Recall:

1) The least square solutions of a matrix equation $A\mathrm{x} = \mathbf{b}$ are the solutions of the equation

$$A\mathrm{x} = \mathrm{proj}_{\mathrm{Col}(A)}\mathbf{b}$$

2) If $A\mathrm{x} = \mathbf{b}$ is a consistent equation, then $\mathbf{b} \in \mathrm{Col}(A)$, and $\mathrm{proj}_{\mathrm{Col}(A)}\mathbf{b} = \mathbf{b}$. In such case the least square solutions of $A\mathrm{x} = \mathbf{b}$ are just the ordinary solutions.

3) If $A\mathrm{x} = \mathbf{b}$ is inconsistent, then the least square solutions are the best substitute for the (nonexistent) ordinary solutions.

4) If $\{\mathrm{v}_1, \ldots, \mathrm{v}_k\}$ is an orthogonal basis of a subspace V of $\mathbb{R}^n$ then

$$\mathrm{proj}_V \mathrm{w} = \left(\frac{\mathrm{w}\cdot\mathrm{v}_1}{\mathrm{v}_1\cdot\mathrm{v}_1}\right)\mathrm{v}_1 + \ldots + \left(\frac{\mathrm{w}\cdot\mathrm{v}_k}{\mathrm{v}_k\cdot\mathrm{v}_k}\right)\mathrm{v}_k$$

5) If $\{\mathrm{v}_1, \ldots, \mathrm{v}_k\}$ is an arbitrary basis of V then we can use the Gram-Schmidt process to obtain an orthogonal basis of V.

How to compute least square solutions of $A\mathbf{x} = \mathbf{b}$

(version 1.0)

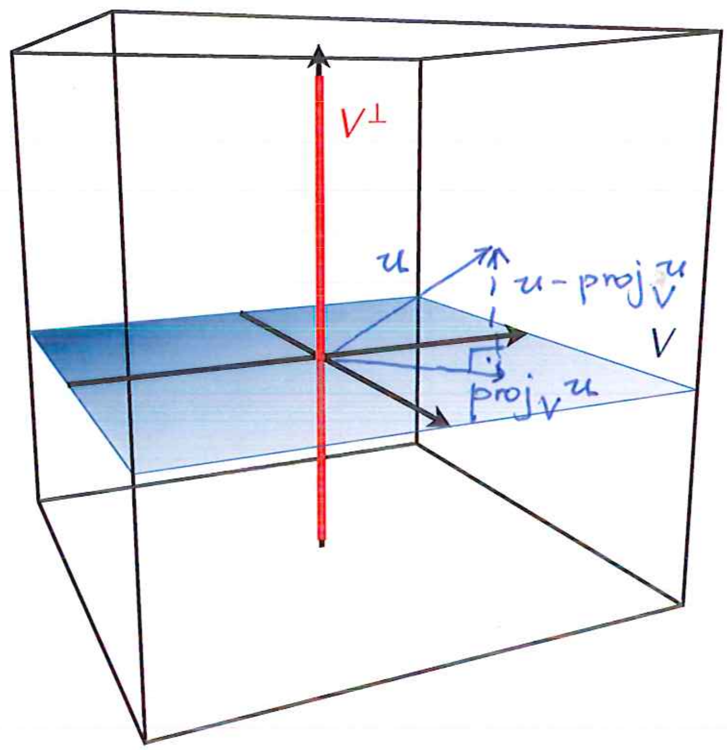
1) Compute a basis of $\mathrm{Col}(A)$.
2) Use the Gram-Schmidt process to get an orthogonal basis of $\mathrm{Col}(A)$.
3) Use the orthogonal basis to compute $\mathrm{proj}_{\mathrm{Col}(A)}\mathbf{b}$.
4) Compute solutions of the equation $A\mathbf{x} = \mathrm{proj}_{\mathrm{Col}(A)}\mathbf{b}$.

Next goal: Simplify this.

