EECS 16A

Least Squares Algorithm

Inner Product

- Provide a measure of “similarity” between vectors
- (Euclidian) inner product is also called ‘dot product’

For two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$, the inner product is:

$$\boxed{< \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} >= \vec{\boldsymbol{x}}^{\boldsymbol{T}} \vec{\boldsymbol{y}}}$$

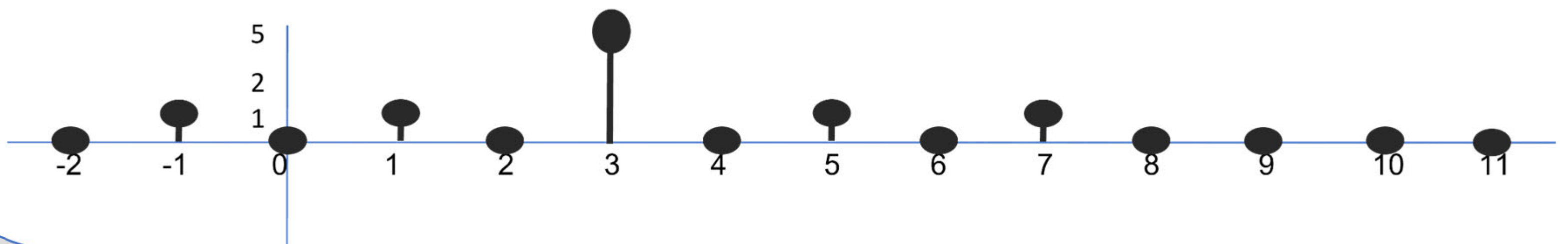
Norm

- Provides a measure of “length” of elements in the vector space

$$\boxed{||\vec{v}|| = \sqrt{< \vec{v}, \vec{v} >}}$$

The cross-correlation between vectors $\vec{r}$ and $\vec{s}$ is:

$$\text{corr}_{\vec{r}}(\vec{s})[k] = \langle r[n], s[n-k] \rangle = \sum_{n=-\infty}^{\infty} r[n]s[n-k]$$

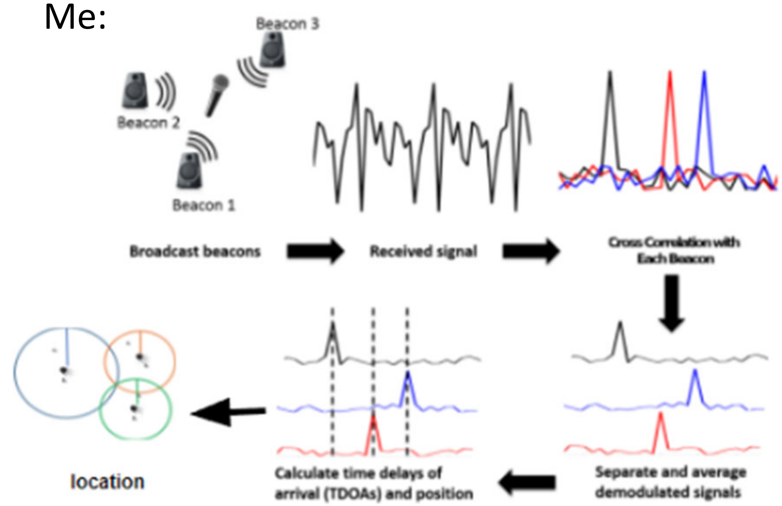


How to solve for GPS coordinates:

1. ***Identify*** which satellites are 'on'
2. Find the ***delay/shift*** for each satellite
3. Use shifts to find ***distances*** to each satellite
4. ***Tri~~lat~~eration*** to find my coordinates

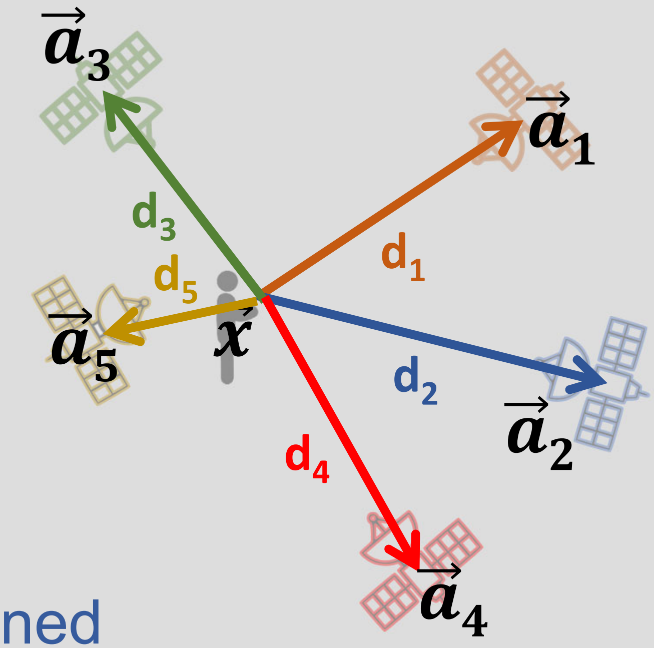
Multilateration

Friend: Come over!
Me: I have no idea where i am and all I have is this recording that sounds like trash
Friend: I have chocolate ☺
Me:

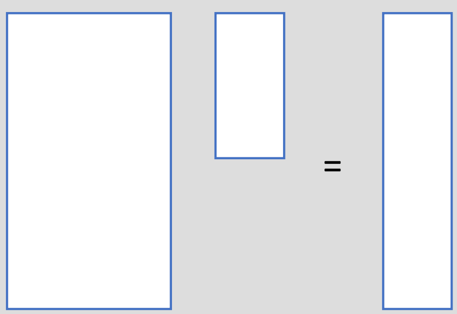


Last time: Multi-Lateration

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



More equations than unknowns!



