Chapter 4 Orthogonality

- Overdetermined linear system $A\vec{x} = \vec{b}$

$A \in \mathbb{R}^{m \times n}$, $m > n$: more equations than unknowns

$$\begin{array}{c} \\ m \end{array} \begin{array}{c} n \\ \boxed{} \\ A \end{array} \begin{array}{c} n\times 1 \\ [\;] \\ \vec{x} \end{array} = \begin{array}{c} m\times 1 \\ [\;] \\ \vec{b} \end{array}$$

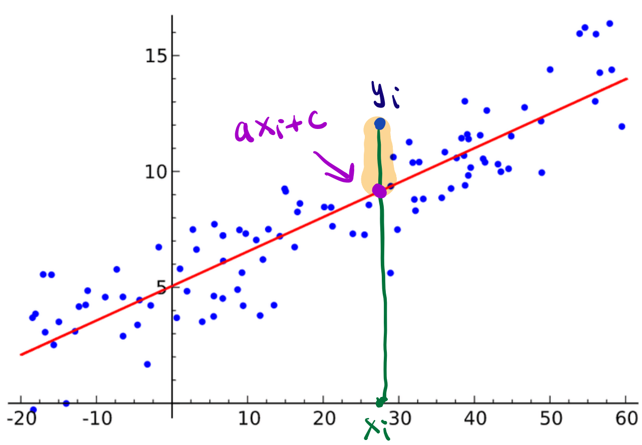
n cols $\Rightarrow$ $\underbrace{\dim(\mathrm{Col}(A))}_{\text{col rank}} \leq n < m$

$\mathrm{Col}(A) \subseteq \mathbb{R}^m \Rightarrow \mathrm{Col}(A) \neq \mathbb{R}^m$

$\Rightarrow$ There are vectors $\vec{b} \in \mathbb{R}^m$ that are not in $\mathrm{Col}(A)$

$\vec{b} \notin \mathrm{Col}(A) \Rightarrow A\vec{x} = \vec{b}$ has no sols.

- It comes from finding Line / Curve fitting data



We are given data of coordinates (x_i, y_i)

$i = 1, \ldots, m$

Want a line $y = ax + c$ s.t.

$$y_1 = ax_1 + c$$
$$y_2 = ax_2 + c$$
$$\vdots$$
$$y_m = ax_m + c$$

$$\overset{A}{\begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix}} \overset{\vec{x}}{\begin{pmatrix} a \\ c \end{pmatrix}} = \overset{\vec{b}}{\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}} \qquad \vec{x} = \begin{pmatrix} a \\ c \end{pmatrix}$$

First, we do not have the perfect line passing all the data points, i.e.,

$A\vec{x} = \vec{b}$ is overdetermined thus no sol.

Second, we only want the line that is the "closest" to all points, i.e., we want to find (a, c) for minimizing the "errors":

$$\text{Line Fitting Error} = \sum_{i=1}^{m} |ax_i + c - y_i|^2$$

$$\underline{A\vec{x}} = \begin{pmatrix} \boxed{\begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{matrix}} & \boxed{\begin{matrix} 1 \\ 1 \\ \vdots \\ 1 \end{matrix}} \end{pmatrix} \underline{\begin{pmatrix} a \\ c \end{pmatrix}} = \begin{pmatrix} ax_1 + c \\ ax_2 + c \\ \vdots \\ ax_m + c \end{pmatrix}$$

$$\vec{b} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

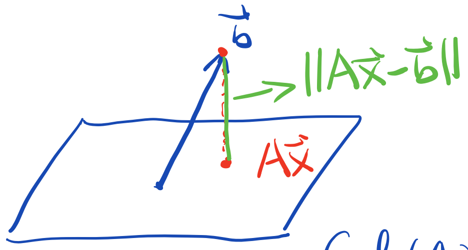
$$A\vec{x} - \vec{b} = \begin{pmatrix} ax_1 + c - y_1 \\ \vdots \\ ax_m + c - y_m \end{pmatrix}$$

$$\text{Line Fitting Error} = \|A\vec{x} - \vec{b}\|^2 \qquad A\vec{x} = \vec{b}$$

$$A\vec{x} = a\begin{pmatrix}x_1\\ \vdots \\ x_m\end{pmatrix} + c\begin{pmatrix}|\\ \vdots \\ |\end{pmatrix} \in Col(A)$$