Definition

If V is a subspace of $\mathbb{R}^n$ then the *orthogonal complement* of V is the set $V^\perp$ of all vectors orthogonal to V:

$$V^\perp = \{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{w} \cdot \mathbf{v} = 0\}$$



Note: If $V \subseteq \mathbb{R}^n$ and $u \in \mathbb{R}^n$ then $(u - \text{proj}_V u) \in V^\perp$

Propositon

If V is a subspace of $\mathbb{R}^n$ then:

1) $V^\perp$ is also a subspace of $\mathbb{R}^n$.
2) For each vector $\mathbf{w} \in \mathbb{R}^n$ there exist unique vectors $\mathbf{v} \in V$ and $\mathbf{z} \in V^\perp$ such that $\mathbf{w} = \mathbf{v} + \mathbf{z}$.

Definition

If A is an $m \times n$ matrix then the *row space* of A is the subspace $\mathrm{Row}(A)$ of $\mathbb{R}^n$ spanned by rows of A.

Example

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$\mathrm{Row}(A) = \mathrm{Span}\left(\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}\right)$$

Propositon

If A is a matrix then

$$\mathrm{Row}(A)^{\perp} = \mathrm{Nul}(A)$$

Note:

If $A = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$ where $r_1, \dots, r_m$ - rows of A

then $Av = \begin{bmatrix} r_1 \cdot v \\ r_2 \cdot v \\ \vdots \\ r_m \cdot v \end{bmatrix}$.

e.g.:

$$\begin{array}{c c} & \\ r_1 & \\ r_2 & \end{array}\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \cdot \overset{v}{\begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}} = \begin{bmatrix} 1\cdot 7 + 2\cdot 8 + 3\cdot 9 \\ 4\cdot 7 + 5 \cdot 8 + 6 \cdot 9 \end{bmatrix} = \begin{bmatrix} r_1 \cdot v \\ r_2 \cdot v \end{bmatrix}$$

Proof of Proposition

Let $A = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$. We have:

$v \in \mathrm{Nul}(A)$ iff $Av = 0$

iff $r_1 v = 0, \dots, r_m \cdot v = 0$

iff $v \in \mathrm{Row}(A)^{\perp}$

Corollary

If A is a matrix then

$$\mathrm{Col}(A)^\perp = \mathrm{Nul}(A^T)$$

Note: $\mathrm{Col}(A) = \mathrm{Row}(A^T)$

e.g.: $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ $\qquad A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

$\mathrm{Col}(A) = \mathrm{Span}\left(\begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \end{bmatrix}\right)$ $\qquad \mathrm{Row}(A^T) = \mathrm{Span}\left(\begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \end{bmatrix}\right)$

Proof of Corollary

$$\mathrm{Col}(A)^\perp = \mathrm{Row}(A^T)^\perp = \mathrm{Nul}(A^T)$$

Back to least square solutions

Theorem

A vector $\hat{\mathbf{x}}$ is a least square solution of a matrix equation

$$A\mathbf{x} = \mathbf{b}$$

if and only if $\hat{\mathbf{x}}$ is an ordinary solution of the equation

$$(A^T A)\mathbf{x} = A^T\mathbf{b}$$

Proof: If $\hat{x}$ is a least square solution of $Ax = b$ then:

$$A\hat{x} = \text{proj}_{Col(A)} b$$

This gives:

$$(b - A\hat{x}) = (b - \text{proj}_{Col(A)} b) \in Col(A)^\perp = Nul(A^T)$$

We obtain:

$$A^T(b - A\hat{x}) = 0$$

$$A^T b - A^T A\hat{x} = 0$$

$$A^T b = A^T A\hat{x}$$

Thus $\hat{x}$ is a solution of the equation

$$A^T A x = A^T b$$

Theorem

The equation

$$(A^T A)\mathbf{x} = A^T\mathbf{b}$$

is called the *normal equation* of $A\mathbf{x} = \mathbf{b}$.

How to compute least square solutions of $Ax = \mathbf{b}$

(version 2.0)

1) Compute $A^T A$, $A^T \mathbf{b}$.
2) Solve the normal equation $(A^T A)\mathrm{x} = A^T \mathbf{b}$.

