Example (When the regression line is not unique)

Simplest possible example: two data points $(2, 0)$ and $(2, 1)$ that lie on a vertical line.

$X\boldsymbol{\beta} = \boldsymbol{y}$ is $\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}\boldsymbol{\beta} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and the normal equation is

$X^T X\boldsymbol{\beta} = X^T\boldsymbol{y}$ $\begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}\boldsymbol{\beta} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$: infinitely many solutions

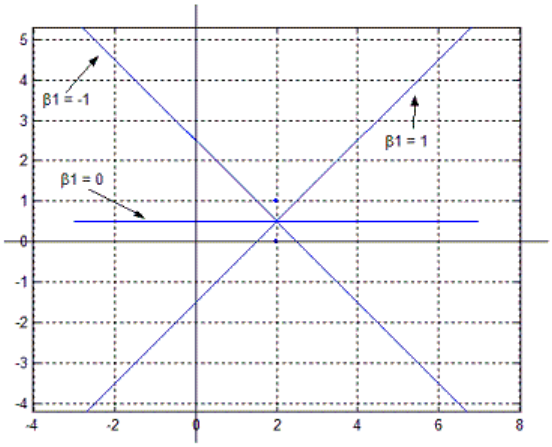
$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} + \widehat{\beta}_1 \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \quad \text{so that}$$

$\widehat{\boldsymbol{\beta}}_0 = \frac{1}{2} - 2\widehat{\beta}_1$ (where $\widehat{\beta}_1$ is free)

Three of the regression lines

$\widehat{\boldsymbol{\beta}}_1 = -1$ $\widehat{\boldsymbol{\beta}}_1 = 0$ $\widehat{\boldsymbol{\beta}}_1 = 1$

$\frac{5}{2} - 1x$ $\frac{1}{2} + 0x$ $-\frac{3}{2} + 1x$



For each regression line: when $x = 2$, the "predicted" observation $\widehat{y} = \frac{1}{2}$
so each residual error $= y_i - \widehat{y}_i = \frac{1}{2}$, so the least squares error
for any of these lines

$$= \sqrt{(0 - \tfrac{1}{2})^2 + (1 - \tfrac{1}{2})^2} = \sqrt{\tfrac{1}{2}} = \tfrac{\sqrt{2}}{2} \approx 0.707$$

Example Curve best fitting data points in $\mathbb{R}^2$

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: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

Setup: "Exact fit" would require

$$\begin{aligned} \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3 &= y_1 \\ \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \beta_3 x_2^3 &= y_2 \\ &\vdots \\ \beta_0 + \beta_1 x_5 + \beta_2 x_5^2 + \beta_3 x_5^3 &= y_5 \end{aligned} \qquad \text{or}$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_5 & x_5^2 & x_5^3 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_5 \end{bmatrix}$$

$$\uparrow \qquad \uparrow \qquad \uparrow$$

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

Solve $X^T X\beta = X^T y$

Example

$(x_1, y_1, z_1), \quad ..., \quad (x_n, y_n, z_n)$ in $\mathbb{R}^3$.

Find the function $z = \beta_0 + \beta_1 x + \beta_2 y$ that best fits the data points. (*Note: the graph of the function is a plane in* $\mathbb{R}^3$)

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \quad \text{where}$$

$$\uparrow$$

To be specific: Find the function $z = \beta_0 + \beta_1 x + \beta_2 y$ that best fits the points $(1, 2, 3),\ (0, 2, 4),\ (-1, -2, 1)$, and $(1, 1, 0)$

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & 2 \\ 1 & -1 & -2 \\ 1 & 1 & 1 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 3 \\ 4 \\ 1 \\ 0 \end{bmatrix}$$

The normal equation is

$$\begin{bmatrix} 4 & 1 & 3 \\ 1 & 3 & 5 \\ 3 & 5 & 13 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 8 \\ 2 \\ 12 \end{bmatrix}$$

The least squares solution is

$$\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}} \approx \begin{bmatrix} 1.4783 \\ -2.2174 \\ 1.4348 \end{bmatrix}$$

(*rounded to 4 decimal places*)

The best fitting plane is

$$z = 1.4783 - 2.2174x + 1.4348y$$