Over-determined

Q: What if equations are inconsistent due to noise?

A: Find closest solution with Least-Squares!

Last time: Orthogonal projections

- finding the component along a particular direction
- key idea in Machine learning, Signal processing

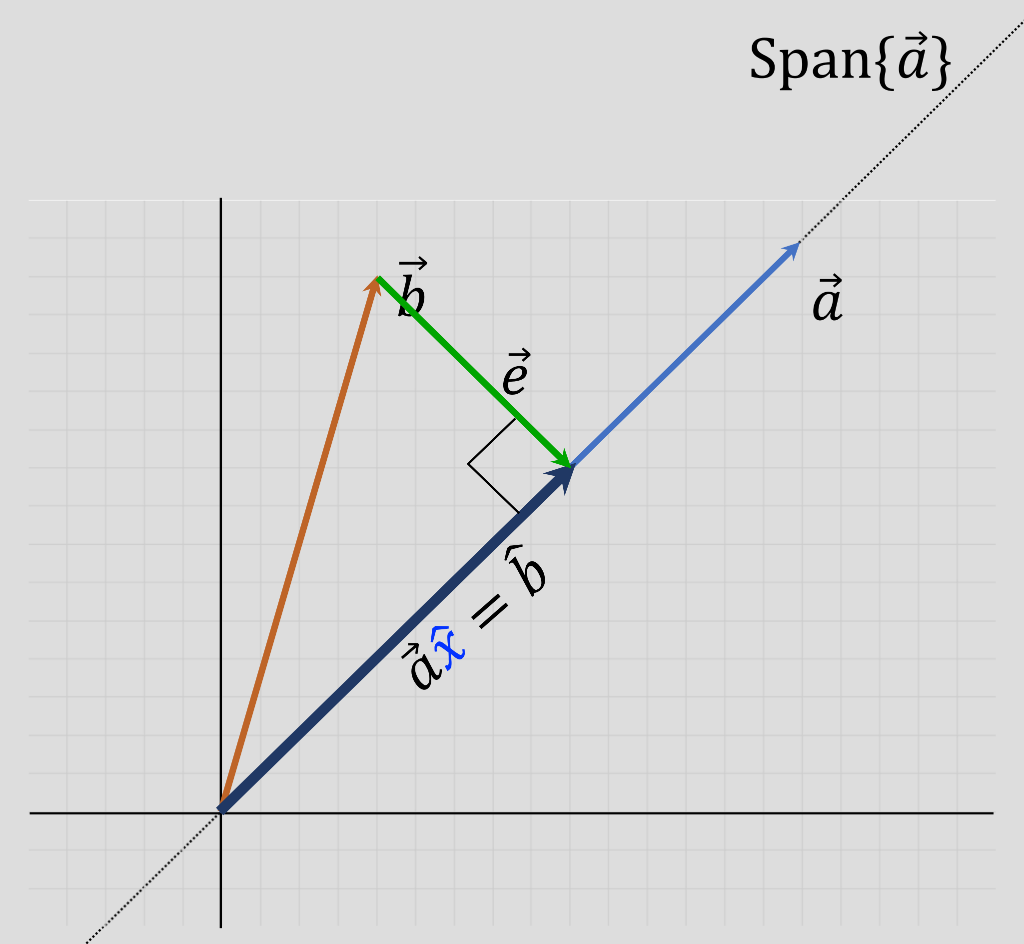
Given vectors $\vec{a}, \vec{b}$,we say that the orthogonal projection of $\vec{b}$ onto $\vec{a}$ is:

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

$$\hat{b} = \frac{\vec{a}^T\vec{b}}{\|\vec{a}\|^2}\vec{a} \qquad \hat{b} = \vec{a}\hat{x}$$

$$\boxed{\hat{x} = \frac{\vec{a}^T\vec{b}}{\|\vec{a}\|^2}}$$

Least squares solution!



find $\hat{x}$ that has the smallest error $\| \vec{e} \| = \| \vec{a}\hat{x} - \vec{b} \|$

Least squares in 2D

3 equations 2 unknowns:

$$\underset{A}{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}} \underset{\vec{x}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} = \underset{\vec{b}}{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}} \quad \text{+error}$$

