$\sum_{i=1}^{n} b_o = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} b_1 x_i$$

$$n b_o = \sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i$$

$$b_o = \frac{\sum_{i=1}^{n} y_i}{n} - \frac{b_1 \sum_{i=1}^{n} x_i}{n}$$

$$b_o = \bar{y} - b_1 \bar{x}$$

,where $\bar{y}$and $\bar{x}$ are averages of the scattered coordinates