$\vec{x} = \begin{pmatrix}a\\ c\end{pmatrix}$

$\begin{cases} m=3 \\ n=2 \end{cases}$



$$\min_{\vec{x}\in\mathbb{R}^2} \|A\vec{x}-\vec{b}\|^2$$

$Col(A) \subseteq \mathbb{R}^m$

① $A\vec{x}$ is a vector in $Col(A)$

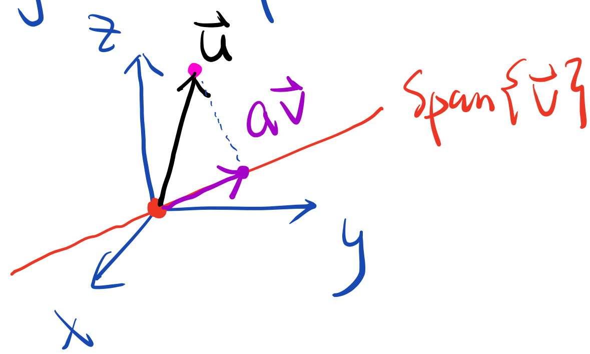
② Any vector in $Col(A)$ can be written as

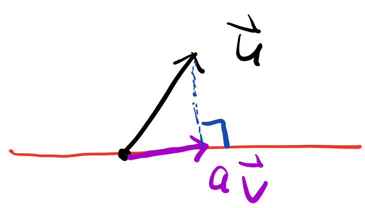
$$s\begin{pmatrix}x_1\\ \vdots \\ x_n\end{pmatrix} + t\begin{pmatrix}|\\ \vdots \\ |\end{pmatrix} = \begin{pmatrix}x_1 & | \\ \vdots & \vdots \\ x_n & |\end{pmatrix}\begin{pmatrix}s\\ t\end{pmatrix} = A\hat{x} \quad \text{where } \hat{x} = \begin{pmatrix}s\\ t\end{pmatrix}$$

So minimizing the fitting error is the same as find the shortest distance between $\vec{b}$ and the subspace $Col(A)$.

Review of Projection Onto a Line

Projection of a vector onto another one:





So we look for a vector $a\vec{v}$ satisfying

$$(\vec{u} - a\vec{v}) \perp \vec{v}$$

$$\Leftrightarrow \langle \vec{u} - a\vec{v}, \vec{v} \rangle = 0$$

$$\Leftrightarrow \langle \vec{u}, \vec{v} \rangle - a\langle \vec{v}, \vec{v} \rangle = 0$$

$$\Leftrightarrow a\langle \vec{v}, \vec{v} \rangle = \langle \vec{u}, \vec{v} \rangle$$

$$\Leftrightarrow a = \frac{\langle \vec{u}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle} = \frac{\vec{v}^T \vec{u}}{\vec{v}^T \vec{v}} = \frac{\vec{u} \cdot \vec{v}}{\vec{v} \cdot \vec{v}}$$

$$= \frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2}$$

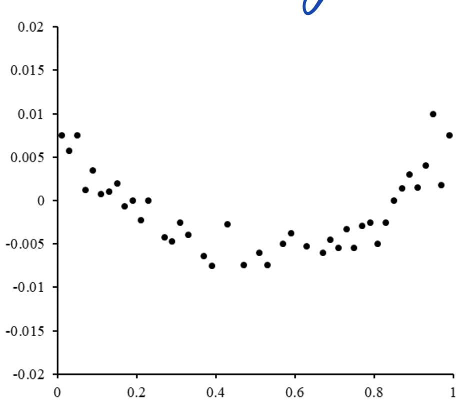
For $\vec{u}, \vec{v} \in \mathbb{R}^n$, the projection of $\vec{u}$ onto $\vec{v}$ is

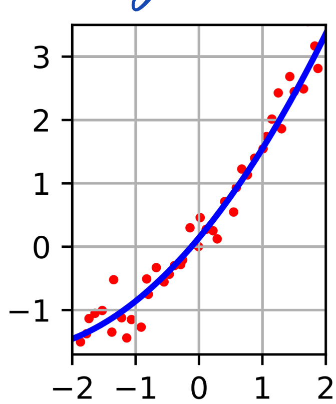
$$P_{\vec{v}}(\vec{u}) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2} \vec{v}$$

Ex: Projection of $\vec{u} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ onto $\vec{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

$$P_{\vec{v}}(\vec{u}) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2} \vec{v} = \frac{6}{(1+1+1)} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}.$$

- Curve Fitting: in general we can consider curves by using polynomials, trigonometrics, etc.