Example. Compute least square solutions of the following equation:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}}_{A} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}_{b}$$

Note:

$$\begin{cases} x_1 + x_2 = 1 \\ 2x_2 = 2 \\ \underbrace{0 = 3}_{\text{no solutions!}} \end{cases}$$

Solution:

$$A^T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

The normal equation:

$$A^T A x = A^T b$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

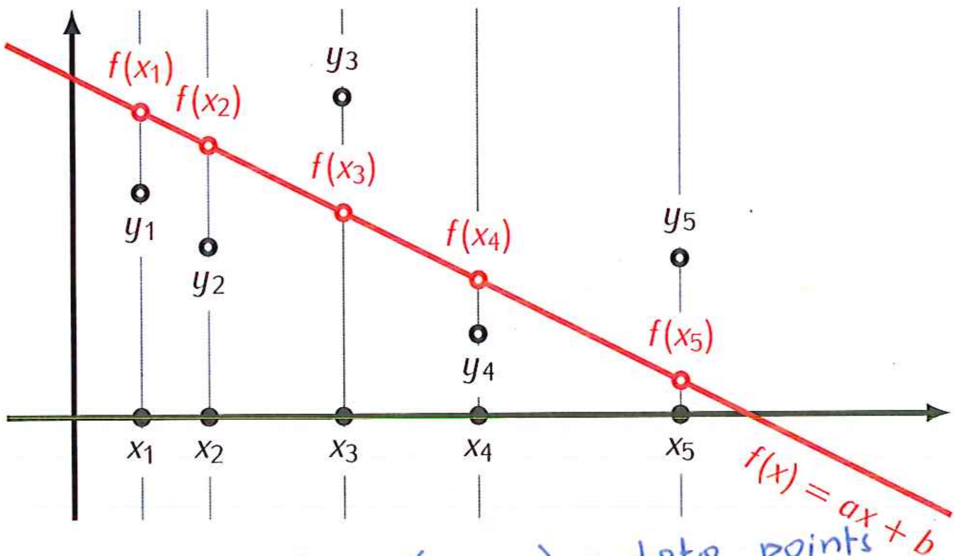
aug. matrix:

$$\left[\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 5 & 5 \end{array}\right] \xrightarrow{\text{row red.}} \begin{array}{c} \begin{array}{cc} x_1 & x_2 \end{array} \\ \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 1 \end{array}\right] \end{array}$$

least square solution:

$$\boxed{\hat{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}}$$

Application: Least square lines



$(x_1, y_1), (x_2, y_2), \dots, (x_5, y_5)$ - data points

data values, computed values

x_1	y_1	$f(x_1)$
x_2	y_2	$f(x_2)$
x_3	y_3	$f(x_3)$
x_4	y_4	$f(x_4)$
x_5	y_5	$f(x_5)$

Definition

If $(x_1, y_1), \dots, (x_p, y_p)$ are points on the plane then the *least square line* for these points is the line given by an equation $f(x) = ax + b$ such that the number

$$\operatorname{dist}\left(\begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_p) \end{bmatrix}\right) = \sqrt{(y_1 - f(x_1))^2 + \ldots + (y_p - f(x_p))^2}$$

is the smallest possible.

Note: If $f(x) = ax+b$ then

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_p) \end{bmatrix} = \begin{bmatrix} ax_1+b \\ ax_2+b \\ \vdots \\ ax_p+b \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_p & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

This gives: $f(x) = ax+b$ is the least square line if

$$\operatorname{dist}\left(\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_p & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} \right)$$

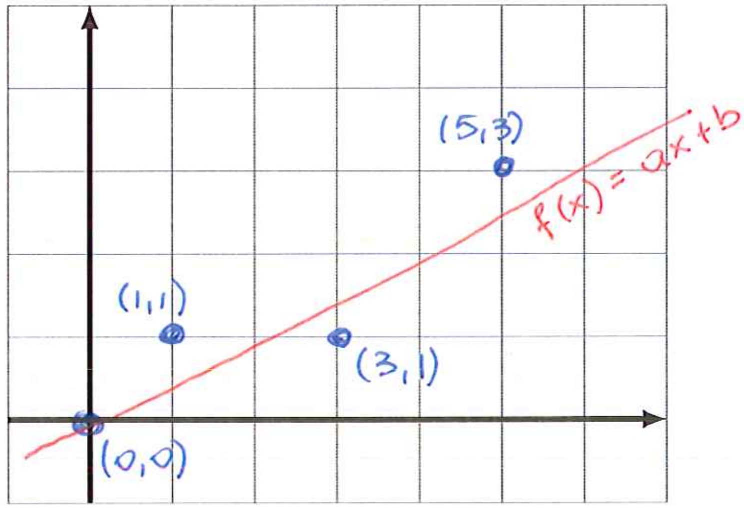
is as small as possible.