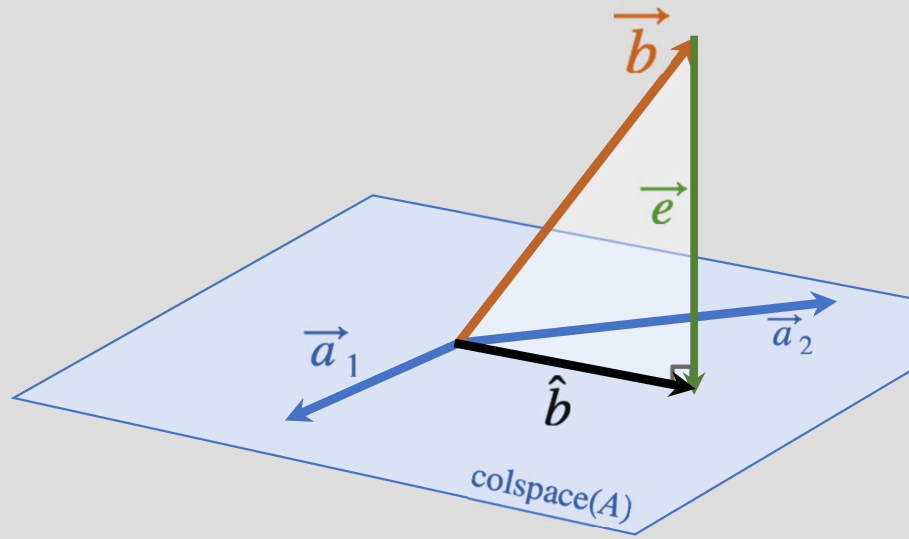
$\vec{a}_1$ $\vec{a}_2$

$\vec{b} \notin \text{colspace}(A) \rightarrow$ No solution! (inconsistent)

Find $\hat{x}$ that is a solution and has smallest error:

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

sol'n will be orthogonal proj. onto colspace(A)



Least squares in many-D

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

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

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

Find $\hat{b} = A\hat{x}$ estimate that min. error

closest measurements that fit model

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

The error vector will be $\perp$ col(A), so $\langle \vec{a}_i, \vec{e} \rangle = 0$

for all i

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

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

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

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

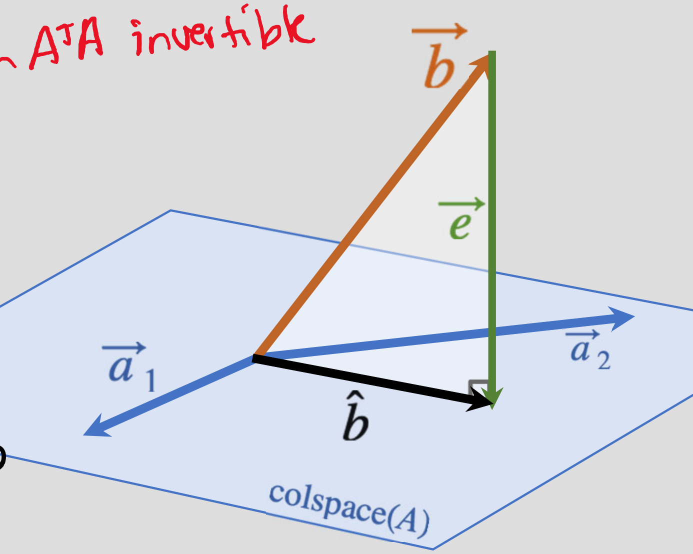
$$A^T\vec{b} - A^TA\hat{x} = \vec{0}$$

$$\underbrace{A^TA}_{\text{if A full rank, then } A^TA \text{ invertible}}\hat{x} = A^T\vec{b}$$

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

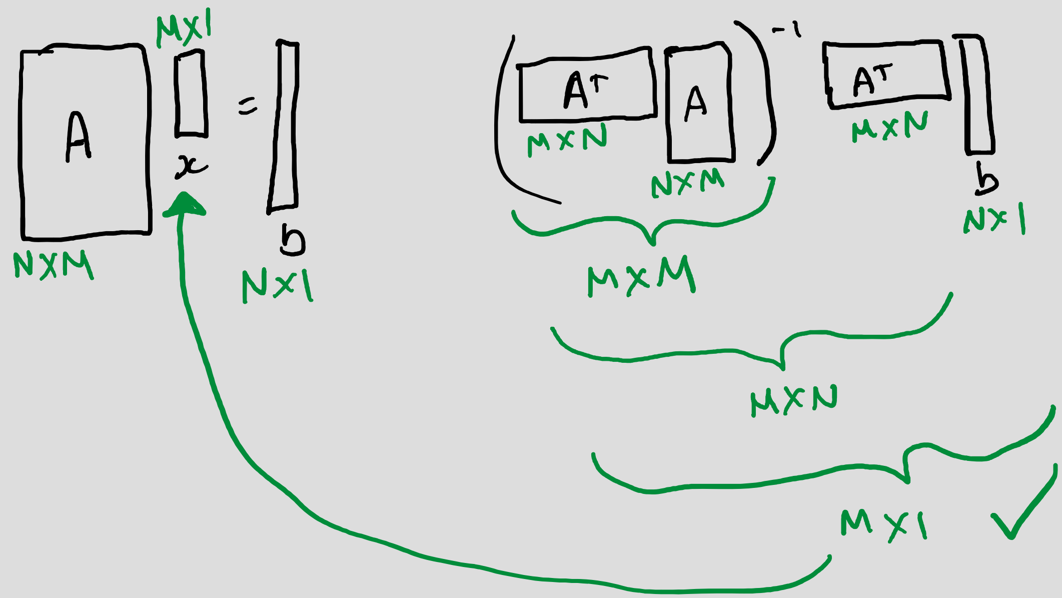
Least squares solution!