We are given data of coordinates (x_i, y_i)
$i = 1, \ldots, m$

Want a quadratic fit:

$$y_1 = dx_1^2 + ax_1 + c$$
$$y_2 = dx_2^2 + ax_2 + c$$
$$\vdots$$
$$y_m = dx_m^2 + ax_m + c$$

$$\begin{pmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_m^2 & x_m & 1 \end{pmatrix} \begin{pmatrix} d \\ a \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

$$A \quad \vec{x} \quad = \quad \vec{b}$$

Want a cubic fit

$$y_1 = ex_1^3 + dx_1^2 + ax_1 + c$$
$$y_2 = ex_1^3 + dx_2^2 + ax_2 + c$$
$$\vdots$$
$$y_m = ex_m^3 + dx_m^2 + ax_m + c$$

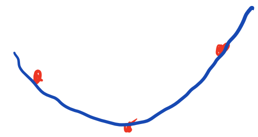
$$\begin{pmatrix} x_1^3 & x_1^2 & x_1 & 1 \\ & \vdots & & \\ x_m^3 & x_m^2 & x_m & 1 \end{pmatrix} \begin{pmatrix} e \\ d \\ a \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

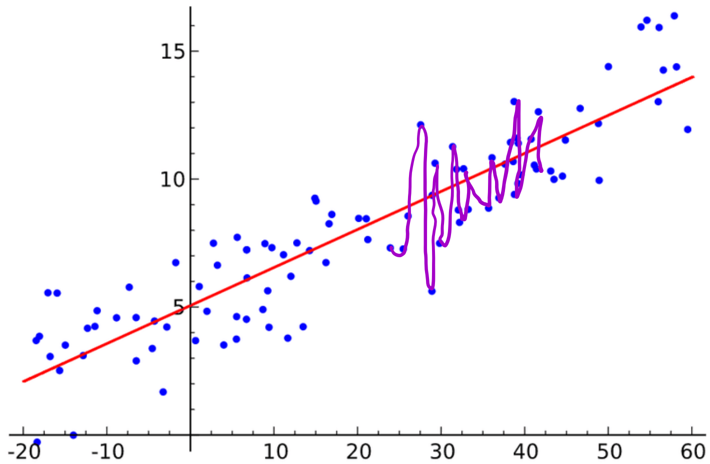
$$A \quad \vec{x} \quad = \quad \vec{b}$$

$\boxed{\text{Remark}}$: For the same data, the best fitting error using quadratics is smaller than the best fitting error using lines.

For m points, we can get a perfect fit using polynomial of degree m-1, but that is overfitting, which should be avoided.

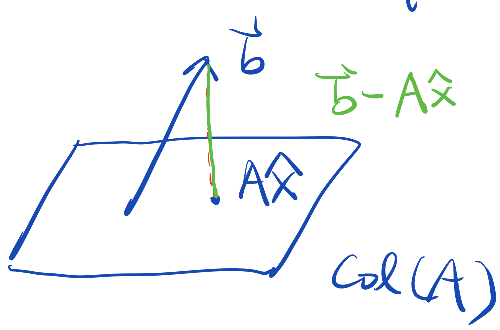
Overfitting means too many terms are used ($n > \sqrt{m}$) and the coefficients from too many terms represent the noise rather than the true model of data





$m = 100$

- Now consider how to find projection of $\vec{b} \in \mathbb{R}^m$ onto the col space of $A \in \mathbb{R}^{m \times n}$.



Want to find $\hat{x} = \begin{pmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{pmatrix}$ s.t.

$$\|A\hat{x} - \vec{b}\| = \min_{\vec{x} \in \mathbb{R}^n} \|A\vec{x} - \vec{b}\|$$

Intuition : $\vec{b} - A\hat{x}$ is $90°$ to the plane $\mathrm{Col}(A)$.

Let cols of A be $\underbrace{\vec{a}_1, \cdots, \vec{a}_n}_{\text{Assume they are independent.}}$

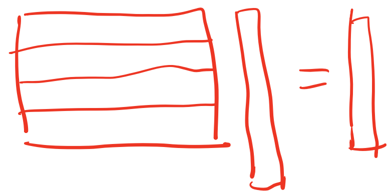
$$\underline{(\vec{b} - A\hat{x}) \perp \mathrm{Col}(A)} \iff (\vec{b} - A\hat{x}) \perp \vec{a}_i, \ \forall i$$

$\langle \vec{b} - A\hat{x}, \vec{a}_i \rangle = 0$

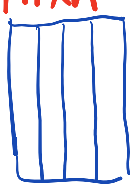
$$\iff \vec{a}_i^{\,T} (\vec{b} - A\hat{x}) = 0, \ \forall i$$

$$\iff \begin{cases} \vec{a}_1^{\,T} (\vec{b} - A\hat{x}) = 0 \\ \quad \vdots \\ \vec{a}_n^{\,T} (\vec{b} - A\hat{x}) = 0 \end{cases}$$

