Example Simple Linear Regression Model: $y = \beta + \beta_1 x$

i	x_i	y_i
1	36	6
2	82	14
3	45	5
4	49	13
5	21	5
6	58	14
7	73	11
8	85	18
9	52	6
10	24	8

$$\begin{bmatrix} 1 & 36 \\ 1 & 82 \\ 1 & 45 \\ 1 & 49 \\ 1 & 21 \\ 1 & 58 \\ 1 & 73 \\ 1 & 85 \\ 1 & 52 \\ 1 & 24 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 6 \\ 14 \\ 5 \\ 13 \\ 5 \\ 14 \\ 11 \\ 18 \\ 6 \\ 8 \end{bmatrix}$$

X (model matrix) $\quad \boldsymbol{y}$ (observation vector)

The normal equations $X^T X\boldsymbol{\beta} = X^T\boldsymbol{y}$ are

$$\begin{bmatrix} 10 & 525 \\ 525 & 32085 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 100 \\ 5980 \end{bmatrix}$$

$X^T X$ is invertible, so there's a unique least squares solution:

$$\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}} = \begin{bmatrix} 10 & 525 \\ 525 & 32085 \end{bmatrix}^{-1} \begin{bmatrix} 100 \\ 5980 \end{bmatrix}$$

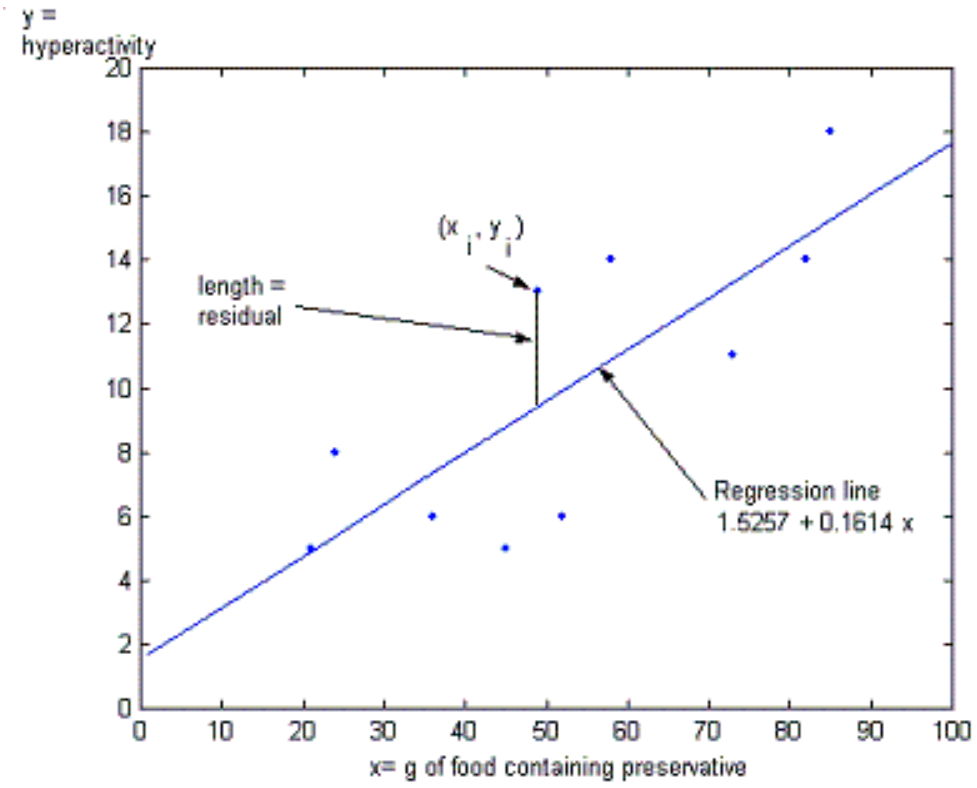
$$= \tfrac{1}{45225} \begin{bmatrix} 32085 & -525 \\ -525 & 10 \end{bmatrix} \begin{bmatrix} 100 \\ 5980 \end{bmatrix} \approx \begin{bmatrix} 1.5257 \\ 0.1614 \end{bmatrix} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix}$$

(*rounded to 4 decimal places*)

The regression line is

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x \approx 1.5257 + 0.1614x$$

The data points and the regression line are pictured below.



