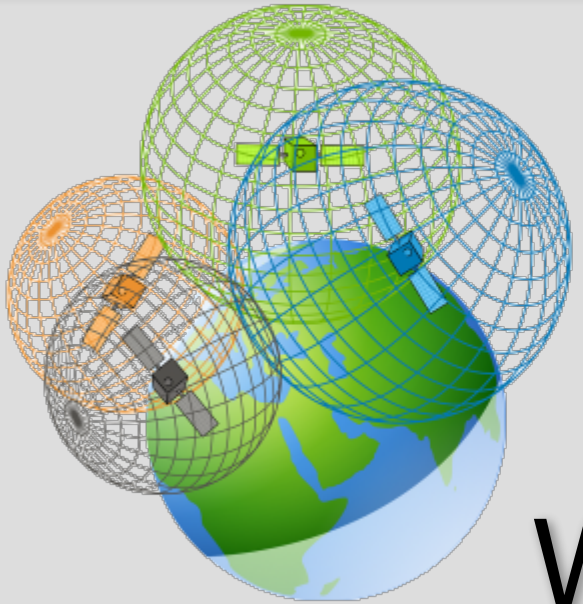
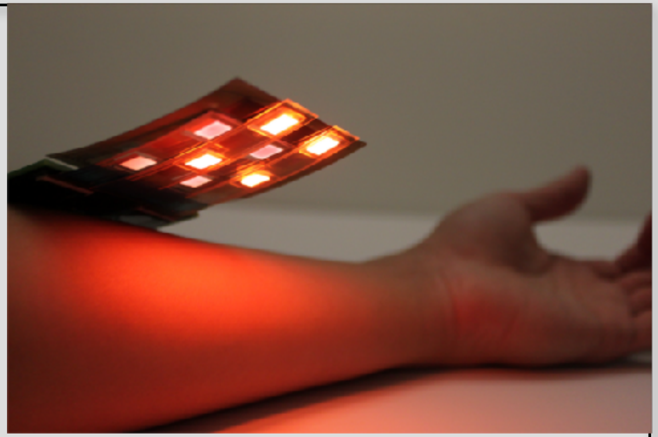




Welcome to EECS 16A!

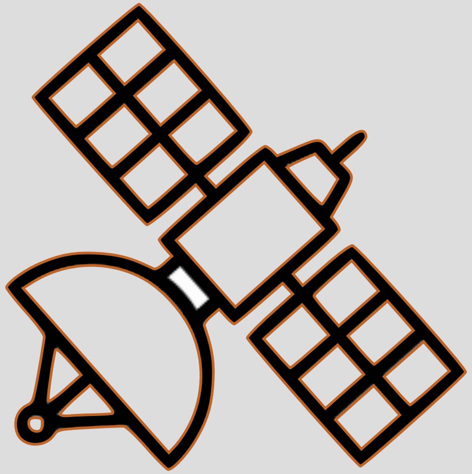
Designing Information Devices and Systems I

Ana Arias and Miki Lustig

Fall 2022

Lecture 13A
Least Squares

GPS



- 24 satellites
    - Known position
    - Time synchronized
    - 8 usually visible
- Problem:
    - Classify which satellite is transmitting
    - Estimate distance to GPS
    - Estimate position from noisy data
- Tools:
    - Inner product
    - Cross correlation
    - Least Squares

Orthogonal Projections

Given vectors $\overrightarrow{a}$, $\overrightarrow{b}$, we say that the orthogonal projection of $\overrightarrow{b}$ onto $\overrightarrow{a}$ is:

$$\mathrm{Proj}_{\overrightarrow{a}}(\overrightarrow{b}) = \frac{\overrightarrow{a}^T \overrightarrow{b}}{\| \overrightarrow{a} \|^2} \overrightarrow{a}$$

Example 2D

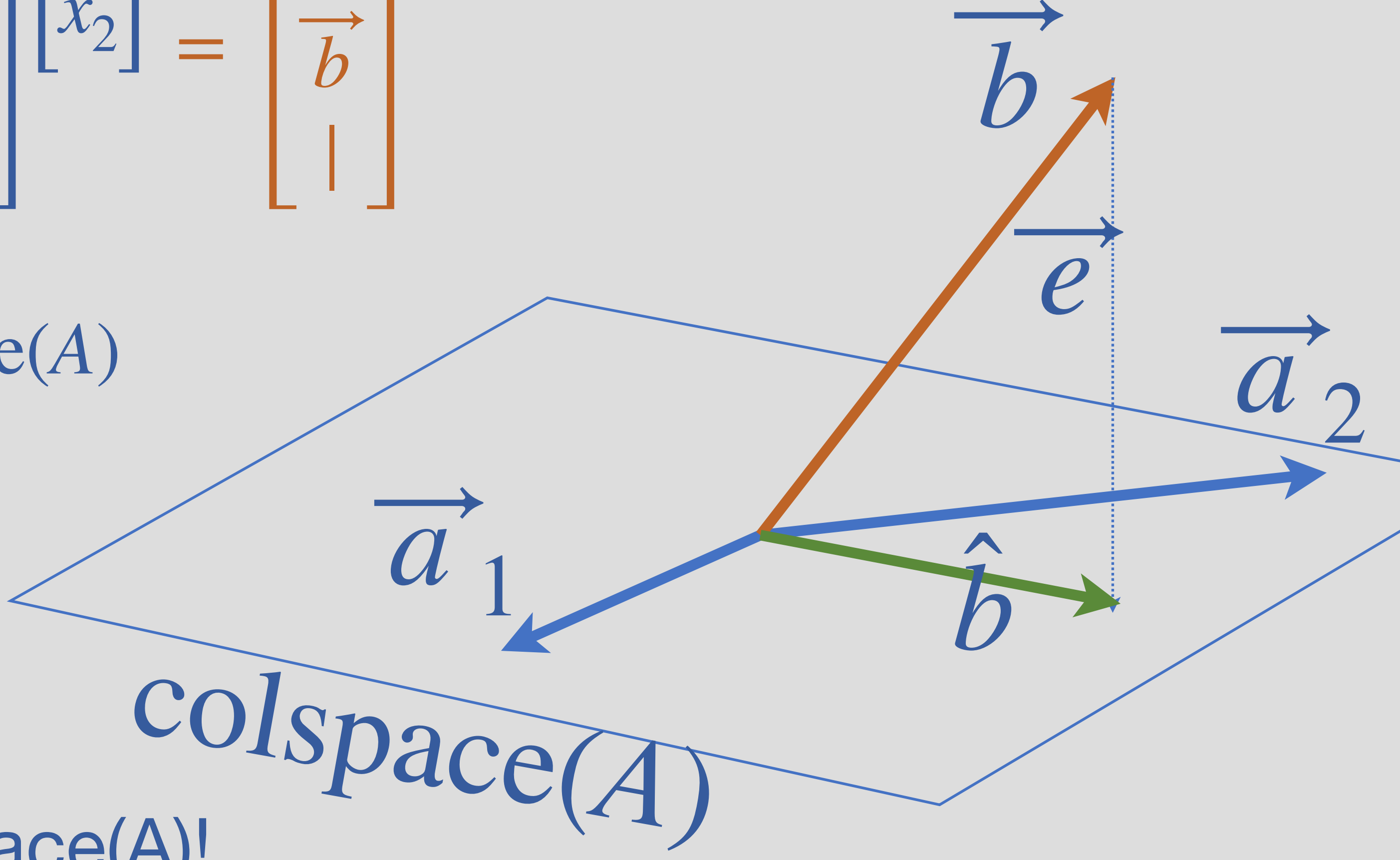
3 equations 2 unknowns:

$$\overset{A}{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}} \overset{\vec{x}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} = \overset{\vec{b}}{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}} \qquad \begin{bmatrix} | & | \\ \vec{a}_1 & \vec{a}_2 \\ | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} | \\ \vec{b} \\ | \end{bmatrix}$$

No solution means: $\vec{b} \notin \text{colspace}(A)$

Find $\hat{x}$ that has the smallest error

$$\|\vec{e}\| = \|A\hat{x} - \vec{b}\| \leq \|Ax - \vec{b}\|$$



Orthogonal projection onto colspace(A)!

Theorem: Consider matrix A, and $\vec{y} \in \mathrm{colspace}(A)$

If $\exists \vec{z}$, such that $\langle \vec{z}, \vec{a}_i \rangle = 0$, then $\langle \vec{z}, \vec{y} \rangle = 0$.

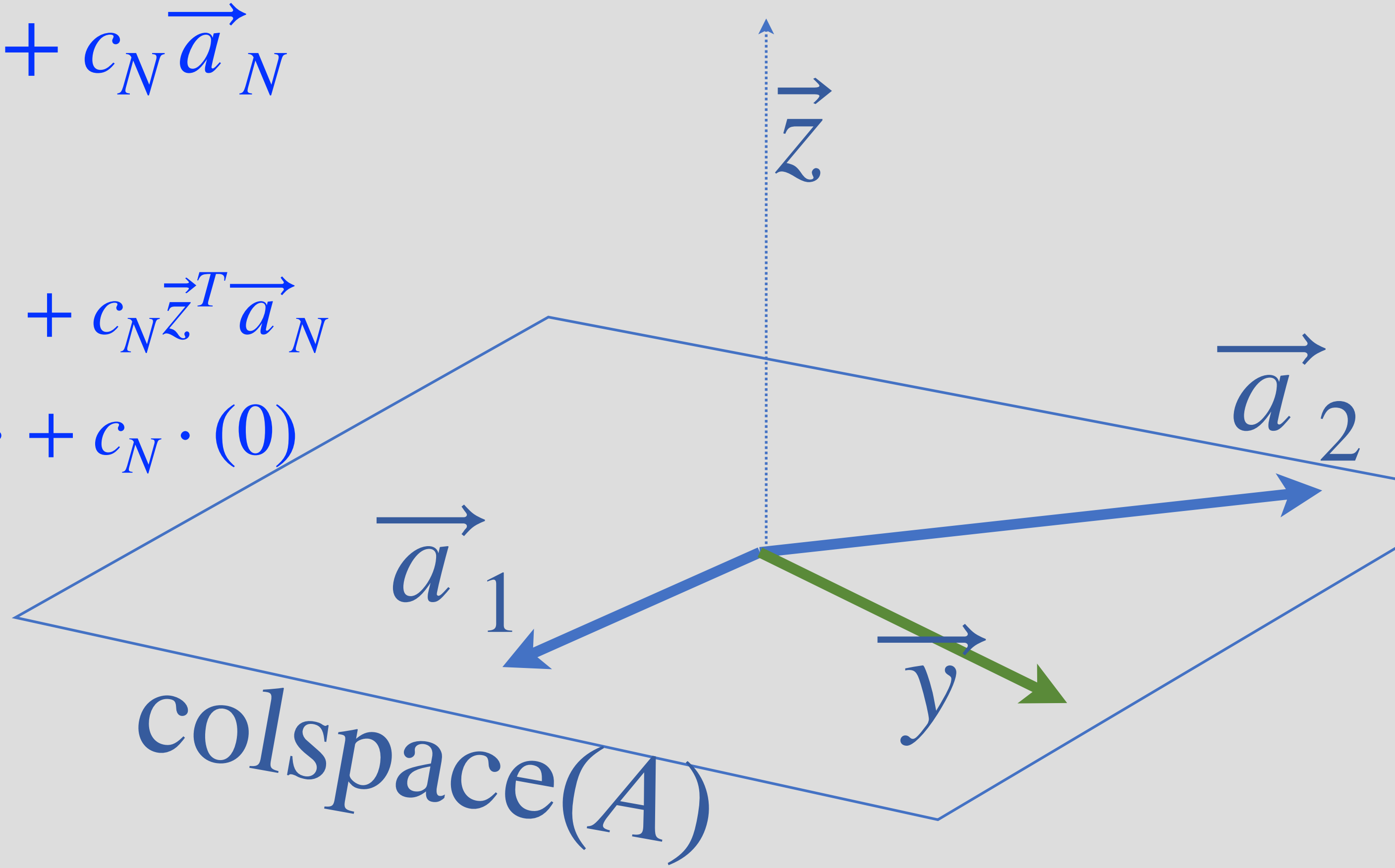
$$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_N \\ | & | & & | \end{bmatrix}$$

Proof:

Know: $\vec{y} = c_1 \vec{a}_1 + c_2 \vec{a}_2 + \cdots + c_N \vec{a}_N$

Show: $\langle \vec{z}, \vec{y} \rangle = 0$

$$\begin{aligned} \langle \vec{z}, c_1 \vec{a}_1 + \cdots + c_N \vec{a}_N \rangle &= c_1 \vec{z}^T \vec{a}_1 + \cdots + c_N \vec{z}^T \vec{a}_N \\ &= c_1 \cdot (0) + \cdots + c_N \cdot (0) \\ &= 0 \end{aligned}$$



Least Squares

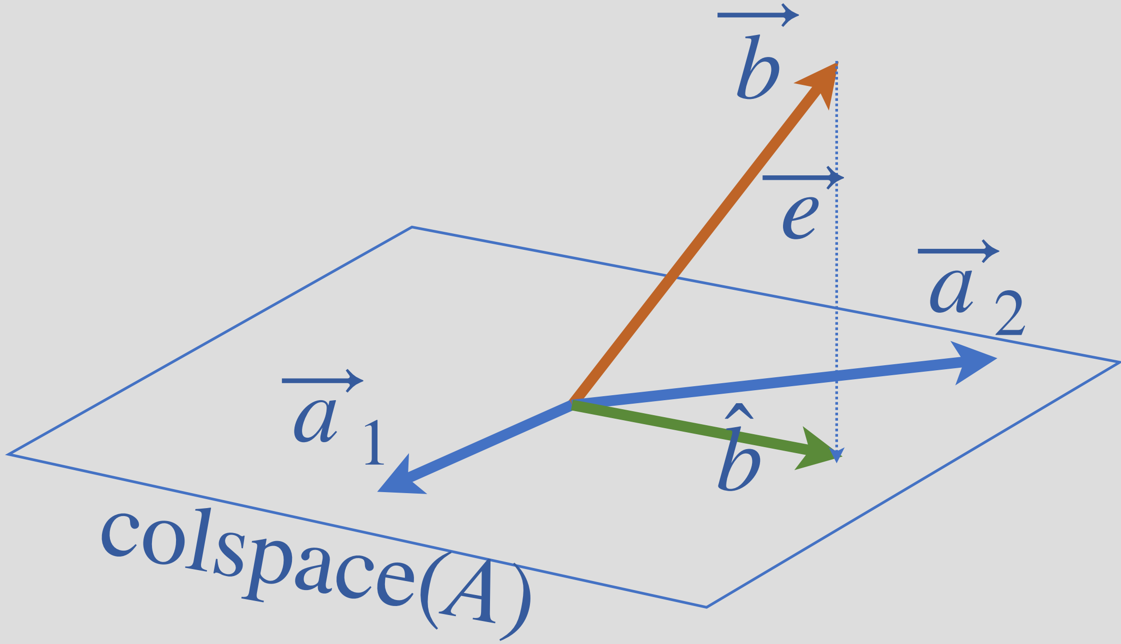
$$\mathrm{argmin}_{\vec{x}} \ \|\vec{e}\| = \|A\vec{x} - \vec{b}\|$$

$$\vec{e} = \vec{b} - \hat{b}$$

Since $\vec{e} \perp \mathrm{col}(A)$, $\langle \vec{a}_i, \vec{e} \rangle = 0$

$$\left\langle \vec{a}_i, \vec{b} - \hat{b} \right\rangle = 0$$

$$\vec{a}_i^T(\vec{b} - \hat{b}) = 0$$



$$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_N \\ | & | & & | \end{bmatrix}$$