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



Least Squares Algorithm

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

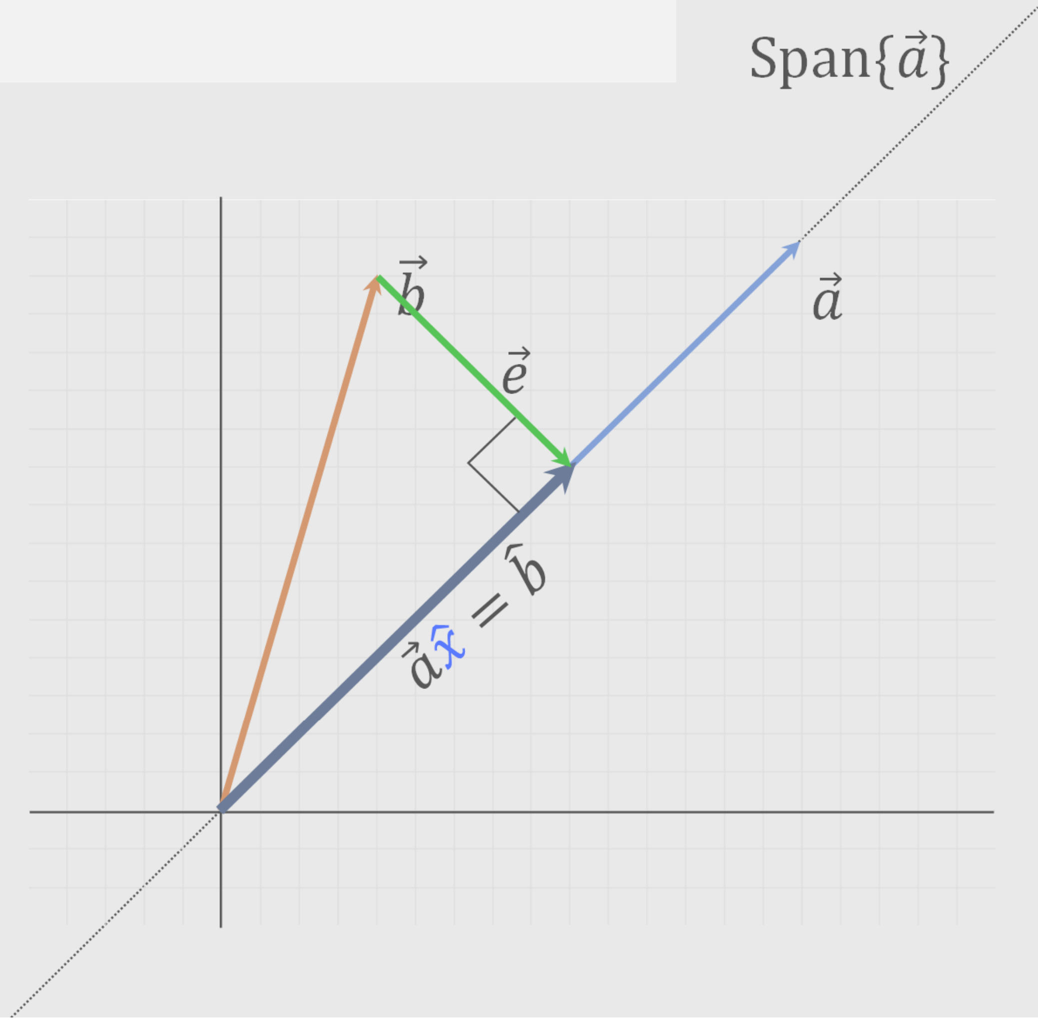


Does it fit with the 1D least squares solution we derived earlier??

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

$$\hat{x} = \frac{\vec{a}^T \vec{b}}{\|\vec{a}\|^2}$$

Least squares solution!



Check:

$$(A^T A)^{-1} A^T \vec{b} = \left(\begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \right)^{-1} \vec{a}^T \vec{b} = (a_1^2 + a_2^2)^{-1} \vec{a}^T \vec{b}$$

$$= \vec{a}^T \vec{b} / \|\vec{a}\|^2 \checkmark$$

A vector ⊥ cols (A) is also ⊥ to anything else in colspace(A)

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

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

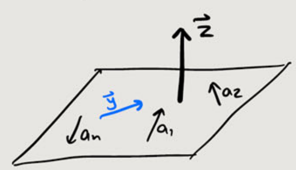
Theorem: Consider matrix A, vector $\vec{y} \in$ colspace(A)

Then, consider vector $\vec{z}$

$$\left.\begin{aligned} \langle \vec{z}, \vec{a}_1 \rangle &= 0 \\ \langle \vec{z}, \vec{a}_2 \rangle &= 0 \\ &\vdots \\ \langle \vec{z}, \vec{a}_n \rangle &= 0 \end{aligned}\right\} \vec{z} \text{ is orthogonal to all vectors in colspace}(A)$$

$\downarrow$ then

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



Proof: we $\boxed{\text{know}}$ $\vec{y} \in$ colspace(A), so it's a lin. combo. of cols:

$$\vec{y} = c_1 \cdot \vec{a}_1 + c_2 \cdot \vec{a}_2 + c_3 \cdot \vec{a}_3 \dots c_n \cdot \vec{a}_n$$

scalar coeffs

we $\boxed{\text{want}}$ $\langle \vec{z}, \vec{y} \rangle = 0$ $\longrightarrow$ $\langle \vec{z}, \vec{y} \rangle = \langle \vec{z}, c_1\vec{a}_1 + c_2\vec{a}_2 + \dots + c_n\vec{a}_n \rangle$