Proposition

The line $f(x) = ax+b$ is the least square line for points $(x_1, y_1), \ldots, (x_p, y_p)$ if the vector $\begin{bmatrix} a \\ b \end{bmatrix}$ is the least square solution of the equation

$$\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_p & 1 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$$

Example. Find the equation of the least square line for the points $(0, 0)$, $(1, 1)$, $(3, 1)$, $(5, 3)$.



Solution: The least square line is given by

$$f(x) = ax + b$$

where $\begin{bmatrix} a \\ b \end{bmatrix}$ is a least square solution of

$$\underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 3 & 1 \\ 5 & 1 \end{bmatrix}}_{A} \begin{bmatrix} a \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \\ 3 \end{bmatrix}}_{b}$$

$$A^T A = \begin{bmatrix} 35 & 9 \\ 9 & 4 \end{bmatrix} \qquad A^T b = \begin{bmatrix} 19 \\ 15 \end{bmatrix}$$

Normal equation:

$$\begin{bmatrix} 35 & 9 \\ 9 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 19 \\ 5 \end{bmatrix}$$

aug. matrix:

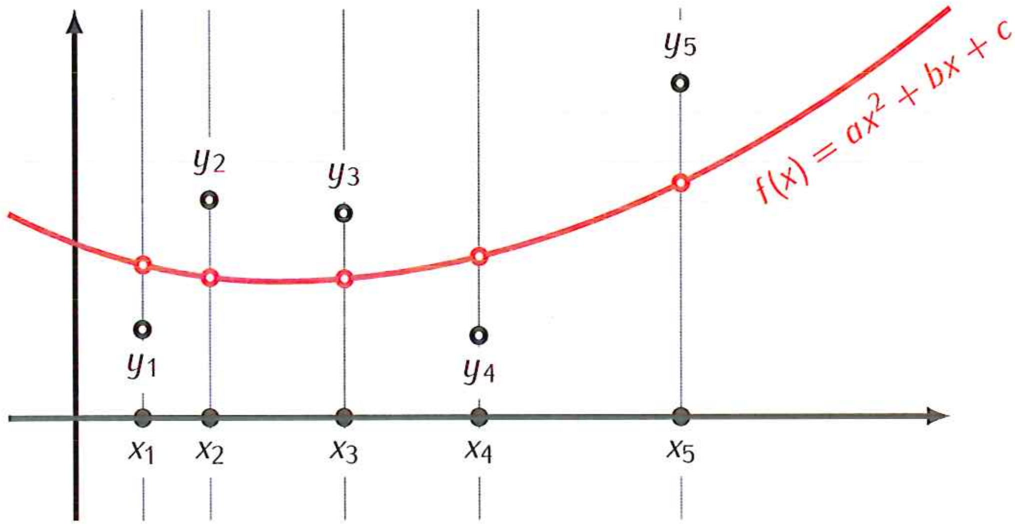
$$\left[\begin{array}{cc|c} 35 & 9 & 19 \\ 9 & 4 & 5 \end{array}\right] \xrightarrow{\text{row red}} \begin{array}{c} \begin{array}{cc} a & b \end{array} \\ \left[\begin{array}{cc|c} 1 & 0 & 31/59 \\ 0 & 1 & 4/59 \end{array}\right] \end{array} \qquad \begin{cases} a = 31/59 \\ b = 4/59 \end{cases}$$

$$\boxed{\begin{array}{c} \text{the least square line:} \\ f(x) = \dfrac{31}{59} x + \dfrac{4}{59} \end{array}}$$

Application: Least square curves

The above procedure can be used to determine curves other than lines that fit a set of points in the least square sense.

Example: Least square parabolas



Definition

If $(x_1, y_1), \ldots, (x_p, y_p)$ are points on the plane then the *least square parabola* for these points is the parabola given by an equation $f(x) = ax^2 + bx + c$ such that the number

$$\operatorname{dist}\left(\begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_p) \end{bmatrix}\right) = \sqrt{(y_1 - f(x_1))^2 + \ldots + (y_p - f(x_p))^2}$$

is the smallest possible.