$A\vec{x} \in \mathrm{colspace}(A)$

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

Least Squares

$$\mathrm{argmin}_{\vec{x}} \ \|\vec{e}\| = \|A\vec{x} - \vec{b}\|$$

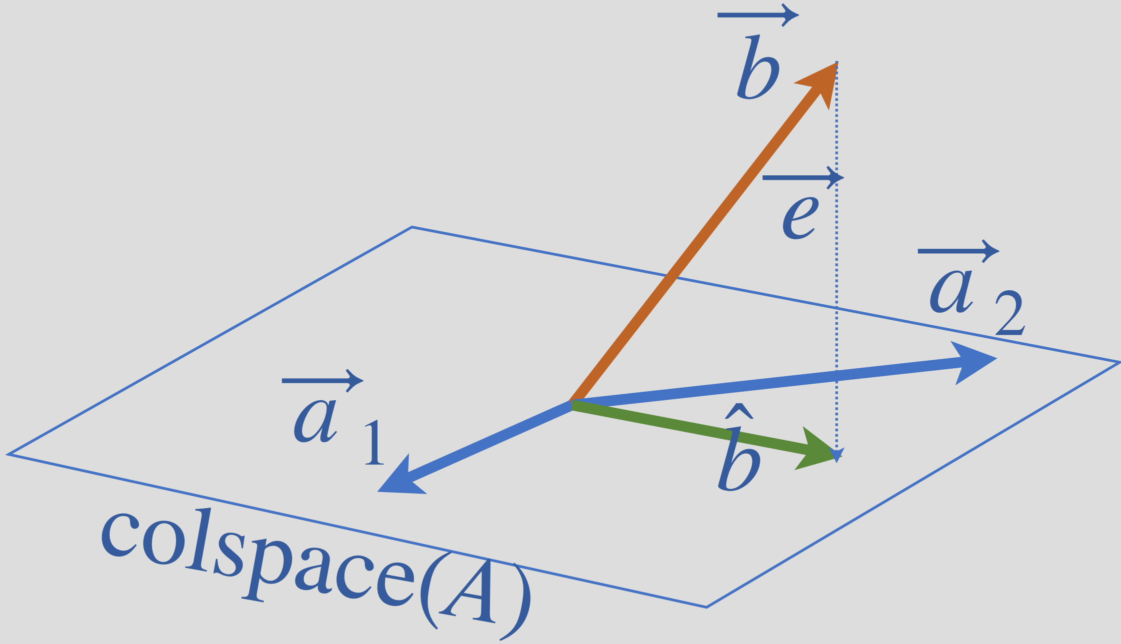
$$\vec{e} = \vec{b} - \hat{b}$$

Since $\vec{e} \perp \mathrm{col}(A)$, $\langle \vec{a}_i, \vec{e} \rangle = 0$

$$\left\langle \vec{a}_i, \vec{b} - \hat{b} \right\rangle = 0$$

$$\vec{a}_i^T(\vec{b} - \hat{b}) = 0$$

$$\begin{bmatrix} - & \vec{a}_1^T & - \\ - & \vec{a}_2^T & - \\ & \vdots & \\ - & \vec{a}_N^T & - \end{bmatrix} \begin{bmatrix} | \\ \vec{b} - \hat{b} \\ | \end{bmatrix} = 0$$



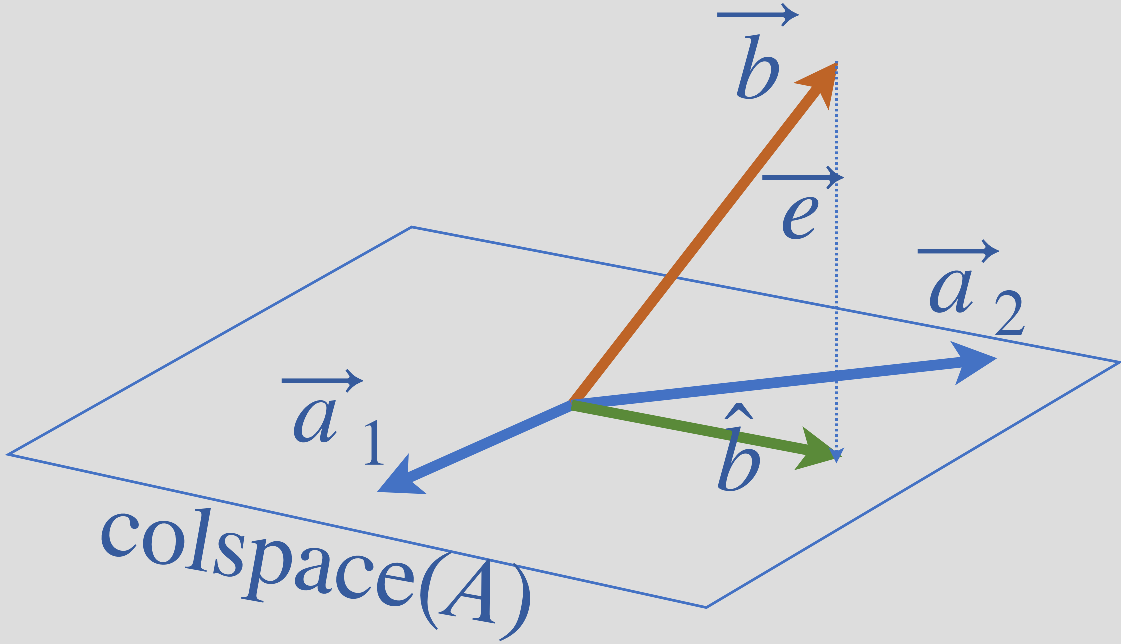
$$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_N \\ | & | & & | \end{bmatrix}$$

$A\vec{x} \in \mathrm{colspace}(A)$

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

Least Squares

$$\begin{bmatrix} - & \vec{a}_1^T & - \\ - & \vec{a}_2^T & - \\ & \vdots & \\ - & \vec{a}_N^T & - \end{bmatrix} \begin{bmatrix} | \\ \vec{b} - \hat{b} \\ | \end{bmatrix} = 0$$



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

$$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_N \\ | & | & & | \end{bmatrix}$$

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

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

$A\vec{x} \in \text{colspace}(A)$

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

If A is full Rank, then $A^T A$ is invertible

$$\boxed{\hat{x} = (A^T A)^{-1} A^T \vec{b}}$$

$$\hat{b} = A(A^T A)^{-1} A^T \vec{b}$$

⌘3+⌘4

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}$$

M

1

1

M $\vec{x}$

N A

$\approx$ N $\vec{b}$

$(A^T A)^{-1}$ A^T $\vec{b}$ = $\hat{x}$

Like this....

and like that!


15
vineyard vines
$\vec{x}=(A^TA)^{-1}A^T\vec{y}$

Example 1

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

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\left.\begin{aligned} 2x &= 1 \\ 1 \cdot x &= 1 \end{aligned}\right\} \quad \begin{bmatrix} & \Big| & \end{bmatrix}$$

Example 1

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

$$\begin{bmatrix}2\\1\end{bmatrix}\begin{bmatrix}x\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$$

$$\left.\begin{aligned} 2x &= 1 \\ 1\cdot x &= 1 \end{aligned}\right\} \quad \left[\begin{array}{c|c} 2 & 1 \\ 1 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{c|c} \; & \; \end{array}\right]$$

Example 1

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

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} [x] = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\left.\begin{aligned} 2x &= 1 \\ 1 \cdot x &= 1 \end{aligned}\right\} \quad \left[\begin{array}{c|c} 2 & 1 \\ 1 & 1 \end{array}\right] \Rightarrow \left[\begin{array}{c|c} 1 & 1/2 \\ 0 & 1/2 \end{array}\right]$$

Inconsistent! No solution

Least Squares:

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}$$

$$= \left(\frac{1}{5}\right) \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \left(\frac{1}{5}\right) \cdot 3 = \boxed{\frac{3}{5}}$$

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

$$(A^T A)^{-1} = \frac{1}{5}$$

Example 2

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

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$x_1 = 1$

$$\left.\begin{array}{l} x_2 = 2 \\ x_2 = 3 \end{array}\right\} 2.5!$$

Least Squares:

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}$$

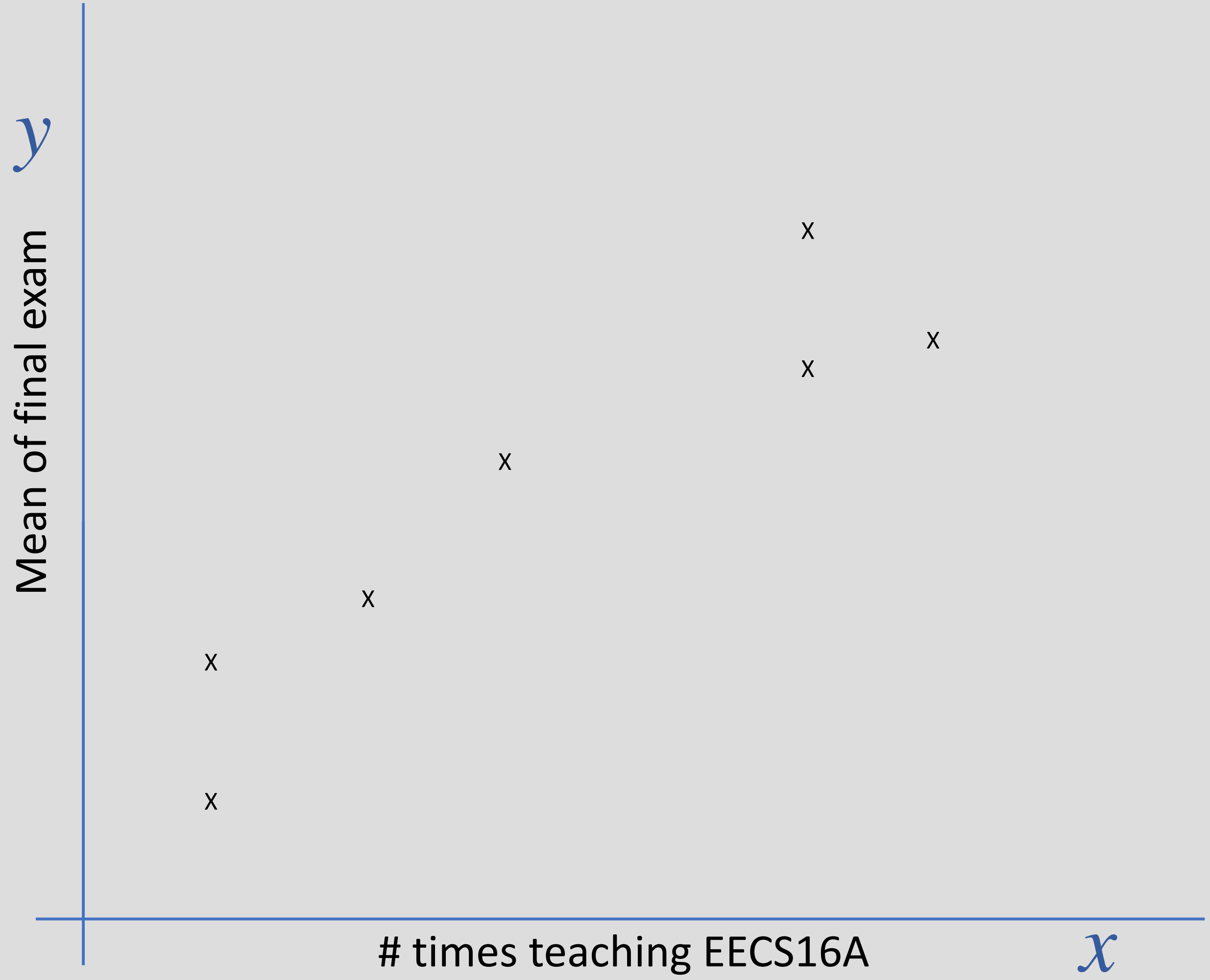
$$= \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 2.5 \end{bmatrix}$$

$$(A^T A)^{-1} = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \right)^{-1}$$

$$= \left(\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$$

Example 3: Linear Regression



Known:

Waller: (x_1, y_1)

Sahai: (x_2, y_2)

Alon: (x_4, y_4)