scalar

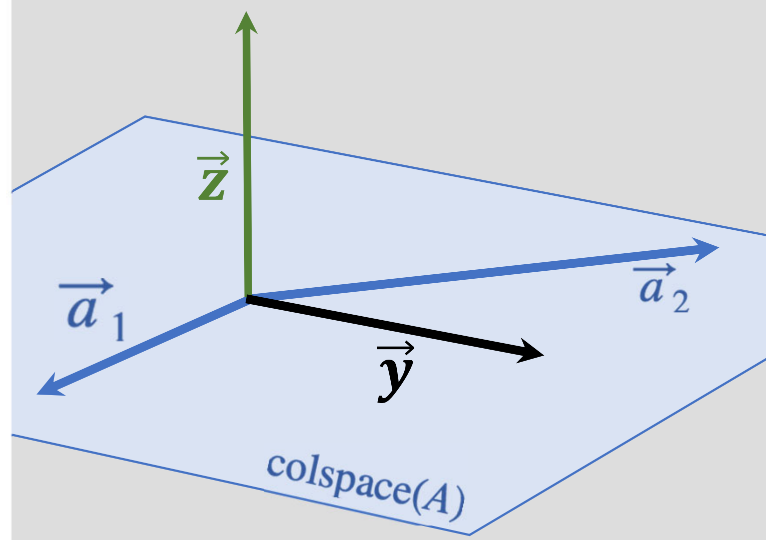
$$= \langle \vec{z}, c_1\vec{a}_1 \rangle + \langle \vec{z}, c_2\vec{a}_2 \rangle + \dots + \langle \vec{z}, c_n\vec{a}_n \rangle$$

$$= c_1 \underbrace{\langle \vec{z}, \vec{a}_1 \rangle}_{\text{given}} + c_2 \langle \vec{z}, \vec{a}_2 \rangle + \dots c_n \langle \vec{z}, \vec{a}_n \rangle$$

$$= c_1(0) + c_2(0) + \dots c_n(0) = 0$$ ✓ yay! ☺

$$\underset{m \times n}{A} = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \\ | & | & & | \end{bmatrix}$$

col view



Least squares: Example 1

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

$$\underset{A}{\begin{bmatrix} 2 \\ 1 \end{bmatrix}} \underset{\hat{x}}{[x]} = \underset{\vec{b}}{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}$$

$$\left.\begin{array}{l} 2x = 1 \rightarrow x = 1 \\ 1x = 1 \quad\;\; x = 0.5 \end{array}\right\} \;??\;$$ Inconsistent

Try Gaussian Elimination: $\left[\begin{array}{c|c} 2 & 1 \\ 1 & 1 \end{array}\right] \rightarrow \left[\begin{array}{c|c} 2 & 1 \\ 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{c|c} 1 & 0.5 \\ 0 & 1 \end{array}\right]$

No sol'n

Inconsistent

Find Least Squares sol'n:

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

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

$$= (5)^{-1}(3) = \boxed{\frac{3}{5}}$$

$$\hat{b} = A\hat{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} 3/5 = \begin{bmatrix} 6/5 \\ 3/5 \end{bmatrix}$$

$$\vec{e} = \vec{b} - \hat{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 6/5 \\ 3/5 \end{bmatrix} = \begin{bmatrix} 1/5 \\ -2/5 \end{bmatrix}$$

$$\|\vec{e}\| = \sqrt{\frac{1}{25} + \frac{4}{25}} = \sqrt{\frac{1}{5}}$$

Why wasn't it average $x = 0.75$?

$$\hat{b}' = \begin{bmatrix} 2 \\ 1 \end{bmatrix} 3/4$$

if $\hat{b}' = \begin{bmatrix} 6/4 \\ 3/4 \end{bmatrix}$

$$\vec{e}' = \begin{bmatrix} 1/2 \\ -1/4 \end{bmatrix}$$

Bigger $\rightarrow$ $\|\vec{e}'\| = \sqrt{\frac{4}{16} + \frac{1}{16}} = \sqrt{\frac{5}{16}}$

Least squares: Example 2

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

$$\begin{array}{ccc} A & \vec{x} & \vec{b} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = & \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \end{array}$$

$x_1 = 1$

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

Gauss. Elim:

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

inconsistent

Least squares sol'n:

$$(A^T A)^{-1} A^T b = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$= \left(\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}}_{\text{invert diagonals}} \right)^{-1} \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

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

Example 3: Linear Regression

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

fit model params (unknown)