Note: If $f(x) = ax^2 + bx + c$ then

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_p) \end{bmatrix} = \begin{bmatrix} ax_1^2 + bx_1 + c \\ ax_2^2 + bx_2 + c \\ \vdots \\ ax_p^2 + bx_p + c \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & & \\ x_p^2 & x_p & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

This gives: $f(x) = ax^2 + bx + c$ is the least square parabola if

$$\operatorname{dist}\left(\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_p^2 & x_p & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} \right)$$

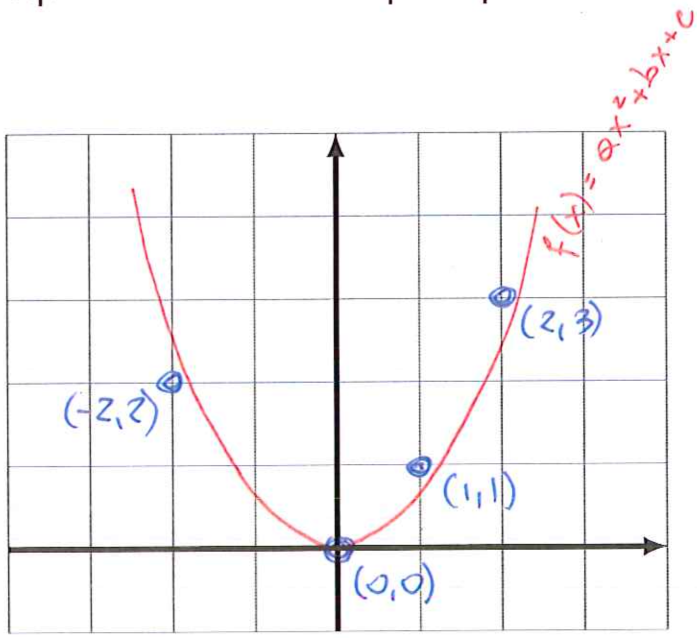
is as small as possible.

Proposition

The parabola $f(x) = ax^2 + bx + c$ is the least square parabola for points $(x_1, y_1), \ldots, (x_p, y_p)$ if the vector $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is the least square solution of the equation

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ \vdots & \vdots & \\ x_p^2 & x_p & 1 \end{bmatrix} \cdot \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$$

Example. Find the equation of the least square parabola for the points $(-2,2)$, $(0,0)$, $(1,1)$, $(2,3)$.



Solution: We need to find a least square solution of the equation

$$\underbrace{\begin{array}{c} \begin{array}{ccc} x^2 & x & 1 \end{array} \\ \begin{bmatrix} (-2)^2 & -2 & 1 \\ 0^2 & 0 & 1 \\ 1^2 & 1 & 1 \\ 2^2 & 2 & 1 \end{bmatrix} \end{array}}_{A} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \underbrace{\begin{array}{c} y \\ \begin{bmatrix} 2 \\ 0 \\ 1 \\ 3 \end{bmatrix} \end{array}}_{b}$$

Normal equation:

$$\overbrace{\begin{bmatrix} 33 & 1 & 9 \\ 1 & 9 & 1 \\ 9 & 1 & 4 \end{bmatrix}}^{A^TA} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \overbrace{\begin{bmatrix} 21 \\ 3 \\ 6 \end{bmatrix}}^{A^Tb}$$

aug. matrix

$$\left[\begin{array}{ccc|c} 31 & 1 & 9 & 21 \\ 1 & 9 & 1 & 3 \\ 9 & 1 & 4 & 6 \end{array}\right] \xrightarrow{\text{row red.}} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \left[\begin{array}{ccc|c} 1 & 0 & 0 & 27/37 \\ 0 & 1 & 0 & 51/185 \\ 0 & 0 & 1 & -39/185 \end{array}\right] \end{array} \qquad \begin{cases} a = 27/37 \\ b = 51/185 \\ c = -39/185 \end{cases}$$

$$\boxed{\begin{array}{c} \text{The least square parabola:} \\ f(x) = \frac{27}{37}x^2 + \frac{51}{185}x - \frac{39}{185} \end{array}}$$