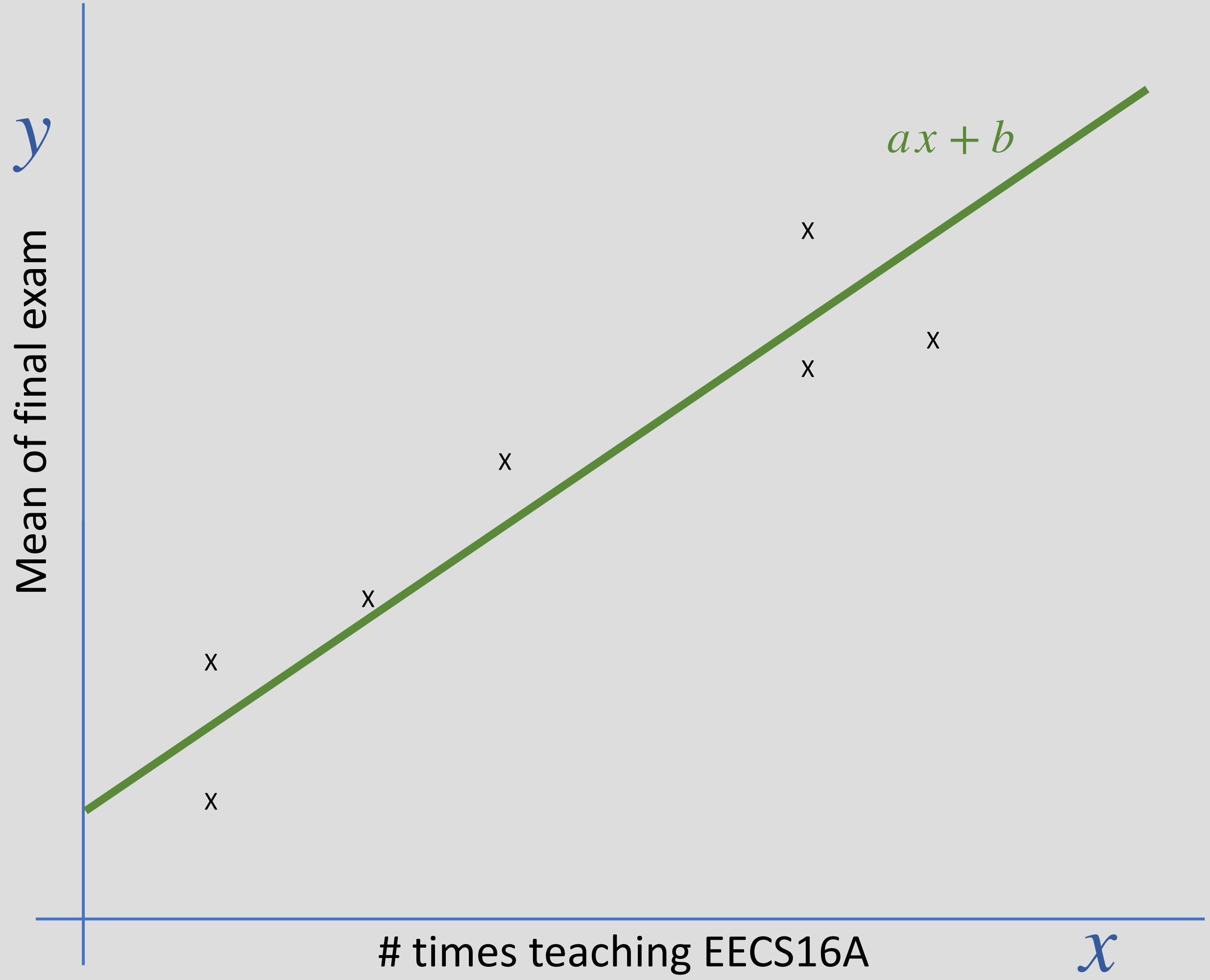
Stojanovic: (x_5, y_5)

Ranade: (x_6, y_6)

Courtade: (x_7, y_7)

Liu: (x_8, y_8)

Example 3: Linear Regression

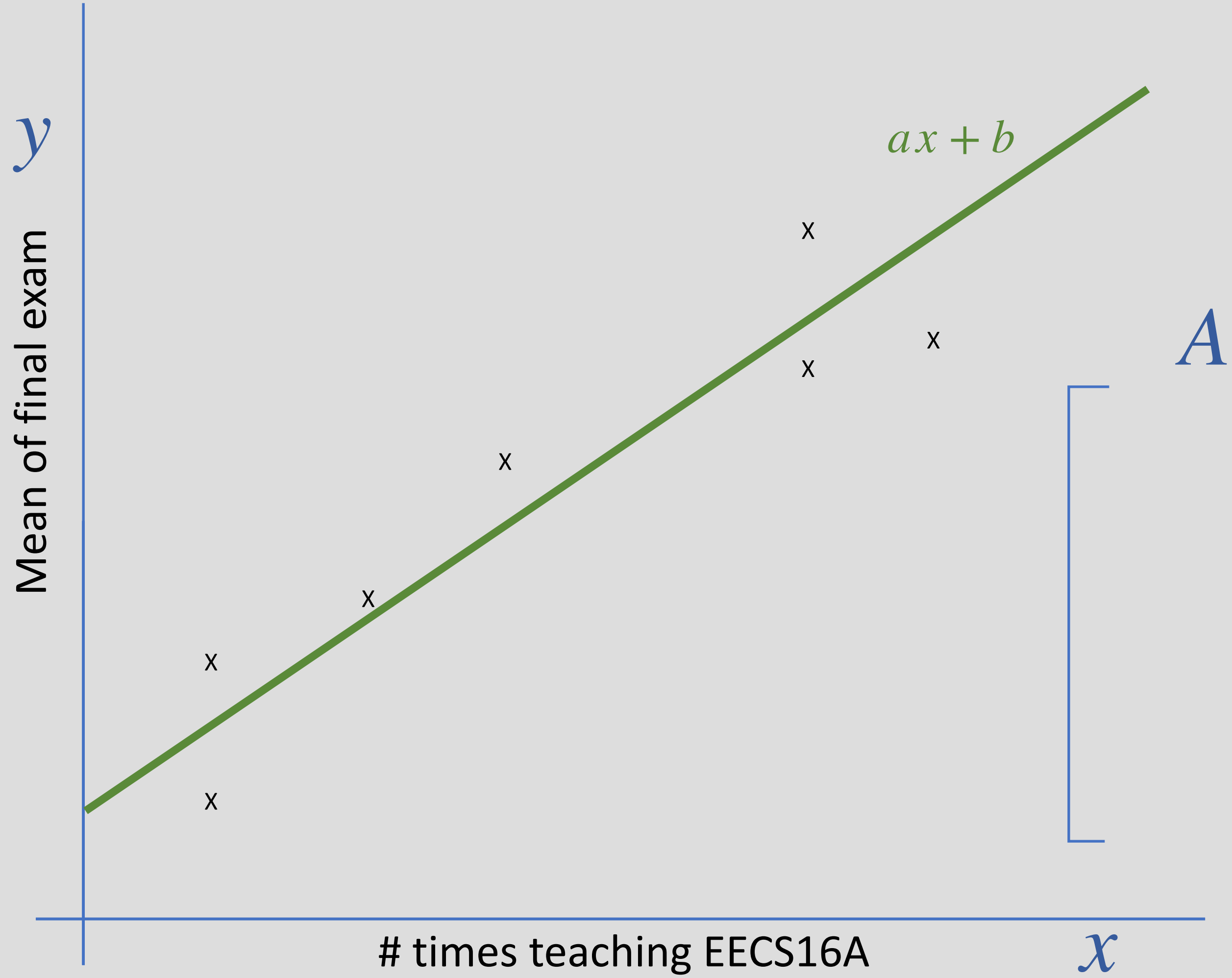


Model: $y = ax + b$

Known:

Waller: (x_1, y_1)
Sahai: (x_2, y_2)
Alon: (x_4, y_4)
Stojanovic: (x_5, y_5)
Ranade: (x_6, y_6)
Courtade: (x_7, y_7)
Liu: (x_8, y_8)

Example 3: Linear Regression



$$\underset{A}{\begin{bmatrix} \ \\ \ \\ \ \end{bmatrix}} \underset{\vec{p}}{\begin{bmatrix} \ \\ \ \end{bmatrix}} = \underset{\vec{y}}{\begin{bmatrix} \ \\ \ \\ \ \end{bmatrix}}$$

Model: $y = ax + b$

Known:

Waller: (x_1, y_1)

Sahai: (x_2, y_2)

Anant + Miki (x_3, y_3)

Alon: (x_4, y_4)

Stojanovic: (x_5, y_5)

Ranade: (x_6, y_6)

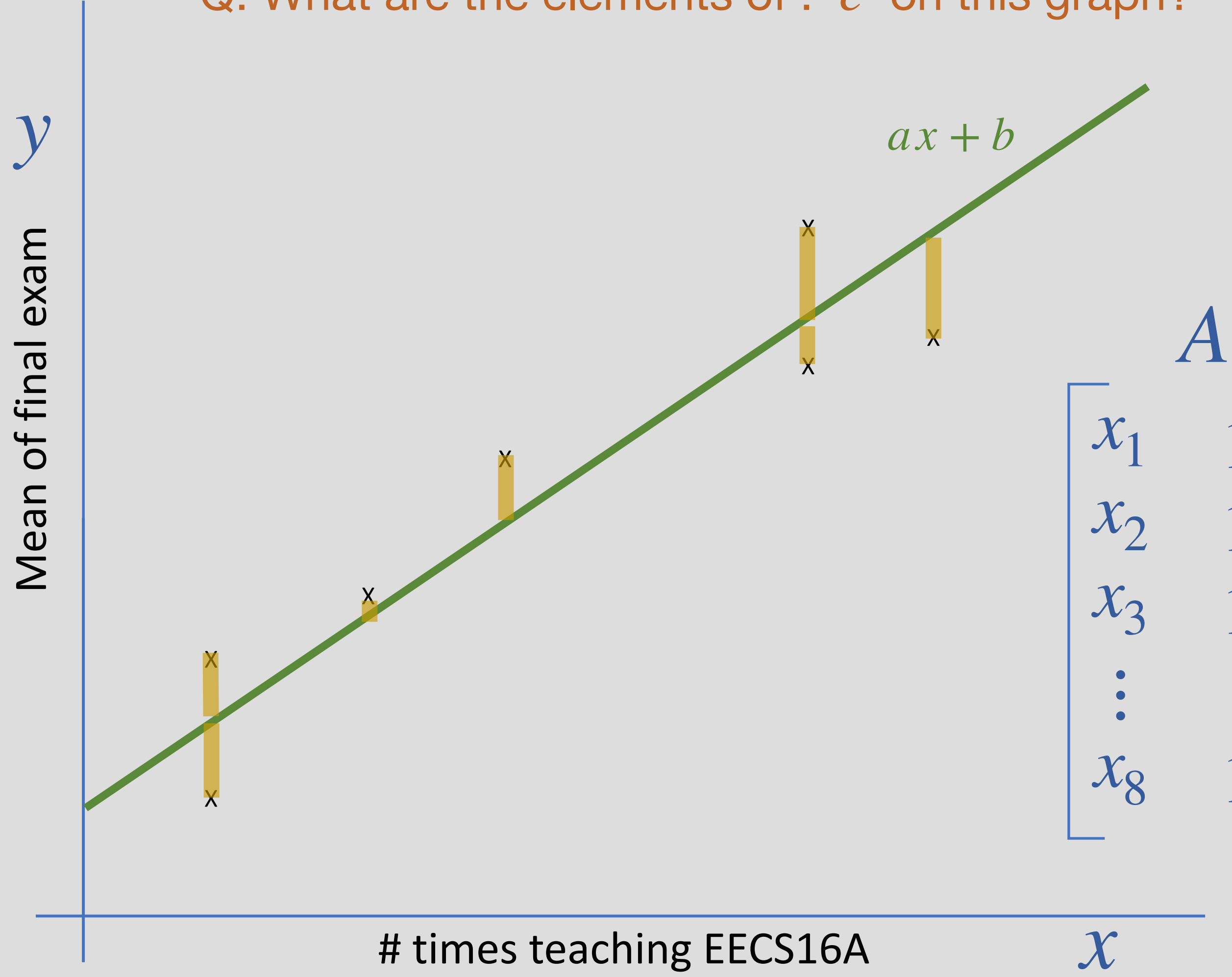
Courtade: (x_7, y_7)

Liu: (x_8, y_8)

Example 3: Linear Regression

$$\overrightarrow{e} = A\hat{p} - \overrightarrow{y}$$

Q: What are the elements of : $\overrightarrow{e}$ on this graph?



Model: $y = ax + b$

Known:

Waller: (x_1, y_1)

Sahai: (x_2, y_2)

Alon: (x_4, y_4)

Stojanovic: (x_5, y_5)

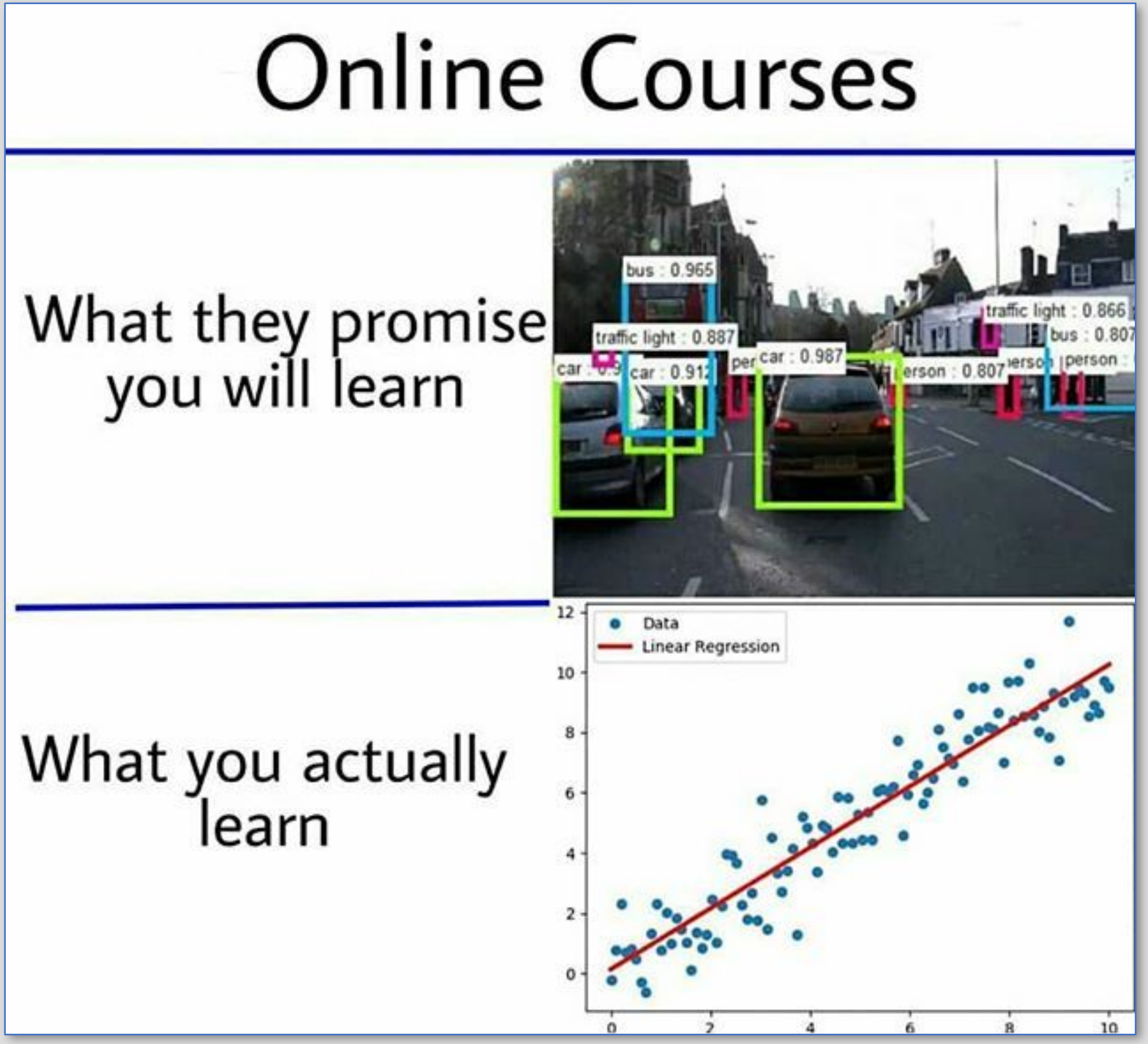
Ranade: (x_6, y_6)

Courtade: (x_7, y_7)

Liu: (x_8, y_8)

$$\underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_8 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\overrightarrow{p}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_8 \end{bmatrix}}_{\overrightarrow{y}}$$

$$\hat{p} = (A^T A)^{-1} A^T \overrightarrow{y}$$



BUT, not everything fits to a line!?!