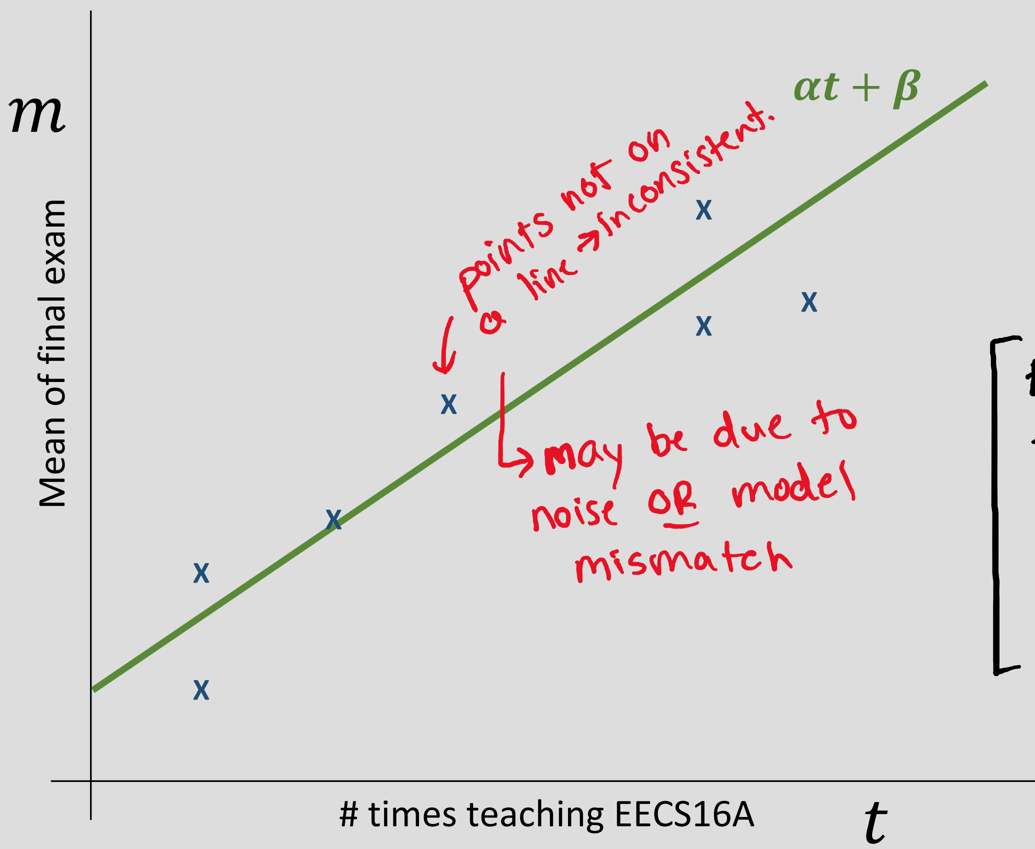
Model: $m = \alpha t + \beta$

Known:
Waller: (t_1, m_1)
Sahai: (t_2, m_2)
Alon: (t_3, m_3)
Stojanovic: (t_4, m_4)
Ranade: (t_5, m_5)
Courtade: (t_6, m_6)
Liu: (t_7, m_7)

known data

times taught

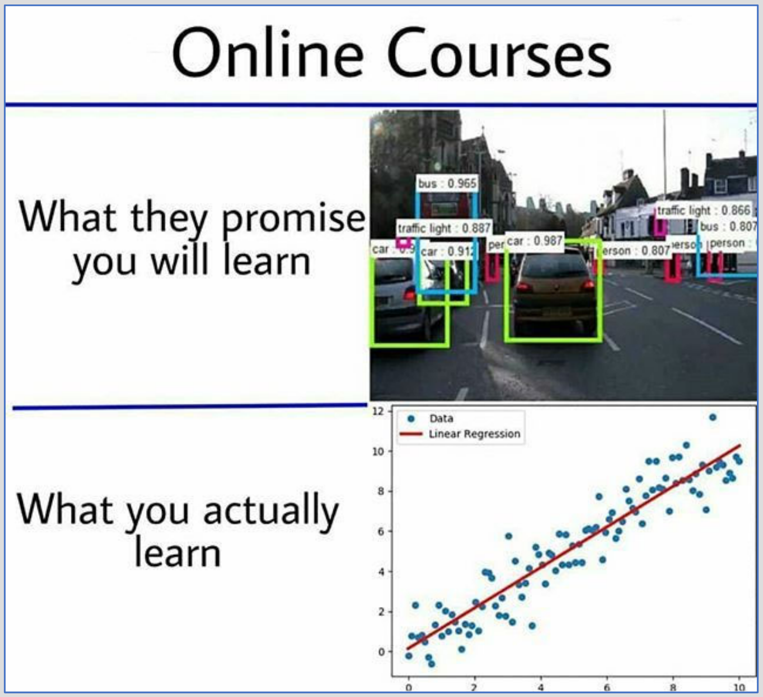
mean of final exam

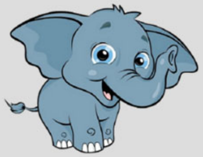


$$\underbrace{\begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ t_3 & 1 \\ \vdots & \vdots \\ t_7 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_7 \end{bmatrix}}_{\vec{b}}$$

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

BUT, not everything fits to a line!?!





“Using a term like nonlinear science is like referring to the bulk of zoology as the study of non-elephant animals.” – S. Ulam

Example 4: Regression

Gauss found Ceres by using Kepler's laws: planet

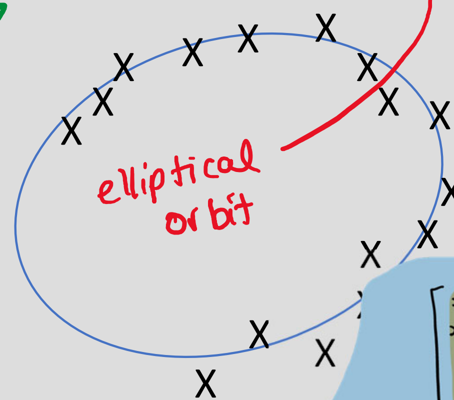
Model: $ax^2 + by^2 + cxy + dx + ey = 1$

unknowns

knowns!

Q: Is this a linear fit?

A: Yes!



$$\underbrace{\begin{bmatrix} x_1^2 & y_1^2 & x_1y_1 & x_1 & y_1 \\ x_2^2 & y_2^2 & x_2y_2 & x_2 & y_2 \\ x_3^2 & y_3^2 & x_3y_3 & x_3 & y_3 \\ & & \vdots & & \\ x_N^2 & y_N^2 & x_Ny_N & x_N & y_N \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix}}_{\vec{p}} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{\vec{y}}$$

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

→ get computer to do it!

$$\underbrace{\begin{bmatrix} x_1^2 & y_1^2 & x_1y_1 & x_1 & y_1 \\ x_2^2 & y_2^2 & x_2y_2 & x_2 & y_2 \\ & & \vdots & & \\ x_{22}^2 & y_{22}^2 & x_{22}y_{22} & x_{22} & y_{22} \end{bmatrix}}_{22 \times 5} \underbrace{\begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix}}_{5 \times 1} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{22 \times 1}$$

in machine learning, cols are called 'features'

solve for me!