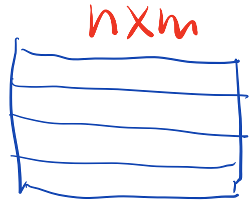
$A^T (\vec{b} - A\hat{x}) = \vec{0}$



$A =$



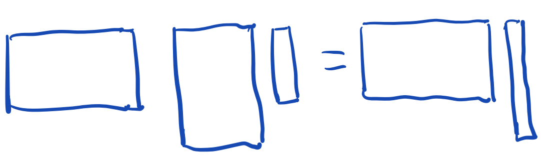
$A^T =$



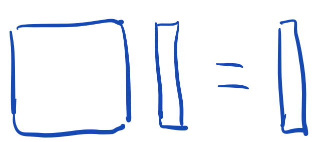
So $A^T (\vec{b} - A\hat{x}) = \vec{0}$

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

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



$$\underbrace{A^T A}_{n \times n} \hat{x} = \underbrace{A^T \vec{b}}_{n \times 1}$$



$n \times m \quad m \times n \qquad n \times m \quad m \times 1 \qquad n \times n \qquad n \times 1$

How to find projection of $\vec{b}$ onto $Col(A)$:

① $\hat{x} = (A^TA)^{-1}A^T\vec{b}$

② The projection of $\vec{b}$ is $A\hat{x} = A(A^TA)^{-1}A^T\vec{b}$

③ Least Square Solution to $A\vec{x} = \vec{b}$ is $\hat{x}$.

If we assume cols of A are independent,

then A^TA is invertible

Proof: Want to show $\underline{(A^TA)\vec{x} = \vec{0}}$ has only zero sol

$$A^T\underbrace{A\vec{x}}_{\vec{y}} = \vec{0} \iff \vec{y} = A\vec{x} \text{ is a sol to } A^T\vec{y} = \vec{0}$$

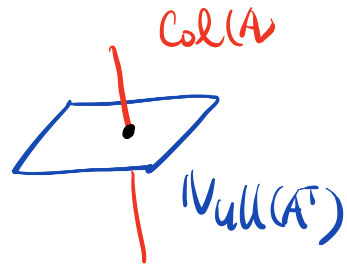
$$\iff A\vec{x} \in Null(A^T)$$

Since $A\vec{x} \in Col(A)$, we know

$$A\vec{x} \in Col(A) \cap Null(A^T)$$

$Col(A) \perp Null(A^T) \Rightarrow$

$$A\vec{x} \perp A\vec{x}$$



$$\Rightarrow \langle A\vec{x}, A\vec{x} \rangle = 0$$

$$\Rightarrow A\vec{x} = \vec{0}$$

$\left.\begin{array}{l}\text{Independence} \\ \text{of Cols of } A\end{array}\right\} \Rightarrow \vec{x} = \vec{0}$. $\square$

Review

① Column Space of A : $Col(A) \subseteq \mathbb{R}^m$

② Row Space of A : $Col(A^T) \subseteq \mathbb{R}^n$

③ Null Space of A : $Null(A) \subseteq \mathbb{R}^n$

④ Left Null Space of A : $Null(A^T) \subseteq \mathbb{R}^m$

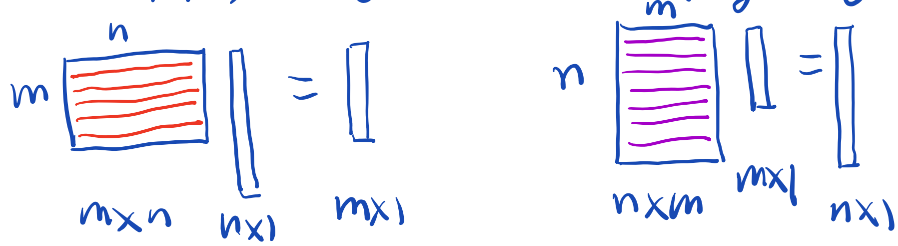
$$Null(A) = \{ \vec{x} \in \mathbb{R}^n : A\vec{x} = \vec{0} \}$$