Example 4: Regression

Gauss found Ceres by using Kepler's laws:

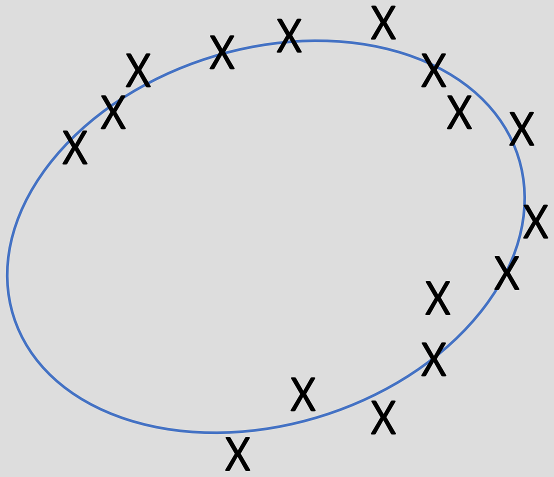
Model: $ax^2+by^2+cxy+dx+ey=1$

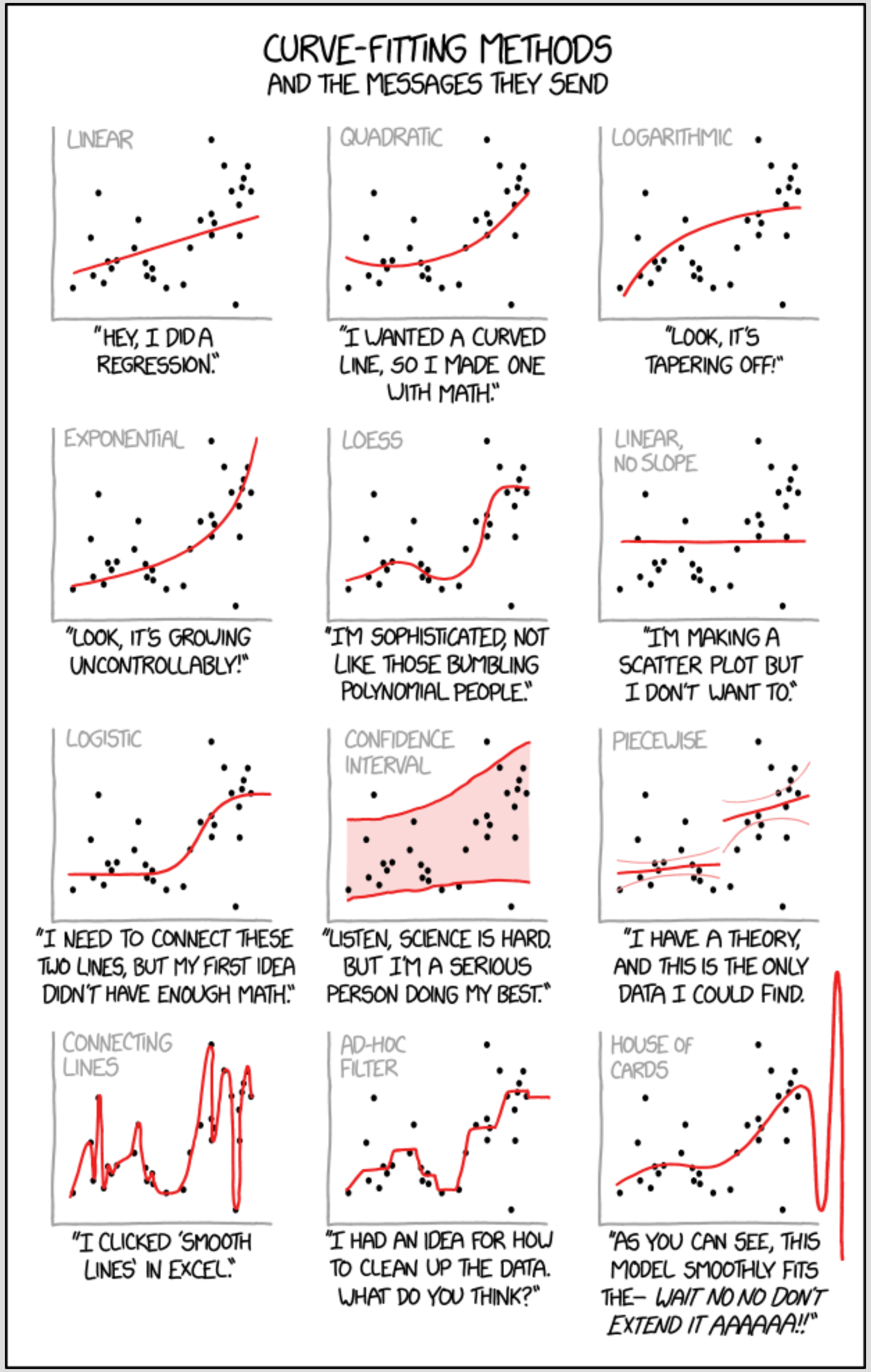
Q: Is this a linear fit?

A: Yes!

Knowns: $(x_1, y_1)\ (x_2, y_2)\ \cdots (x_N, y_N)$

Unknowns: $\vec{p} = [a\ b\ c\ d\ e]^T$





Example 4: Regression

Gauss found Ceres by using Kepler's laws:

Model: $ax^2+by^2+cxy+dx+ey = 1$

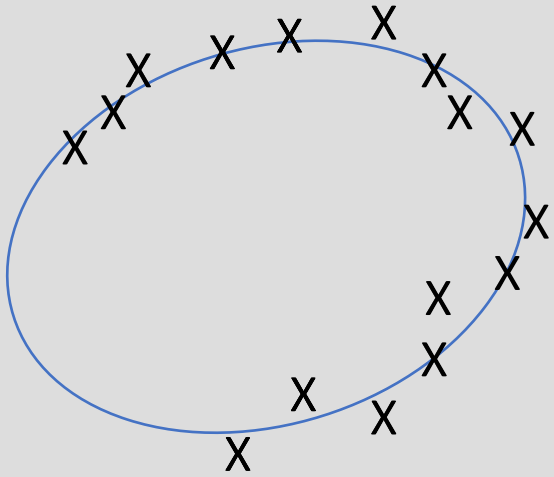
Q: Is this a linear fit?

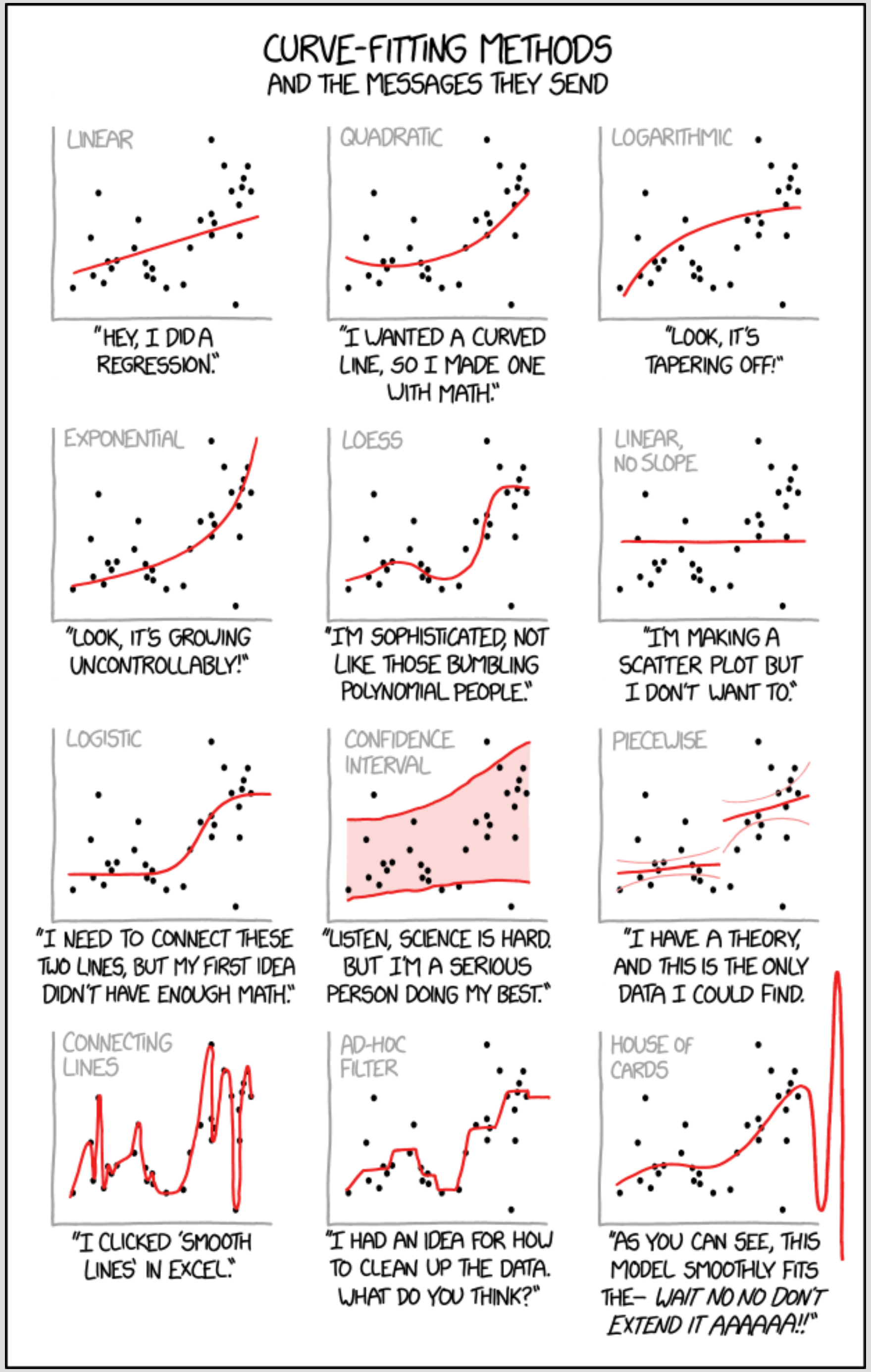
A: Yes!

Knowns: $(x_1, y_1)\ (x_2, y_2)\ \cdots (x_N, y_N)$

Unknowns: $\vec{p} = [a\ b\ c\ d\ e]^T$

$$A \, \vec{p} = \vec{y}$$





Example 4: Regression

Gauss found Ceres by using Kepler's laws:

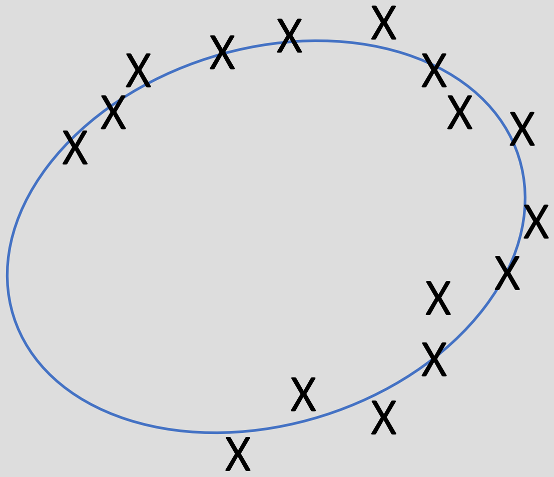
Model: $ax^2+by^2+cxy+dx+ey = 1$

Q: Is this a linear fit?

A: Yes!

