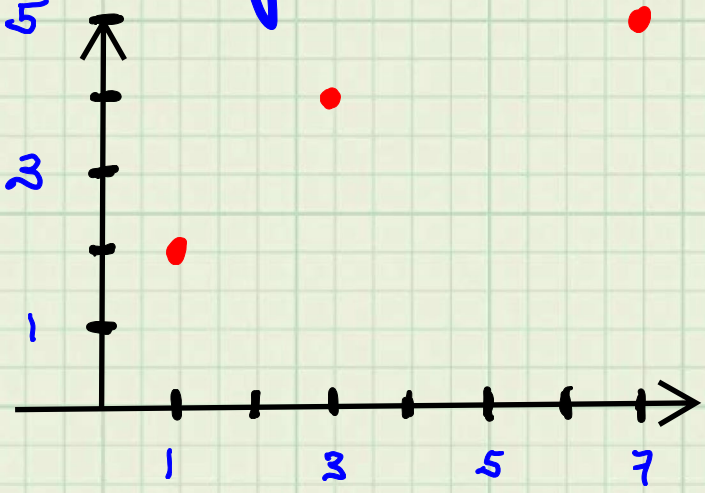
Chapter 8: Least squared error

solutions to $Ax=b$, b not a linear combination of the columns of A! Big

Application: Linear Regression = Fitting Functions to Data

Machine Learning (AI) does a nonlinear version of "regression" = fitting functions to data!

Summary



x_i	y_i
1	2
3	4
7	5

Linear model → two unknowns $\alpha = \begin{bmatrix} m \\ b \end{bmatrix}$

$y_i \approx m x_i + b$

$$\begin{aligned} y_1 &= m x_1 + b \\ y_2 &= m x_2 + b \\ y_3 &= m x_3 + b \end{aligned} \iff \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} m \\ b \end{bmatrix}}_{\alpha}$$

$$Y = \Phi \cdot \alpha$$

$$\begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 3 & 1 \\ 7 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix}$$

Y Φ α

When you are first learning something unfamiliar notation throws you off your game!

Re-write as $Y = \Phi \cdot \alpha$

$Ax = b$

$A = \Phi$ (regressor matrix)

$x = \alpha = \begin{bmatrix} m \\ b \end{bmatrix}$ unknowns

$b = Y$ measured data

- There is no line that goes through all data points
- There is no solution to our system of linear equations
- We want the line that "best fits"

the data !!!

Write $e(x) := Ax - b$

= error in the solution, and it depends on the x we select

- For some x, $e(x) := Ax - b$ might be HUGE

For other x, $e(x) := Ax - b$ might be smaller.

Question: Can we find an x that makes the norm of the error as small as possible ??

If so, is it unique ?

Answer: Suppose the columns of A are linearly independent. Then there exists $x^* \in \mathbb{R}^m$ such that

$$\|e(x^*)\| = \|Ax^* - b\| \leq \|Ax - b\| \text{ for all } x \in \mathbb{R}^m.$$

Moreover, x^* is unique.

$A = n \times m$, $b = n \times 1$

Meaning:

$$\|Ax^* - b\| = \min_{x \in \mathbb{R}^m} \|Ax - b\|$$

x^* minimizes the norm of the error!!

One writes $x^* = \arg\min_{x \in \mathbb{R}^m} \|Ax - b\|$

x^* is that vector in $\mathbb{R}^m$ achieving the minimum error.

How to Compute it?

Fact: Suppose the columns of A are linearly independent.

Then TFAE

1. $x^* = \arg\min_{x \in \mathbb{R}^m} \|Ax - b\|$
2. $A^TA x^* = A^Tb$
3. $x^* = (A^TA)^{-1}A^Tb$

Recall: $\det(A^TA) \neq 0 \iff$ columns of A are lin. indep.