Example Simple Linear Regression Model

i	x_i	y_i
1	1	1
2	3	3
3	4	2

best fitting line $y = \beta_0 + \beta_1 x$? (*"predict y from x"*)

$$\begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

X (model matrix) $\quad \boldsymbol{y}$ (observation vector)

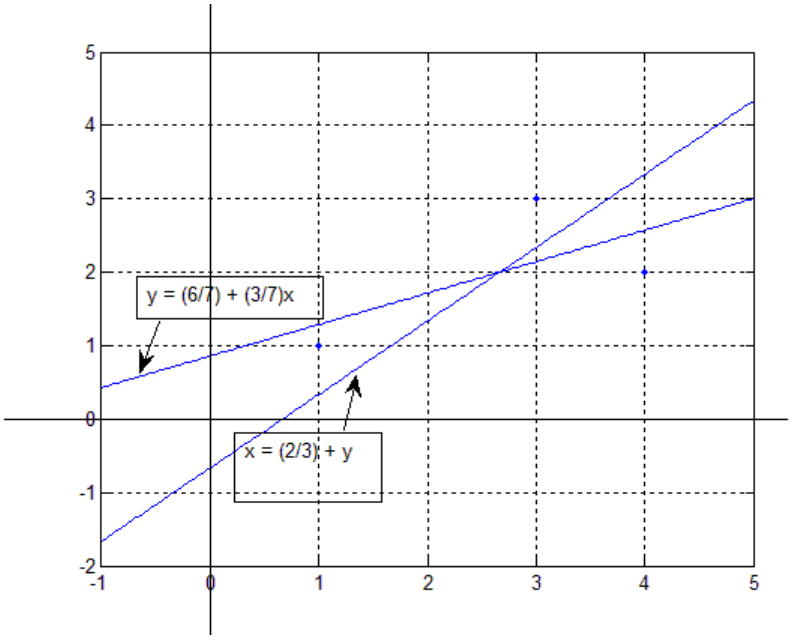
Normal equations: $X^T X\boldsymbol{\beta} = X^T \boldsymbol{y}$ are

$$\begin{bmatrix} 3 & 8 \\ 8 & 26 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 6 \\ 18 \end{bmatrix}$$

$X^T X$ is invertible (so unique least squares solution)

$$\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}} = \begin{bmatrix} 3 & 8 \\ 8 & 26 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 18 \end{bmatrix} = \begin{bmatrix} \frac{6}{7} \\ \frac{3}{7} \end{bmatrix} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix}$$

The regression line for y on x ("for predicting y from x") : $\quad \widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = \frac{6}{7} + \frac{3}{7}x$



Alternate model for the same data:

i	x_i	y_i
1	1	1
2	3	3
3	4	2

best fitting line $x = \alpha_0 + \alpha_1 y$? (*"predict x from y"*)

$$\begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 2 \end{bmatrix} \boldsymbol{\alpha} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$$

Y (model matrix) $\boldsymbol{x}$ (observation vector)

Normal Equation $Y^T Y \boldsymbol{\alpha} = Y^T \boldsymbol{x}$

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \boldsymbol{\alpha} = \begin{bmatrix} 8 \\ 18 \end{bmatrix}$$

$$\boldsymbol{\alpha} = \widehat{\alpha} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 8 \\ 18 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix} = \begin{bmatrix} \widehat{\alpha}_0 \\ \widehat{\alpha}_1 \end{bmatrix}$$

The regression line for x on y
("for predicting x from y")

$$\widehat{x} = \qquad \widehat{\alpha}_0 + \widehat{\alpha}_1 y = \tfrac{2}{3} + 1y$$