Knowns: $(x_1, y_1)\ (x_2, y_2)\ \cdots (x_N, y_N)$

Unknowns: $\vec{p} = [a\ b\ c\ d\ e]^T$



$$\begin{array}{ccc} A & \vec{p} & & \vec{y} \\ \begin{bmatrix} x_1^2 & y_1^2 & x_1y_1 & x_1 & y_1 \\ x_2^2 & y_2^2 & x_2y_2 & x_2 & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N^2 & y_N^2 & x_Ny_N & x_N & y_N \end{bmatrix} & \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} & = & \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \end{array}$$

$$\hat{p} = (A^TA)^{-1}A^T\vec{y}$$

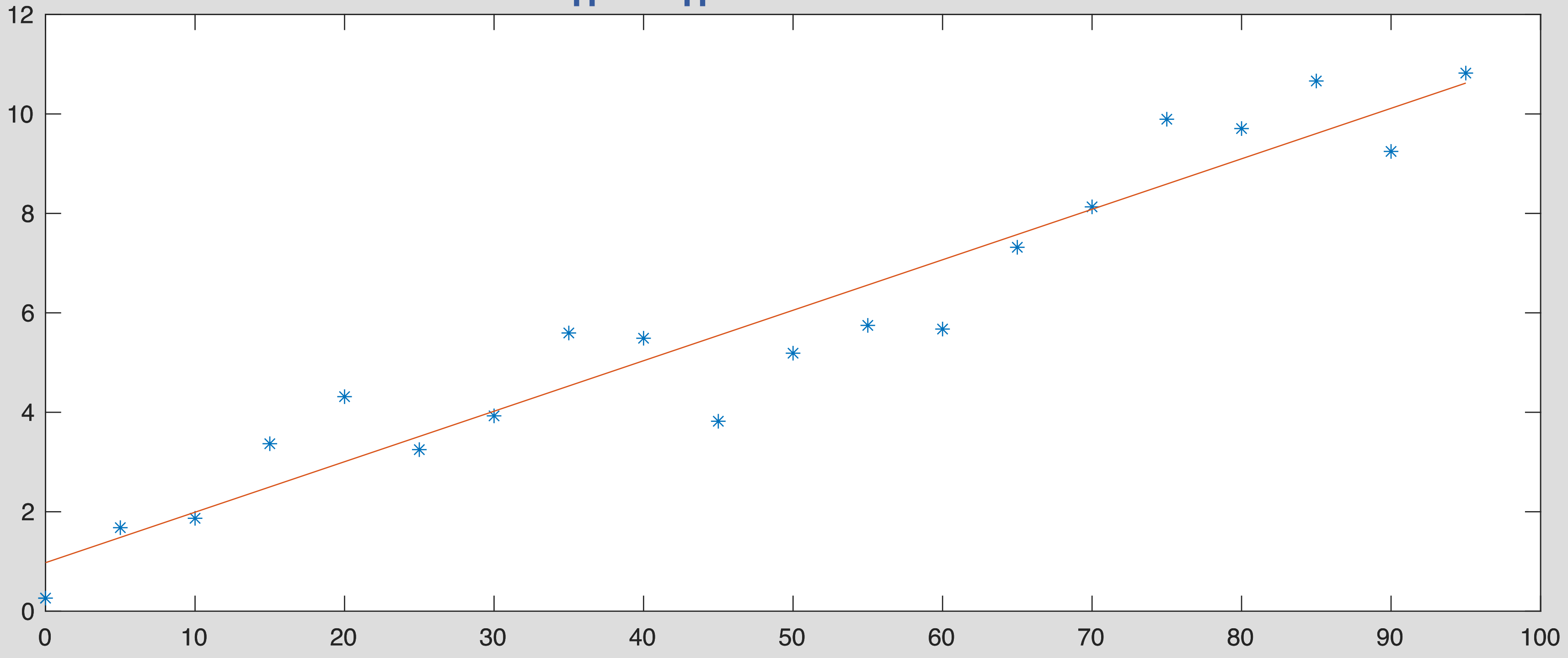
Example 5: Over Fitting

- Consider noisy measurements of $y = 0.1x + 1$:

$$\vec{p} = [0.1015 \quad 0.9757]^T$$

Model: $y = ax + b$

$$\|\vec{e}\| = 3.85$$

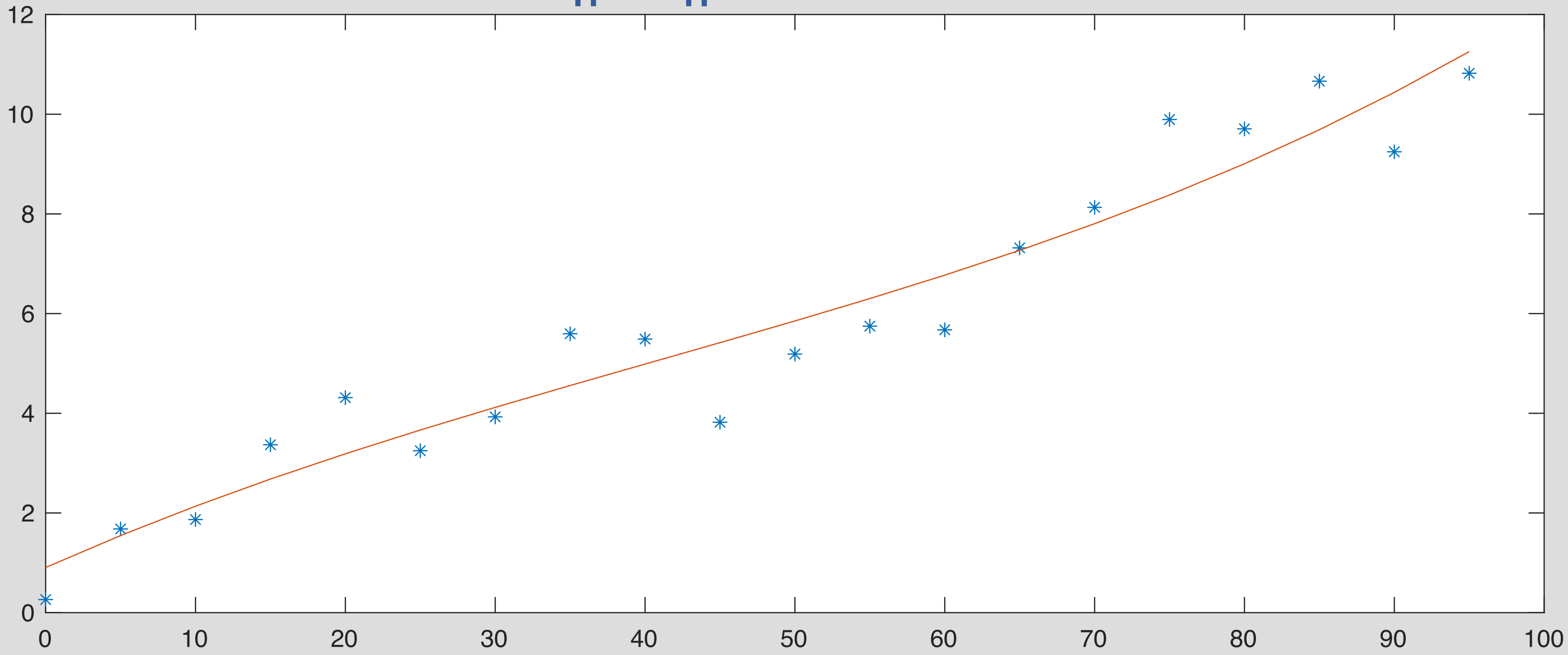


Example 5: Over Fitting

- Consider noisy measurements of $y = 0.1x + 1$:

Model: $y = ax^3 + bx^2 + cx + d$

$$\|\vec{e}\| = 3.71$$

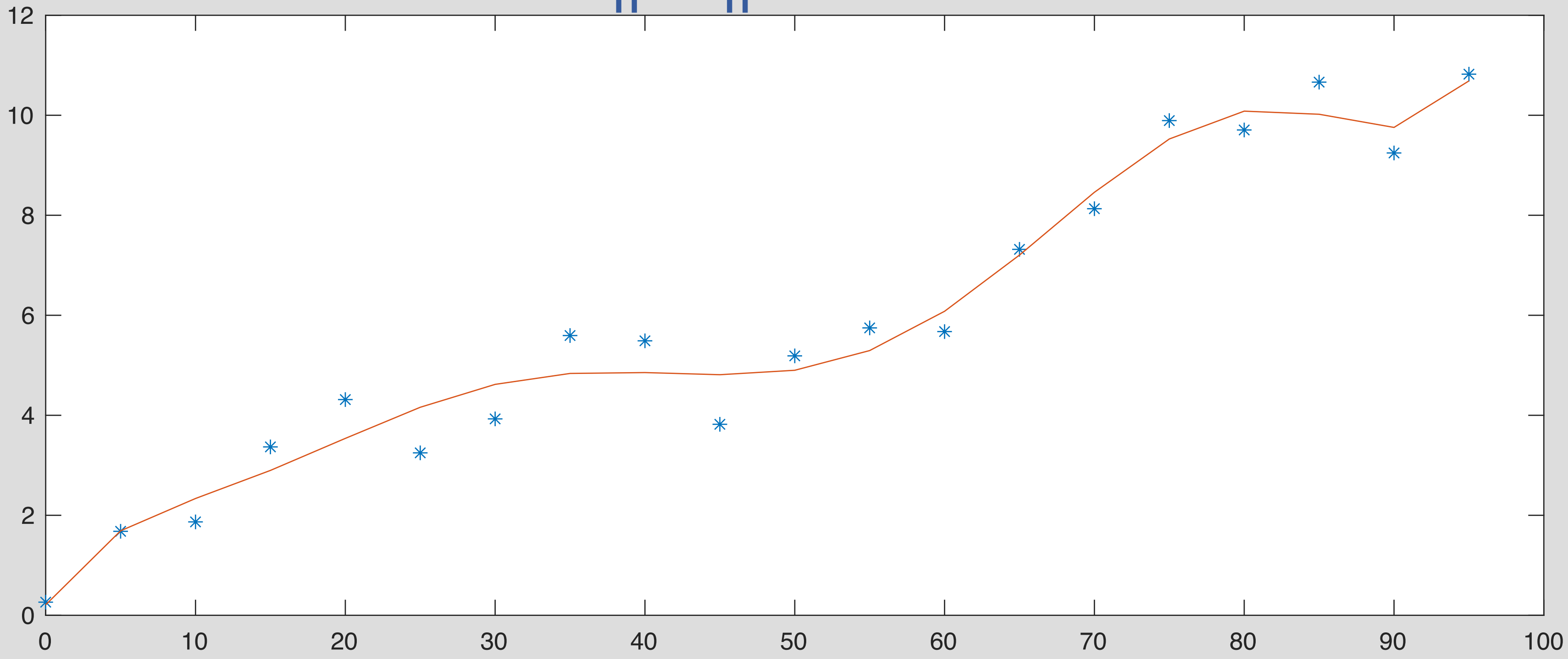


Example 5: Over Fitting

- Consider noisy measurements of $y = 0.1x + 1$:

Model: $y = ax^7 + bx^6 + cx^5 + dx^4 + ex^3 + fx^2 + gx + h$

$$\|\vec{e}\| = 2.42$$



Example 5: Exponential Regression

Model: $y = ce^{ax}$

Q: Is this a linear fit?

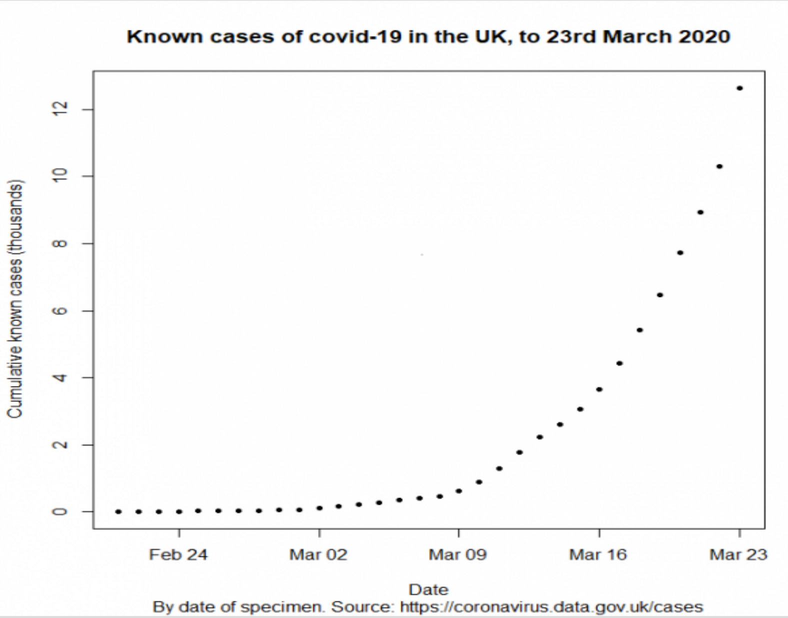
A: No! But, can be made linear…..

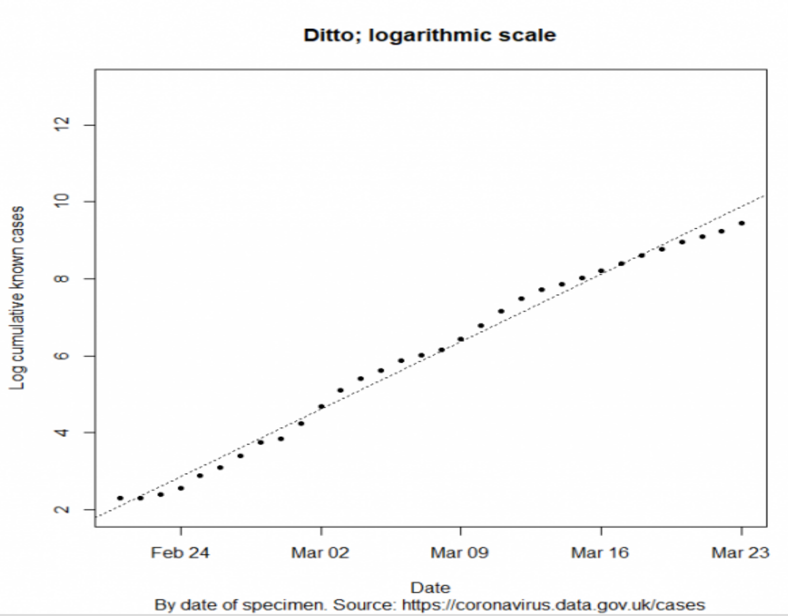
New Model: $\log(y) = \log c + ax = b + ax$

Knowns: $(x_1, \log(y_1))\ \ (x_2, \log(y_2))\ \cdots (x_N, \log(y_N))$

Unknowns: $\vec{p} = [a\ b]^T$

$A \qquad \vec{p} \qquad \vec{y}$





https://www.cebm.net/covid-19/exponential-growth-what-it-is-why-it-matters-and-how-to-spot-it/

Example 5: Exponential Regression

Model: $y = ce^{ax}$

Q: Is this a linear fit?

A: No! But, can be made linear…..