$$Null(A^T) = \{ \vec{y} \in \mathbb{R}^m : A^T\vec{y} = \vec{0} \}$$

$Col(A^T) \perp Null(A)$

$Col(A) \perp Null(A^T)$

$A \vec{x} = \vec{0}$ $\qquad\qquad$ $A^T \vec{y} = \vec{0}$



If $A\vec{x} = \vec{0}$, then each row of A is $\perp$ to $\vec{x}$

$\Rightarrow$ $Col(A^T) \perp Null(A)$

If $A^T\vec{y} = \vec{0}$, then each row of A^T is $\perp$ to $\vec{y}$

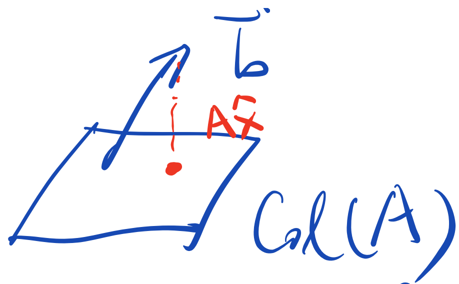
$\Rightarrow$ each col of A is $\perp$ to $\vec{y}$

$\Rightarrow$ $Col(A) \perp Null(A^T)$

Example: Find projection of $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ onto

$Span\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \right\}$

Sol: $A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ $\qquad \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$



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

$$A^TA = \begin{bmatrix} 2 & 4 \\ 4 & 9 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

$$A^TA\hat{x} = A^T\vec{b} \Leftarrow A\hat{x} = \vec{b}$$

$$\left[\begin{array}{cc|c} 2 & 4 & 3 \\ 4 & 9 & 6 \end{array}\right]$$

$$\to \left[\begin{array}{cc|c} 1 & 2 & 3/2 \\ 1 & 9/4 & 3/2 \end{array}\right]$$

$$\to \left[\begin{array}{cc|c} 1 & 2 & 3/2 \\ 0 & 1/4 & 0 \end{array}\right]$$

$$\to \left[\begin{array}{cc|c} 1 & 0 & 3/2 \\ 0 & 1 & 0 \end{array}\right]$$

$$\Rightarrow \hat{x} = \begin{pmatrix} 3/2 \\ 0 \end{pmatrix}$$

$$A\hat{x} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3/2 \\ 0 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 3/2 \\ 0 \end{bmatrix}$$

So projection of $\begin{bmatrix}1\\2\\3\end{bmatrix}$ onto

$\text{Span}\left\{\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}2\\2\\1\end{bmatrix}\right\}$ is $\begin{bmatrix}3/2\\3/2\\0\end{bmatrix}$.

Ex: Given (x_i, y_i) $i=1,2,3$

Want to find best line fit

$$y = ax + c.$$

Solution: Let $\hat{x} = \begin{pmatrix}a\\c\end{pmatrix}$

① $\begin{cases} ax_1 + c = y_1 \\ ax_2 + c = y_2 \\ ax_3 + c = y_3 \end{cases}$ $\begin{bmatrix}x_1 & 1\\x_2 & 1\\x_3 & 1\end{bmatrix}\begin{bmatrix}a\\c\end{bmatrix} = \begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$

$$A\hat{x} = \vec{b}$$

② Find projection of $\vec{b}$ onto

$$\text{Col}(A) = \text{Span}\left\{\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$$

Compute A^TA and A^Tb

Solve $\underbrace{\boxed{A^TA}}_{2\times 2}\hat{x} = \underbrace{A^Tb}_{2\times 1}$

1) Projection of $\vec{b}$ is $A\hat{x}$

2) Line fit is $y = ax + c$

Ex: Given data points $\begin{cases} x_1 = -1 \\ y_1 = 0 \end{cases}$, $\begin{cases} x_2 = 0 \\ y_2 = 1 \end{cases}$, $\begin{cases} x_3 = 1 \\ y_3 = -1 \end{cases}$

Find best line fit $y = ax + c$.

Sol: ① $\begin{cases} ax_1 + c = y_1 \\ ax_2 + c = y_2 \\ ax_3 + c = y_3 \end{cases}$ $\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$

$A\hat{x} = \vec{b}$

$\hat{x} = \begin{pmatrix} a \\ c \end{pmatrix}$

$$\begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

② $A\hat{x} = \vec{b}$

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

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

$$A^T \vec{b} = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$A^T A\hat{x} = A^T \vec{b} \Rightarrow \left[\begin{array}{cc|c} 2 & 0 & -1 \\ 0 & 3 & 0 \end{array}\right]$$

$$\Rightarrow \hat{x} = \begin{pmatrix} -\frac{1}{2} \\ 0 \end{pmatrix}$$

The best line fit is $y = ax + c$

$= -\frac{1}{2}x$.