This is a <u>different</u> line. It it the line that minimizes

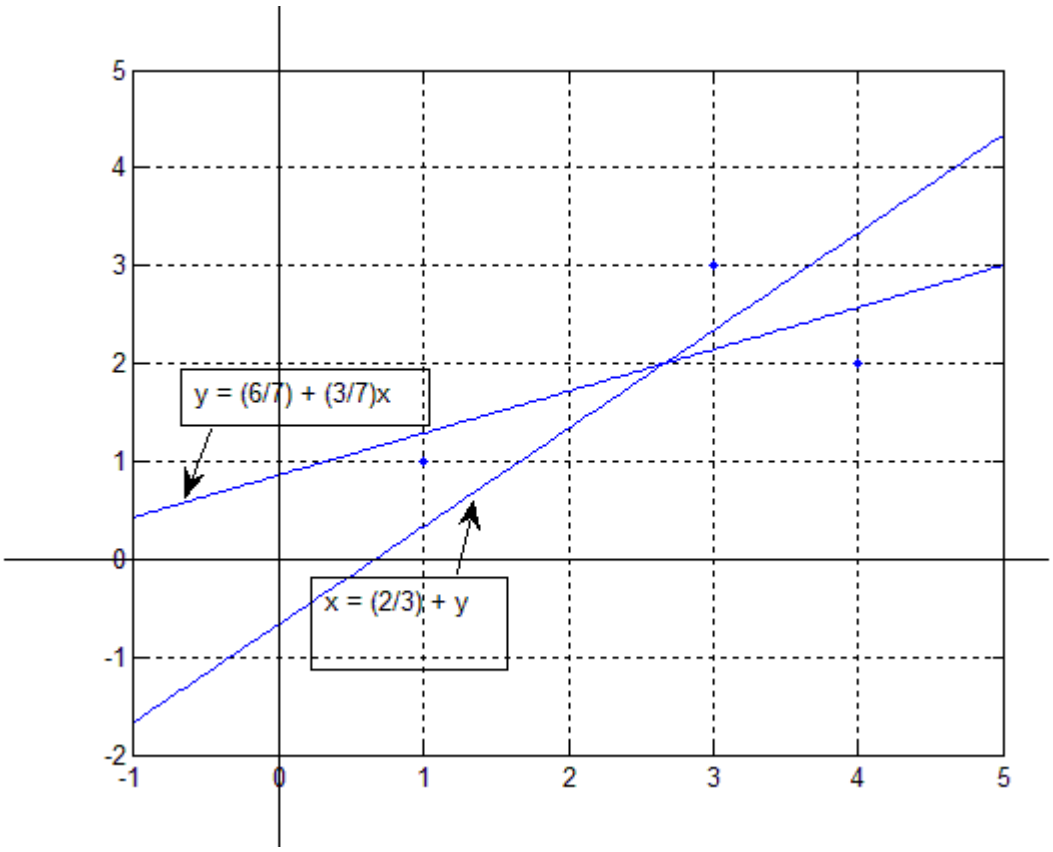
$$\sum_{i=1}^{n} (x_i - ((\ \widehat{\alpha}_0 + \widehat{\alpha}_1 y_i))^2$$

= "sum of squares of horizontal distances from data points to line"

For the first model, we found the line that minimizes $\sum_{i=1}^{n} (y_i - (\ \widehat{\beta}_0 + \widehat{\beta}_1 x_i))^2$