New Model: $\log(y) = \log c + ax = b + ax$

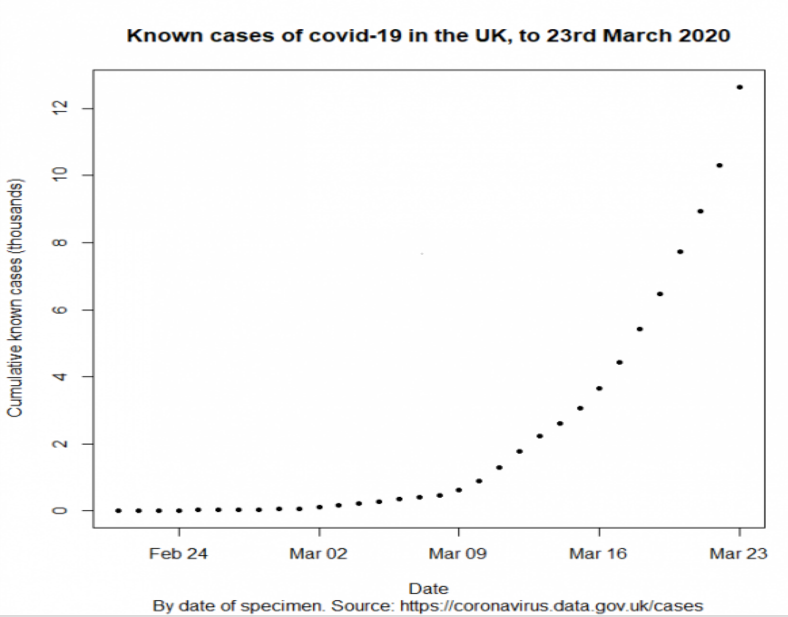
Knowns: $(x_1, \log(y_1))\ (x_2, \log(y_2)) \cdots (x_N, \log(y_N))$

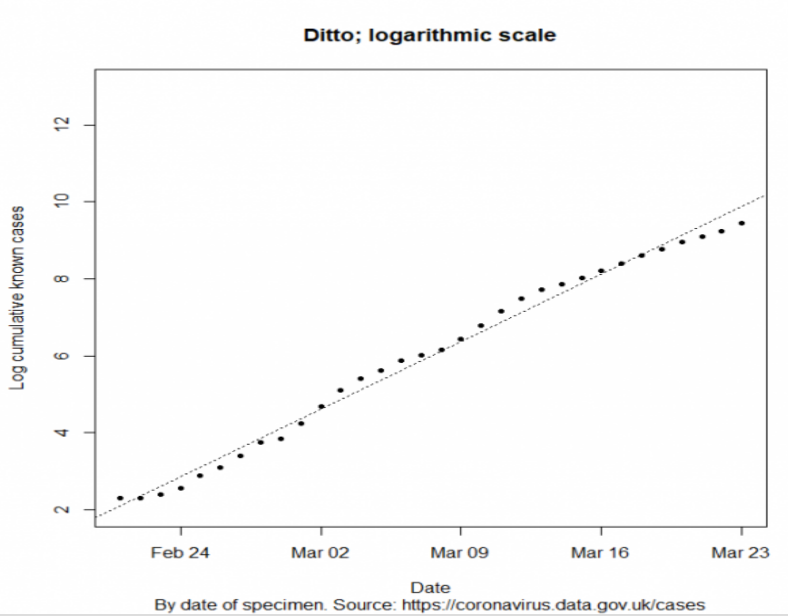
Unknowns: $\vec{p} = [a\ b]^T$

$$\underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\vec{p}} = \underbrace{\begin{bmatrix} \log y_1 \\ \log y_2 \\ \vdots \\ \log y_N \end{bmatrix}}_{\vec{y}}$$

$$\boxed{\hat{p} = (A^T A)^{-1} A^T \vec{y}}$$

$$\boxed{\hat{c} = e^{\hat{b}}}$$





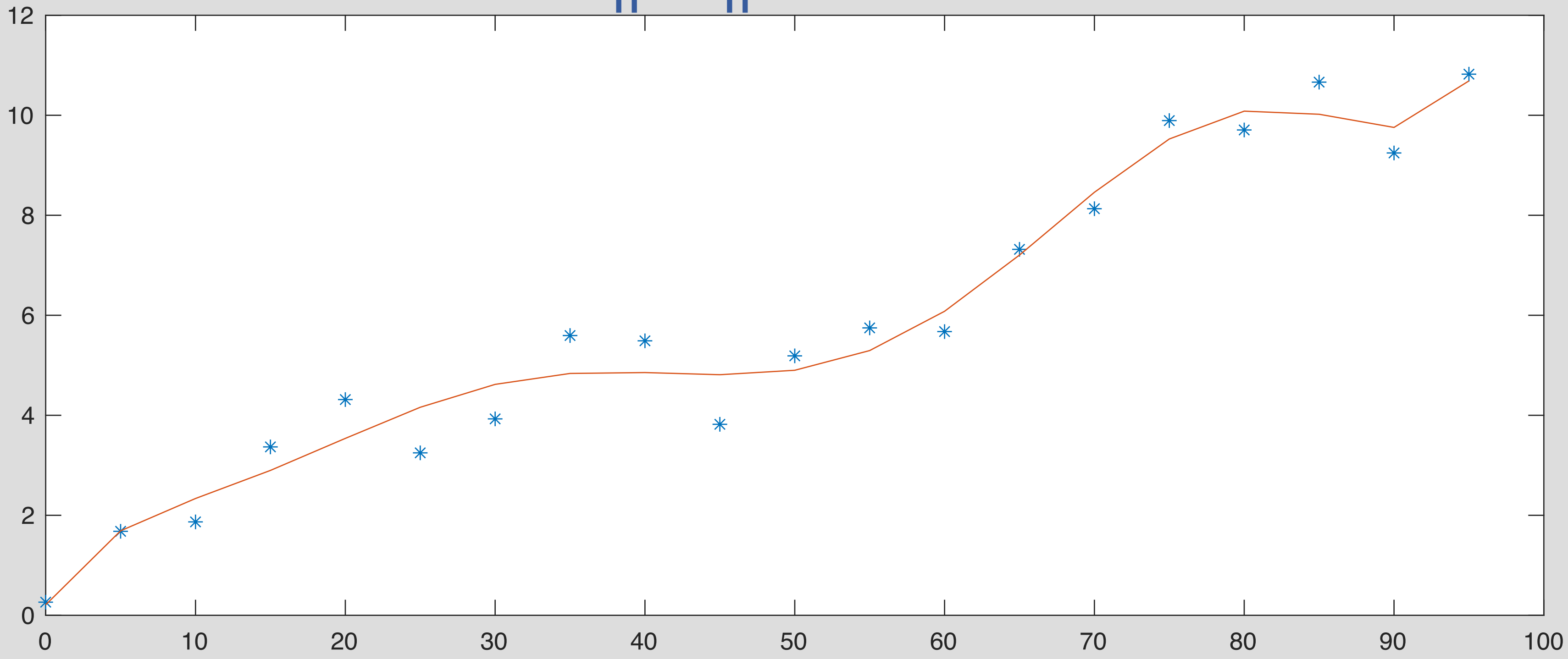
https://www.cebm.net/covid-19/exponential-growth-what-it-is-why-it-matters-and-how-to-spot-it/

Example 6: Over Fitting

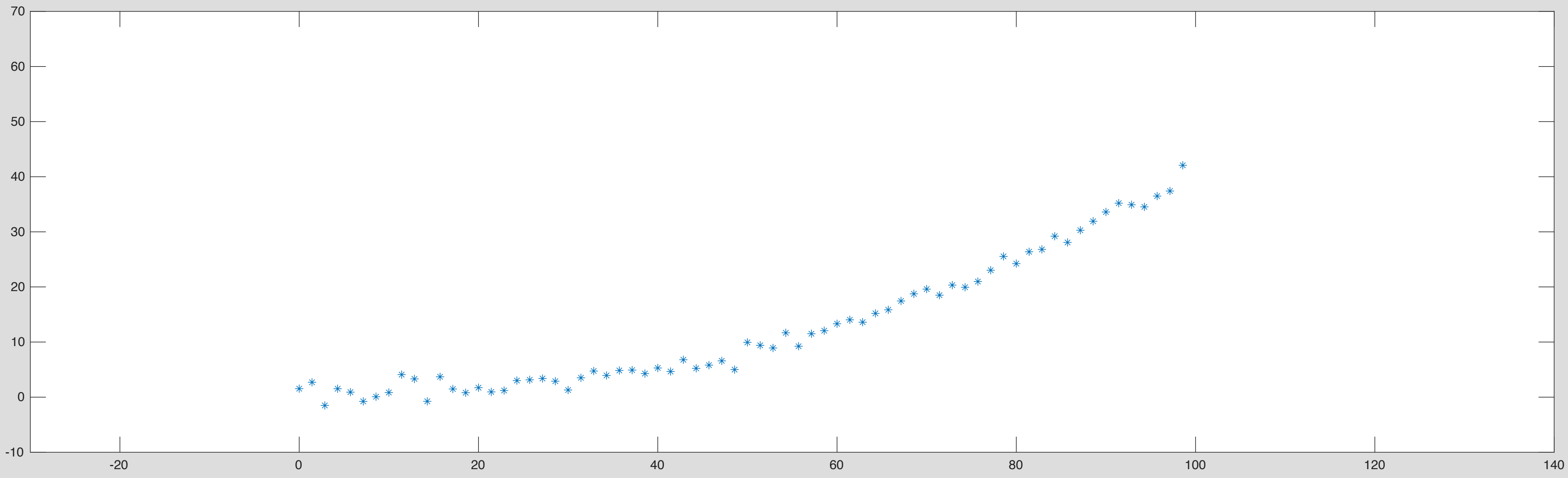
- Consider noisy measurements of $y = 0.1x + 1$:

Model: $y = ax^7 + bx^6 + cx^5 + dx^4 + ex^3 + fx^2 + gx + h$

$$\|\overrightarrow{e}\| = 2.42$$

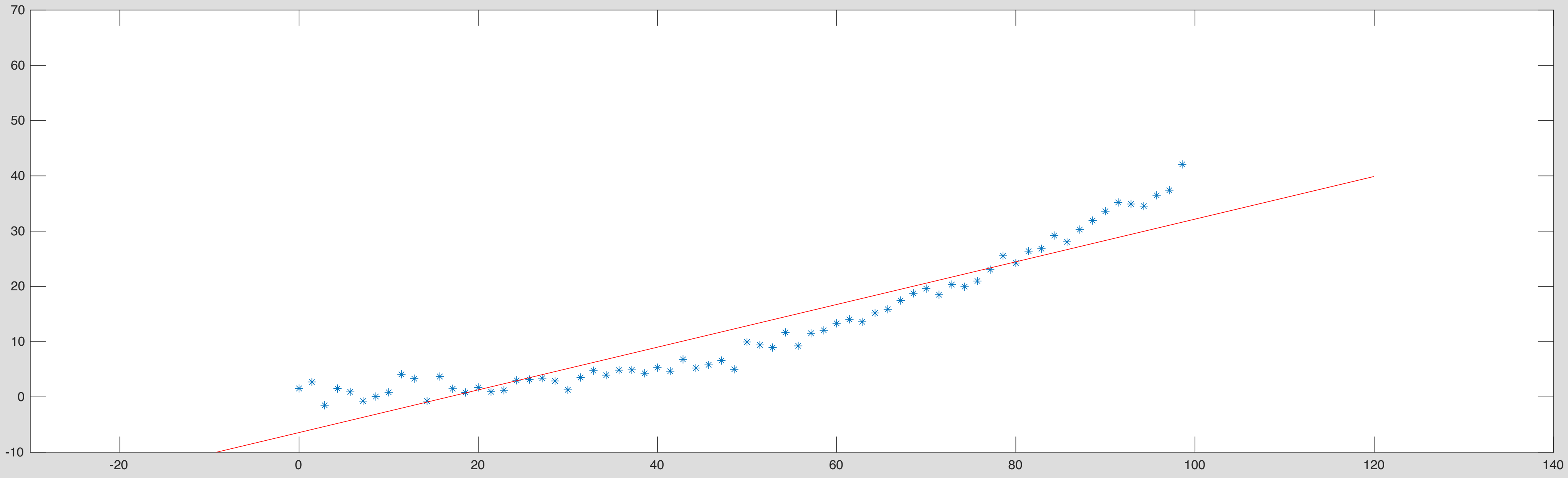


Example 6: Model Order Selection



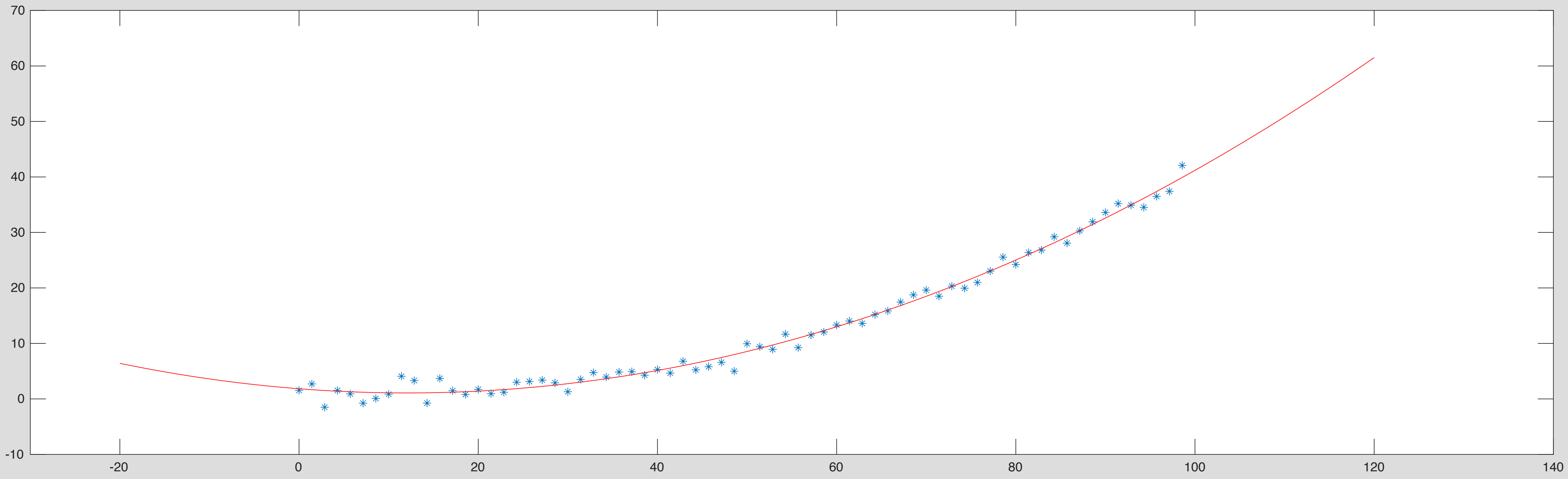
Example 6: Model Order Selection

Model: $y = ax+b$



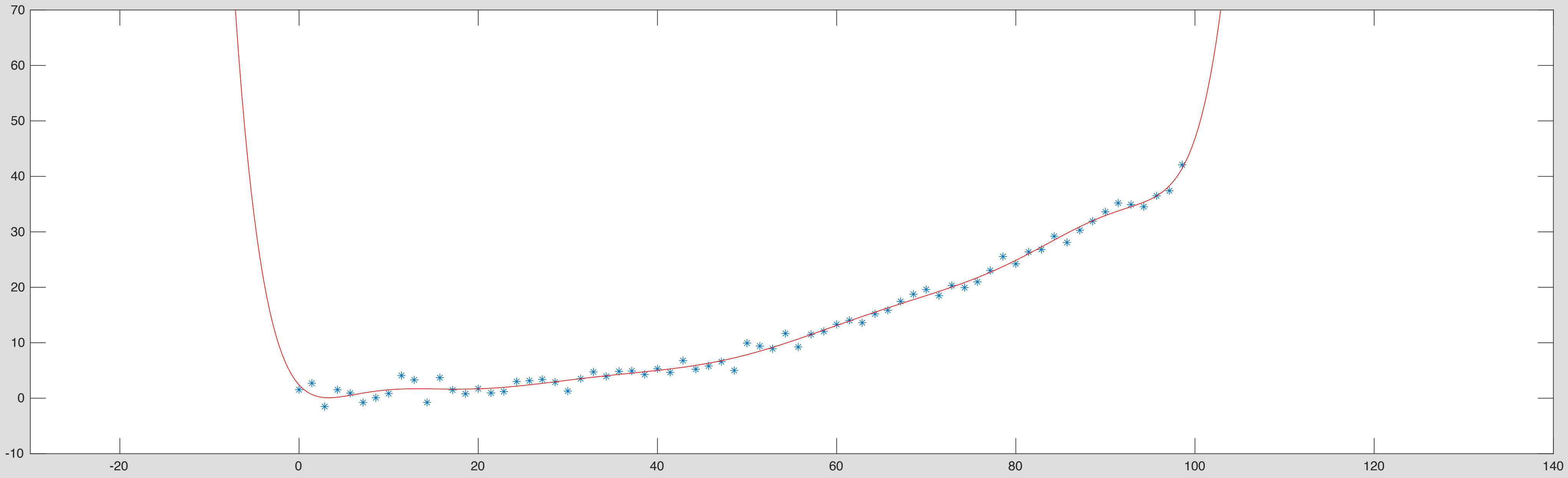
Example 6: Model Order Selection

Model: $y = ax^2 + bx + c$

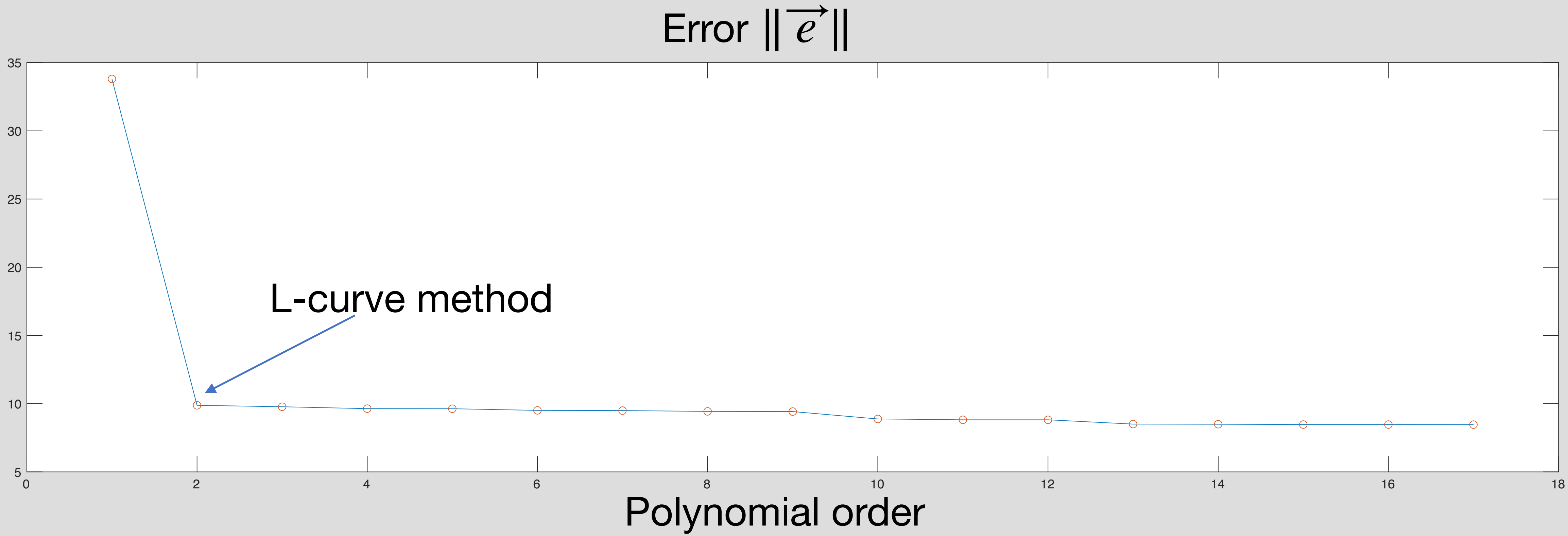


Example 6: Model Order Selection

Model: $y = ax^{10} + bx^9 + cx^8 + dx^7 + ex^6 + fx^5 + gx^4 + hx^3 + ix^2 + jx + k$