22 equations, 5 unknowns

'Overdetermined'

*Gauss did this by hand! We have Jupyter notebooks so can be lazy ☺ yay! (see slides or 'Ceres-orbit' notebook)

Least-squares is building block for all of signal processing / machine learning / pattern matching

Example 5: Exponential Regression

Model: $y = ce^{ax}$

Q: Is this a linear fit?

A: No! But, can be made linear…..

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

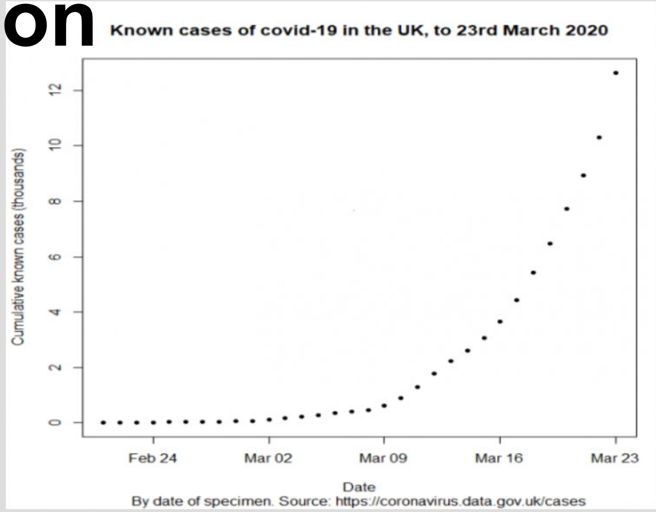
Knowns: $(x_1, \log(y_1))(x_2, \log(y_2))..(x_N, \log(y_N))$

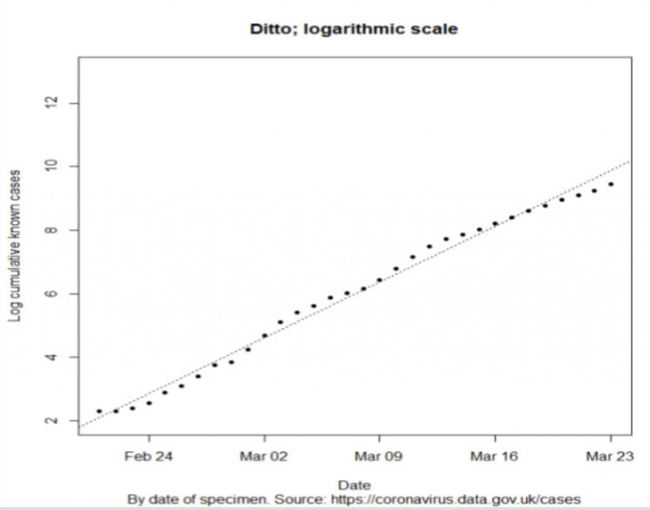
Unknowns: $\vec{p} = [a \;\; b]^T$

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

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

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

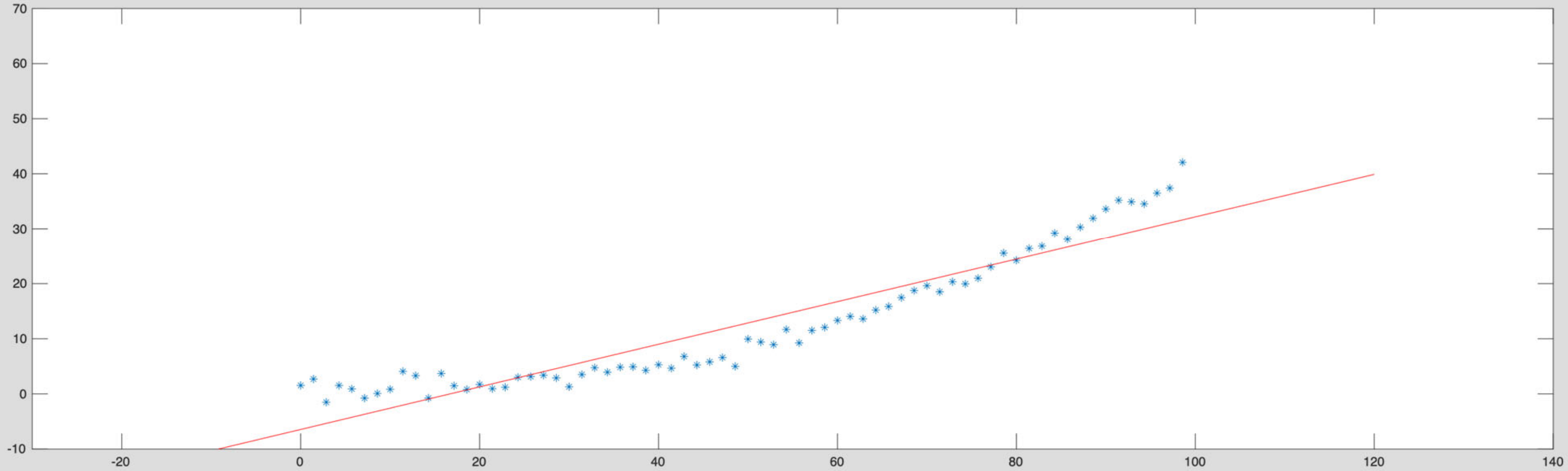




https://www.cebm.net/covid-19/exponential-growth-what-it-is-why-it-matters-and-how-to-spot-it/