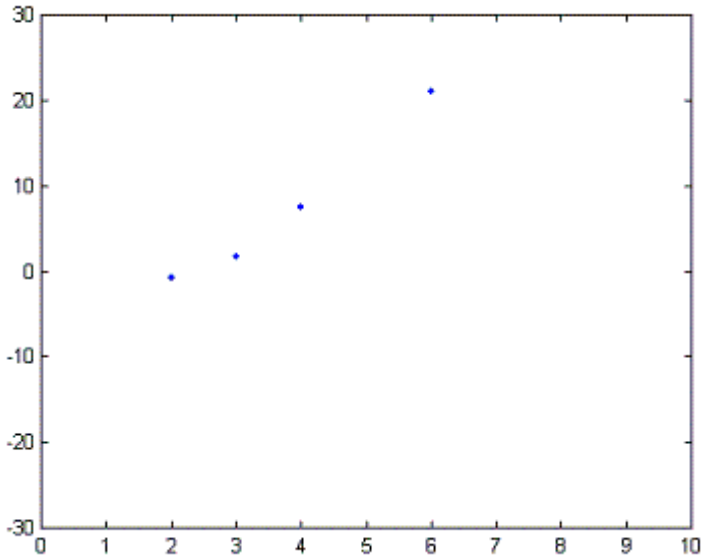
= "sums of squares of vertical distances from data points to line"

Here are the two regression lines.



i	x_i	y_i
1	2.1	-0.8
2	3.2	1.6
3	4.5	7.5
4	6.2	21.1
5	7.1	30.1

Find best fit by: $y = \beta_0 + \beta_1 x + \beta_2 x^2$



"Exact fit" would require

$$\begin{aligned} \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 &= y_1 \\ \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 &= y_2 \\ &\vdots \\ \beta_0 + \beta_1 x_5 + \beta_2 x_5^2 &= y_5 \end{aligned} \qquad \text{or}$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_5 & x_5^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_5 \end{bmatrix}$$

$$X \qquad \boldsymbol{\beta} \qquad = \quad \boldsymbol{y}$$

Inconsistent. Find the $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ so that $||\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}||$ is as small as possible.

$X^T X\boldsymbol{\beta} = X^T\boldsymbol{y}$ is (*approximately, after rounding*)

$$\begin{bmatrix} 5 & 23.1 & 123.8 \\ 23.1 & 123.8 & 729.4 \\ 123.8 & 729.4 & 455.32 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 59.5 \\ 381.7 \\ 2493.2 \end{bmatrix}$$

so that (Using Matlab)

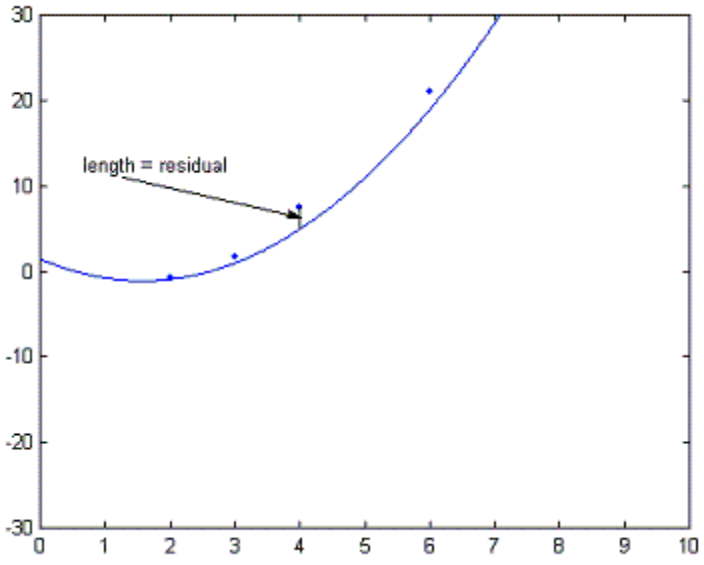
$$\boldsymbol{\beta} = \widehat{\beta} = \begin{bmatrix} 5 & 23.1 & 123.8 \\ 23.1 & 123.8 & 729.4 \\ 123.8 & 729.4 & 455.32 \end{bmatrix}^{-1} \begin{bmatrix} 59.5 \\ 381.7 \\ 2493.2 \end{bmatrix}$$

$$\approx \begin{bmatrix} 1.4759 \\ -3.2591 \\ 1.0295 \end{bmatrix} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix}.$$

The "best fitting quadratic" is

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x + \widehat{\beta}_2 x^2$$

$$\approx 1.4759 - 3.2591x + 1.0295x^2$$



This is the "best fit" in the sense that

it minimizes the sum of the squared (vertical) residuals.