Our Line Fitting Problem

$$A = \underbrace{\begin{bmatrix} 1 & 1 \\ 3 & 1 \\ 7 & 1 \end{bmatrix}}_{\Phi} \qquad b = \underbrace{\begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}}_{Y} \qquad x = \underbrace{\begin{bmatrix} m \\ b \end{bmatrix}}_{\alpha}$$

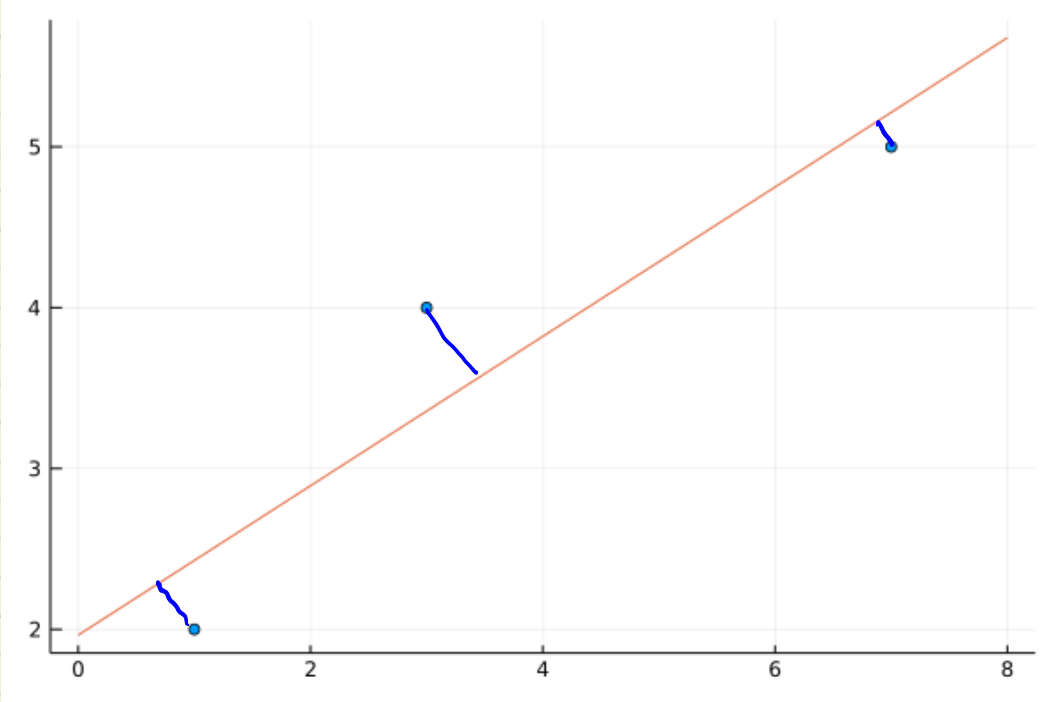
$Ax = b \iff Y = \Phi \cdot \alpha$

$$(A^TA)x^* \simeq A^Tb$$

$$A^TA = \begin{bmatrix} 1 & 3 & 7 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 3 & 1 \\ 7 & 1 \end{bmatrix} = \begin{bmatrix} 59 & 11 \\ 11 & 3 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 1 & 3 & 7 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 49 \\ 11 \end{bmatrix}$$

$$x^* = \begin{bmatrix} 0.46 \\ 1.96 \end{bmatrix} \qquad m \simeq 0.46 \qquad b \simeq 1.96$$



Why True?

(1) $$x^* = \arg\min_{x \in \mathbb{R}^m} \|Ax-b\|$$

$$\Updownarrow$$

$$x^* = \arg\min_{x \in \mathbb{R}^m} \|Ax-b\|^2$$

$$\|Ax-b\|^2 = (Ax-b)^T(Ax-b)$$

Some painful algebra to "complete the square"

$$(Ax-b)^T(Ax-b) = (A^TAx-A^Tb)^T(A^TA)^{-1}(A^TAx-A^Tb) + b^Tb - b^TA(A^TA)^{-1}A^Tb$$

$b^Tb - b^TA(A^TA)^{-1}A^Tb$ does not depend on x

Minimize with $\boxed{A^TAx - A^Tb = 0}$

Someone asked: Why not minimize with a large negative value instead of zero?

Answer: $(A^TA)^{-1}$ is what we call a positive definite matrix (see Appendix of Book)

Such matrices have the property that $y^T(A^TA)y \geq 0$ for all $y \in \mathbb{R}^m$ and $y^T(A^TA)y = 0 \iff y = 0$!!

This a bit advanced for ROB 101.