Example Find the least squares solution for $\begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$A^T A = \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 13 & -1 \\ -1 & 2 \end{bmatrix}$$

$$A^T \boldsymbol{b} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

The system of normal equations is

$$\begin{bmatrix} 13 & -1 \\ -1 & 2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}.$$

Here $A^T A$ is invertible: $\begin{bmatrix} 13 & -1 \\ -1 & 2 \end{bmatrix}^{-1} = \frac{1}{25} \begin{bmatrix} 2 & 1 \\ 1 & 13 \end{bmatrix}$, so there is a unique least squares solution:

$$\widehat{\boldsymbol{x}} = \frac{1}{25} \begin{bmatrix} 2 & 1 \\ 1 & 13 \end{bmatrix} \begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} \\ \frac{1}{5} \end{bmatrix}.$$

In this case, the least squares error is

(*) $\quad || \boldsymbol{b} - A\, \widehat{\boldsymbol{x}} \, || = || \boldsymbol{b} - \widehat{\boldsymbol{b}} \, || = || \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} \frac{2}{5} \\ \frac{1}{5} \end{bmatrix} || = || \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} || = 0 \quad !$

Least squares error $= 0$ means that $A\, \widehat{\boldsymbol{x}} = \widehat{\boldsymbol{b}} = \boldsymbol{b}$: the least squares solution $\widehat{\boldsymbol{x}} = \begin{bmatrix} \frac{2}{5} \\ \frac{1}{5} \end{bmatrix}$ is the exact solution. We didn't notice, in the beginning, that $A\boldsymbol{x} = \boldsymbol{b}$ is a consistent system.

No harm done $-$ but using the least squares method wasn't necessary (*and if I had not computed the least squares error, I would probably not have noticed that I really found an exact solution to the original system*).

Example Fitting data to a simple linear regression model: $y = \beta_0 + \beta_1 x$

To see whether a widely used food preservative contributes to the hyperactivity of preschool children, a dietitian chose a random sample of 10 four-year-olds known to be fairly hyperactive from various nursery schools and observed their behavior 45 minutes after having eaten measured amounts of food containing the preservative. For the i^{th} child, the table shows the amount x_i of food containing the preservative consumed (measured in g) and y_i shows a subjective rating of hyperactivity, on a 1-20 scale, based on the child's restlessness and interaction with other children:

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

In the notation of the preceding discussion, we have

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

(*the data points shown in the figure below)*

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

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

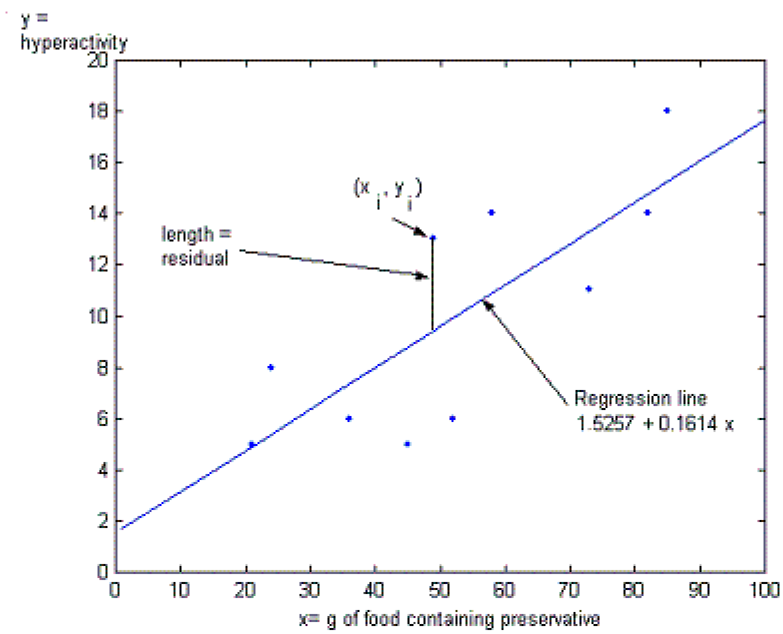
$X^T X$ is invertible, so there is a unique least squares solution:

$$\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}} = \begin{bmatrix} 10 & 525 \\ 525 & 32085 \end{bmatrix}^{-1} \begin{bmatrix} 100 \\ 5980 \end{bmatrix} = \frac{1}{45225} \begin{bmatrix} 32085 & -525 \\ -525 & 10 \end{bmatrix} \begin{bmatrix} 100 \\ 5980 \end{bmatrix}$$

$$\approx \begin{bmatrix} 1.5257 \\ 0.1614 \end{bmatrix} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} \text{ (\textit{rounded to 4 decimal places}).}$$

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.



$\beta_0 + \beta_1 x$:   the regression line $\widehat{\beta}_0 + \widehat{\beta}_1 x$ corresponds to choosing $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix}$.

This makes $||X\widehat{\boldsymbol{\beta}} - \boldsymbol{y}|| \leq ||X\boldsymbol{\beta} - \boldsymbol{y}||$ for all choices of $\boldsymbol{\beta}$

$$||$$

$$\sqrt{\sum (\textit{residuals})^2}$$