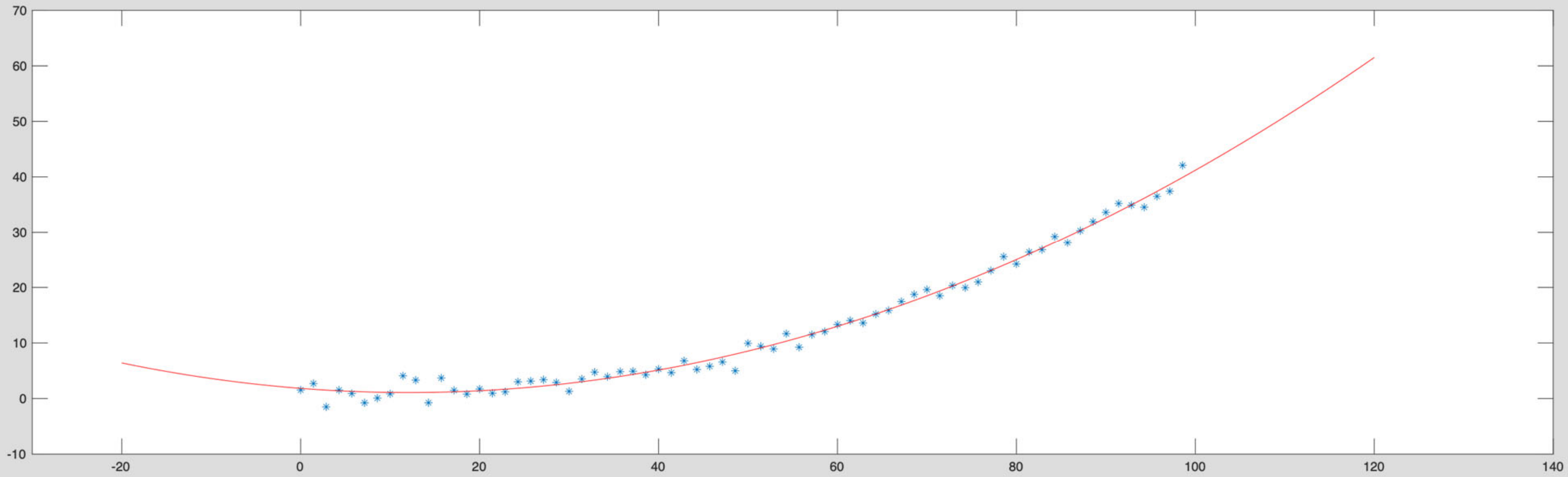
Example 6: Model Order Selection

Model: $y = ax + b$



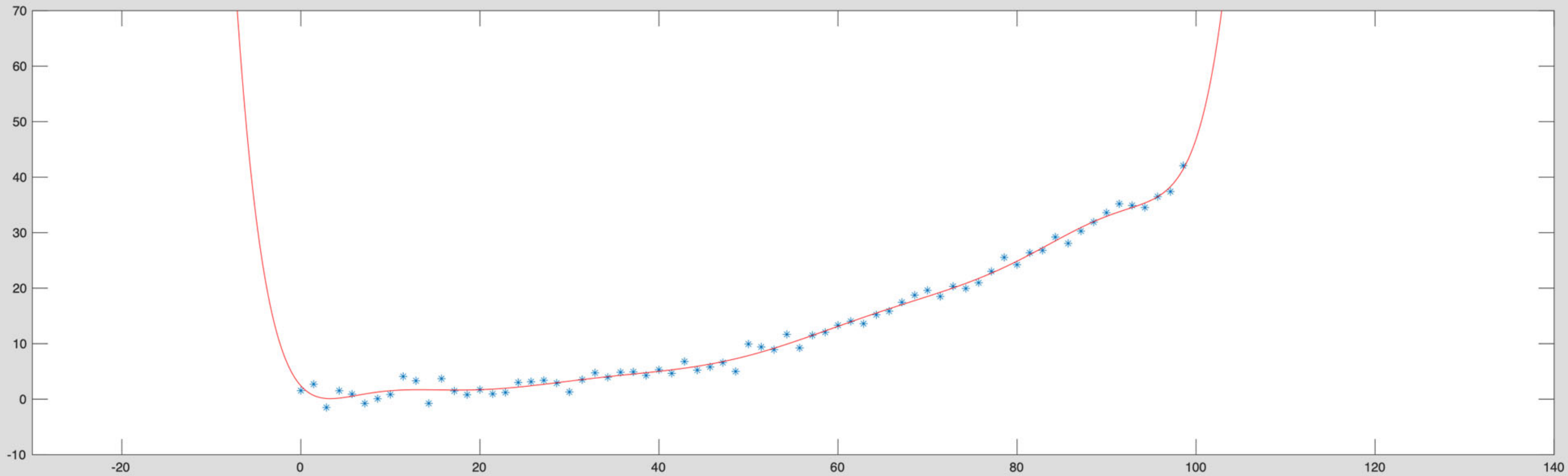
Example 6: Model Order Selection

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



Example 6: Model Order Selection

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



Example 6: Model Order Selection