Example 6: Model Order Selection

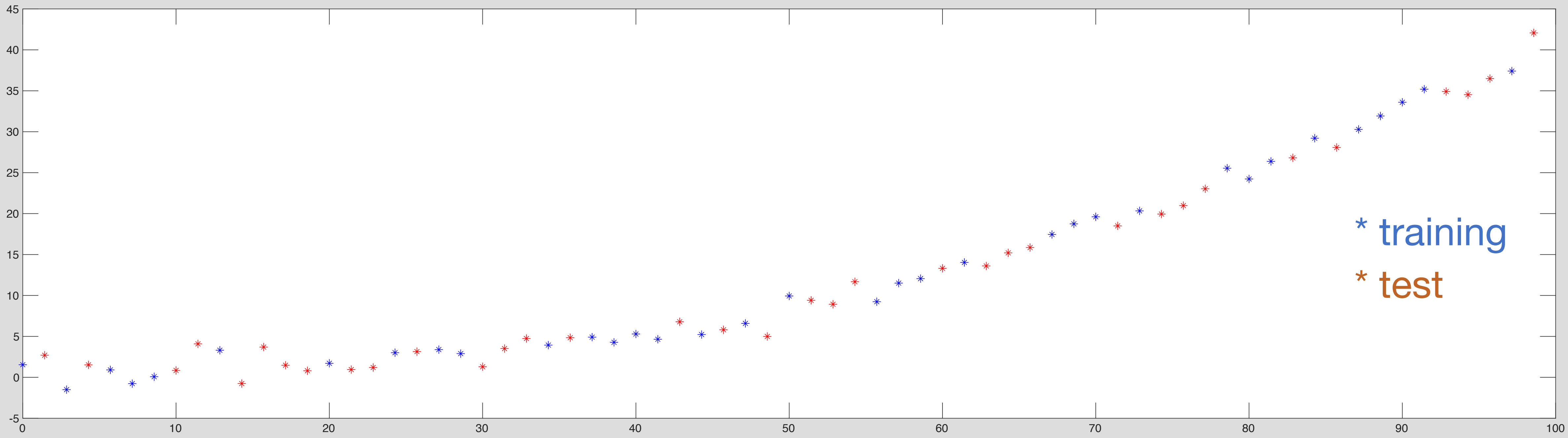


Example 6: Model Order Selection

Split data into training / test sets

Fit model on training set

Evaluate error on test set

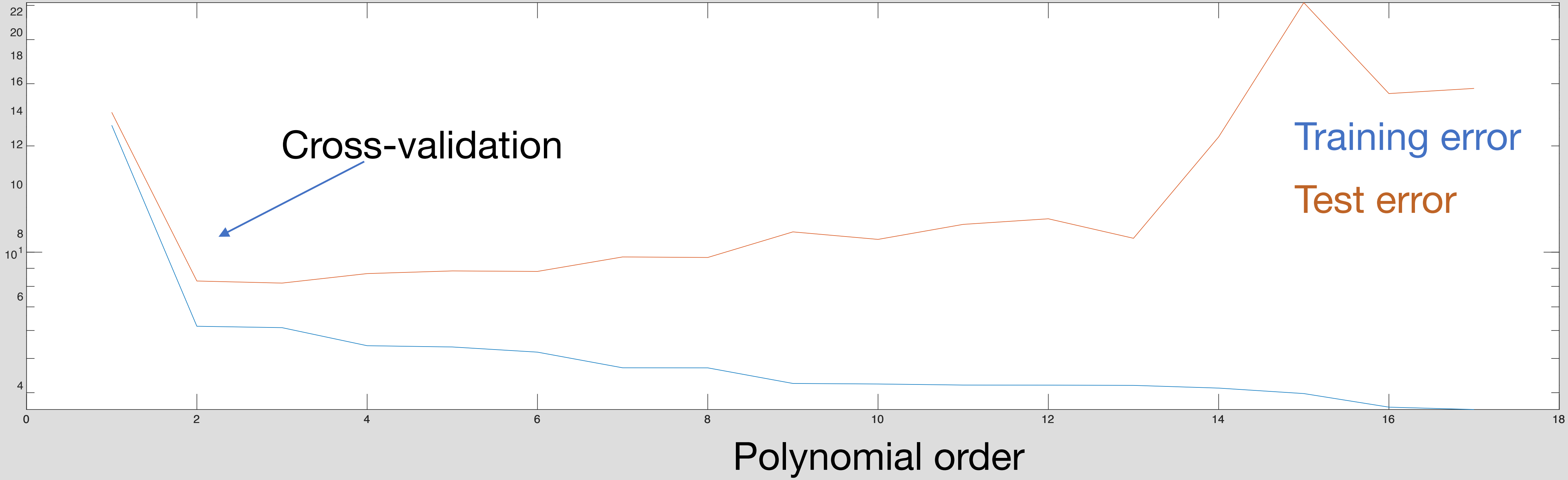


Example 6: Model Order Selection

Split data into training / test sets

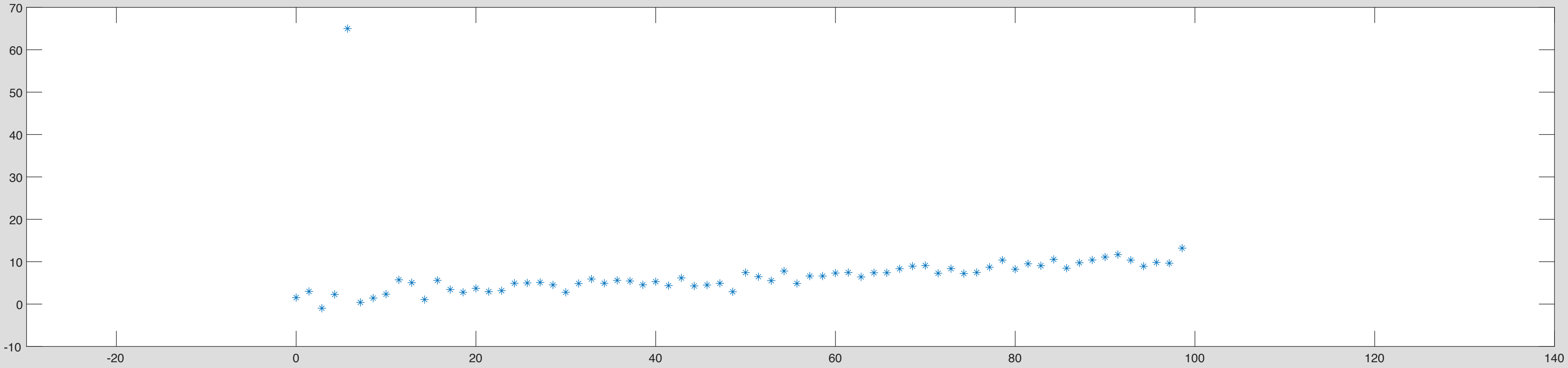
Fit model on training set

Evaluate error on test set



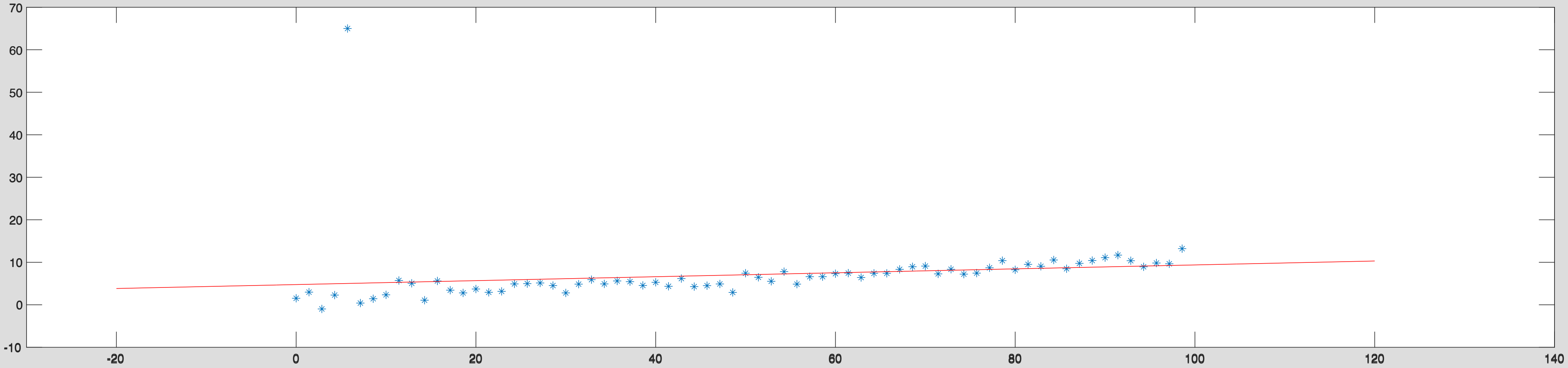
Example 7: Outlier

Model: $y = ax + b$



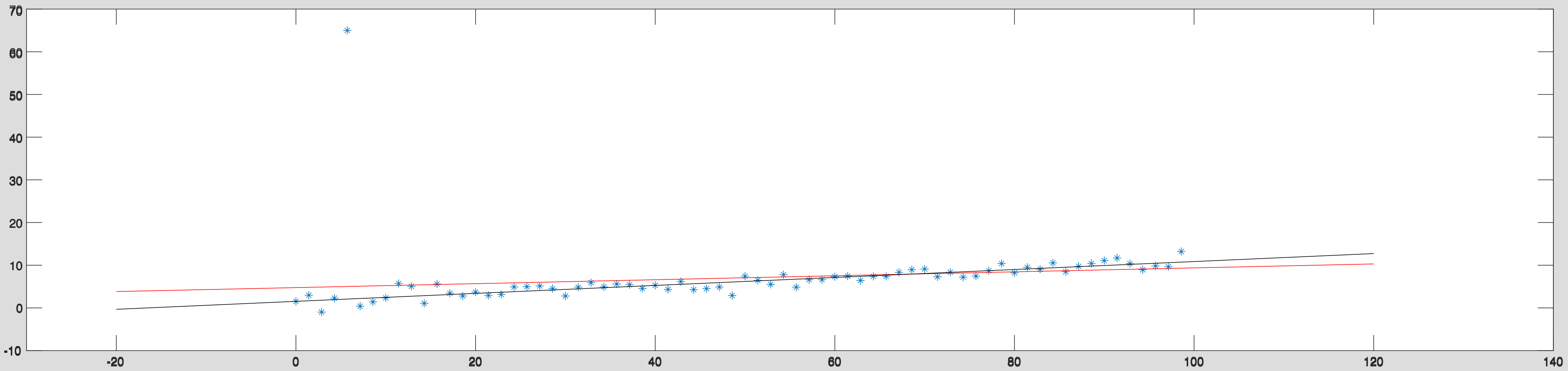
Example 7: Outlier

Model: $y = ax+b$



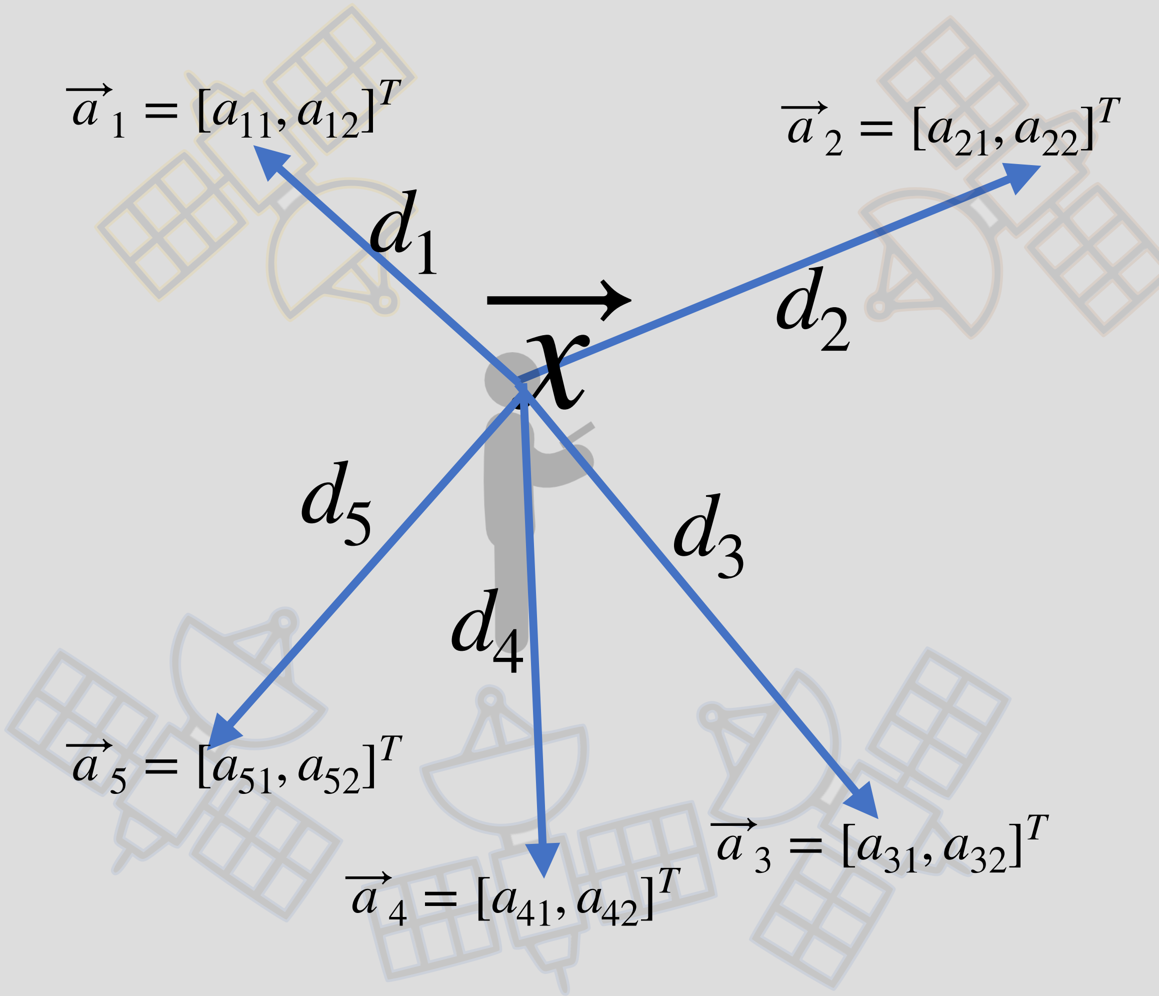
Example 7: Outlier

Model: $y = ax + b$



Multi-Lateration

$$2(\vec{a}_1 - \vec{a}_2)^T \vec{x} - 2C^2 \Delta\tau_2 \tau_1 = \|\vec{a}_1\|^2 - \|\vec{a}_2\|^2 + C^2(\Delta\tau_2)^2$$
$$2(\vec{a}_1 - \vec{a}_3)^T \vec{x} - 2C^2 \Delta\tau_3 \tau_1 = \|\vec{a}_1\|^2 - \|\vec{a}_3\|^2 + C^2(\Delta\tau_3)^2$$
$$2(\vec{a}_1 - \vec{a}_4)^T \vec{x} - 2C^2 \Delta\tau_4 \tau_1 = \|\vec{a}_1\|^2 - \|\vec{a}_4\|^2 + C^2(\Delta\tau_4)^2$$
$$2(\vec{a}_1 - \vec{a}_5)^T \vec{x} - 2C^2 \Delta\tau_5 \tau_1 = \|\vec{a}_1\|^2 - \|\vec{a}_5\|^2 + C^2(\Delta\tau_5)^2$$



More equations than unknowns

$$A \vec{p} = \vec{b}$$

Over-determined — Solve via Least-Squares

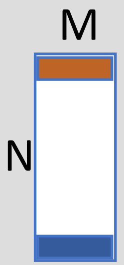
Q: How do we know if $A^T A$ is invertible?

A: if A is full rank!?!?

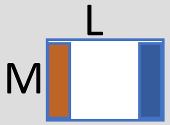
$$\boxed{\hat{p} = (A^T A)^{-1} A^T \vec{y}}$$

Matrix Transposes

$A \in \mathbb{R}^{N\times M}$

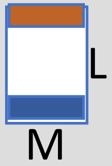


$B \in \mathbb{R}^{M\times L}$

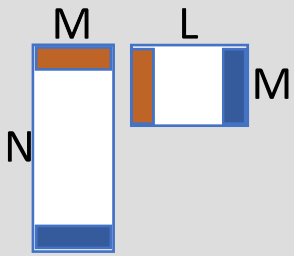


$A^T \in \mathbb{R}^{M\times N}$

$B^T \in \mathbb{R}^{L\times M}$

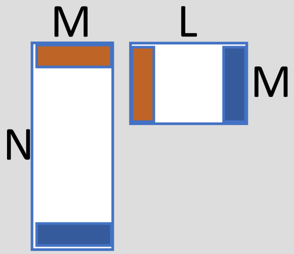


AB



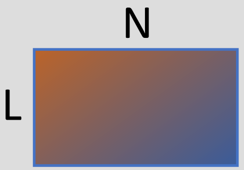
$\in \mathbb{R}^{N\times L}$

$(AB)^T$ $\Bigg($



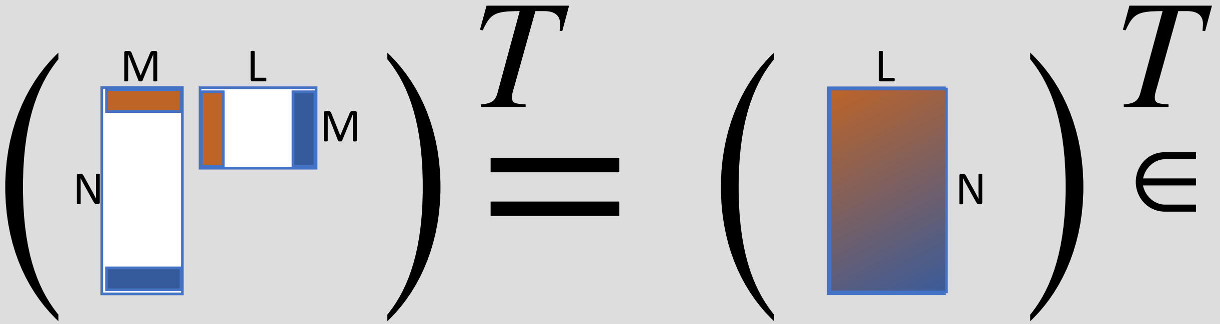
$\Bigg)^T = \Bigg($

$\Bigg)^T \in \mathbb{R}^{L\times N}$

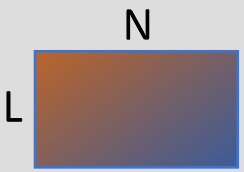


Matrix Transposes

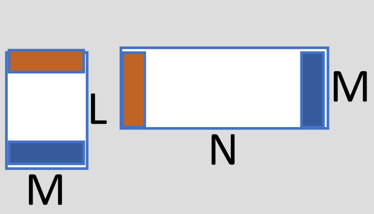
$(AB)^T$



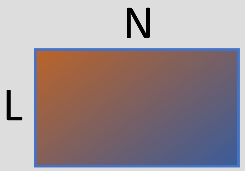
$\in \mathbb{R}^{L \times N}$



$B^T A^T$



$\in \mathbb{R}^{L \times N}$



$$(AB)^T = B^T A^T$$

Invertibility of A^TA

- Invertible $\Rightarrow$ Trivial null space $\Rightarrow$ Linear independent cols/rows….

The matrix A^TA is invertible iff $\text{Null}\left(A^TA\right) = \vec{0}$

Theorem: $\text{Null}\left(A^TA\right) = \text{Null}\left(A\right)$

Proof: (1) show that if $\vec{w} \in \text{Null}(A)$, then $\vec{w} \in \text{Null}(A^TA)$
(2) show that if $\vec{v} \in \text{Null}(A^TA)$, then $\vec{v} \in \text{Null}(A)$

(1). $\vec{w} \in \text{Null}(A)$

$$A\vec{w} = \vec{0}$$

$$A^TA\vec{w} = A^T\vec{0}$$

$$A^TA\vec{w} = \vec{0} \quad ✅$$

(2). $\vec{v} \in \text{Null}(A^TA)$

$$A^TA\vec{v} = \vec{0}$$

Need to show $A\vec{v} = \vec{0}$

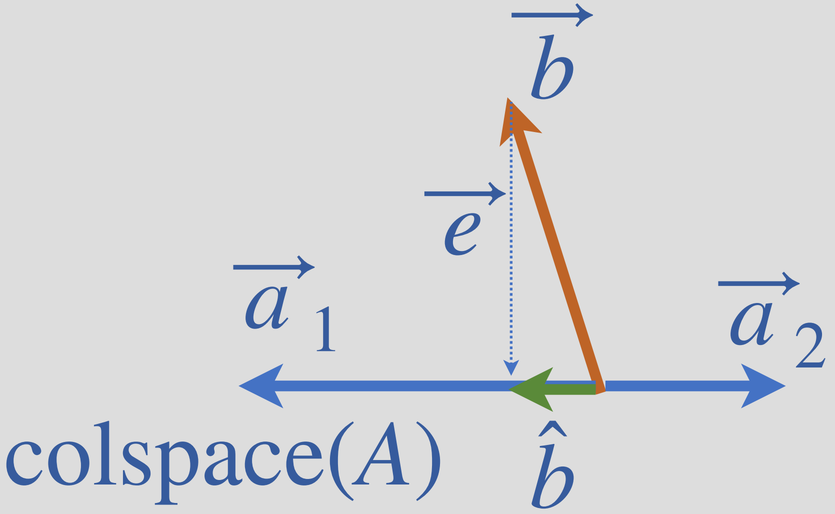
Or… $\|A\vec{v}\| = 0$

$$\begin{aligned}\|A\vec{v}\|^2 &= (A\vec{v})^T(A\vec{v}) \\ &= \vec{v}^TA^T(A\vec{v}) \\ &= \vec{v}^T(A^TA\vec{v}) = 0 \quad ✅\end{aligned}$$

Invertibility of $A^T A$

- What if $A^T A$ is not invertible

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



A: $\hat{x}$ will have infinite solutions with the same $\vec{e} = A\hat{x} - \vec{b}$