4 weeks of class remaining.

Final Project (due Dec 1):

Write a point form summary of what you learned in this course.

For the remainder of the course I want to show you the two most common applications of Linear Algebra:

1. Least Squares Approximation
2. Diagonalization ("Spectral Analysis")

This Week: Least Squares.

Suppose that a linear system $A\vec{x} = \vec{b}$ has NO SOLUTION.

[In particular, this means that A^{-1} does not exist.]

So what can we do? In this case we will look for the "best" approximate solution $\hat{x}$:

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

"Best" in what sense? The OLS appoximation (Ordinary Least Squares) will minimize the length

$$\| A\hat{x} - \vec{b} \|.$$

Without explaining the details, let me just tell you the solution:

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

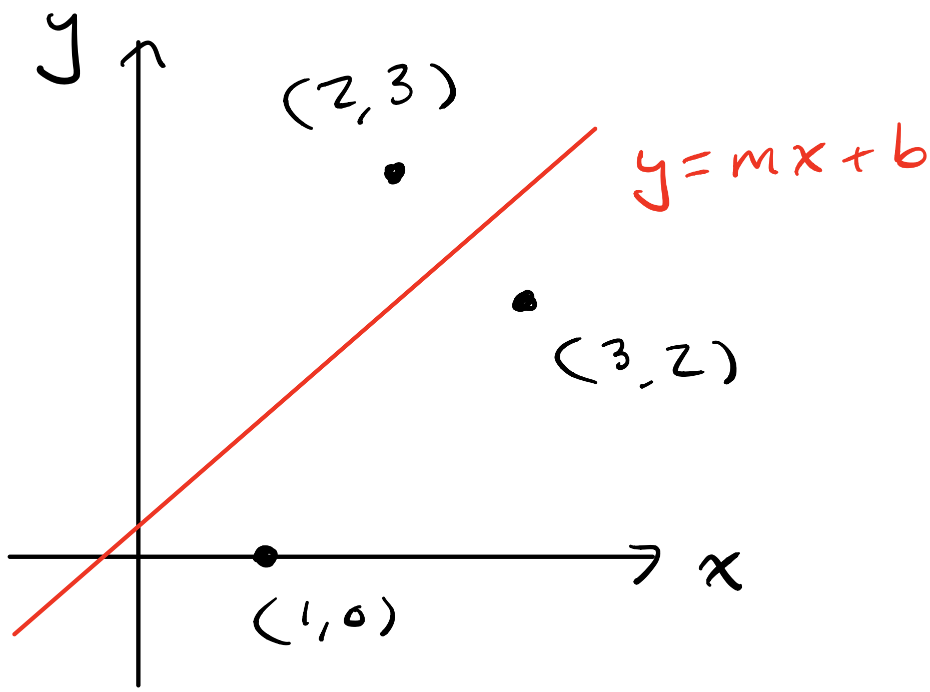
$$\hat{x} = (A^T A)^{-1} A^T \vec{b}.$$

That's it!

The Prototypical Example of OLS:

Find the line $y = mx + b$ that is "closest" to the 3 data points

$$(x, y) = (1, 0),\ (2, 3),\ (3, 2).$$



Solve for m & b. How?

Each data point gives us one linear equation in unknowns m & b:

data point (x,y)	equation $y=mx+b$
$(1,0)$	$0 = 1m + b$
$(2,3)$	$3 = 2m + b$
$(3,2)$	$2 = 3m + b$

Write this as a system of 3 linear equations in 2 unknowns:

$$\begin{cases} 1m + b = 0 \\ 2m + b = 3 \\ 3m + b = 2 \end{cases} \rightsquigarrow \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix}$$

Does this have a solution? NO!

Reason:

- More than 2 equations in 2 unknowns probably has no solution.
- More than 2 points in the plane probably do not lie on the same line.

The general method stated above tells us that the best approximate solutions $\hat{m}$ & $\hat{b}$ satisfy the so-called "normal equation":

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} \hat{m} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix}$$

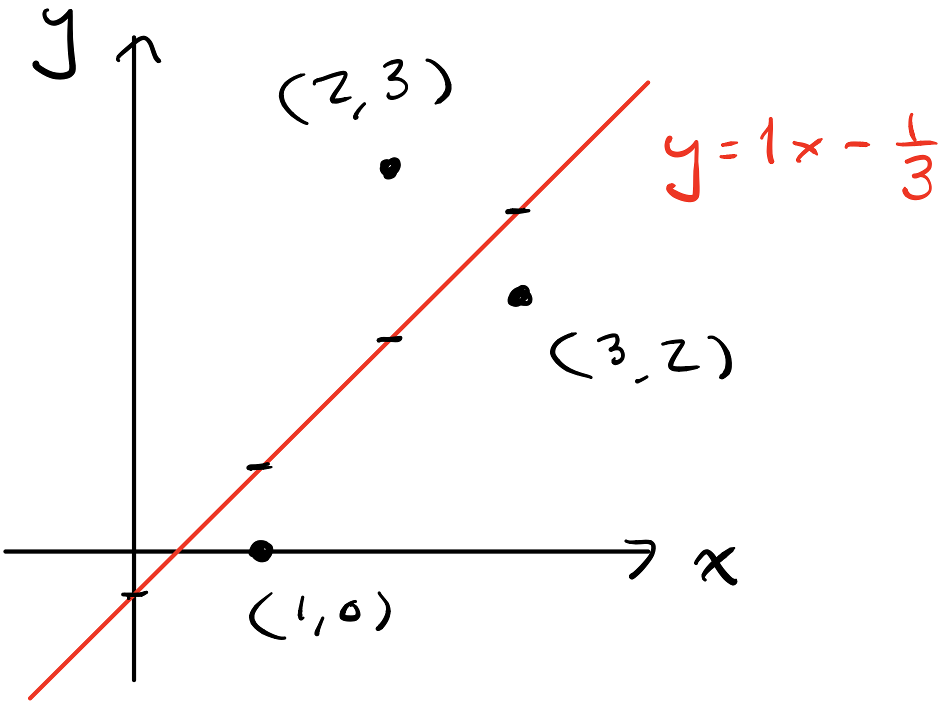
$$\begin{pmatrix} 14 & 6 \\ 6 & 3 \end{pmatrix} \begin{pmatrix} \hat{m} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} 12 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} \hat{m} \\ \hat{b} \end{pmatrix} = \frac{1}{14 \cdot 3 - 6 \cdot 6} \begin{pmatrix} 3 & -6 \\ -6 & 14 \end{pmatrix} \begin{pmatrix} 12 \\ 5 \end{pmatrix}$$

$$= \frac{1}{6} \begin{pmatrix} 6 \\ -2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ -1/3 \end{pmatrix}$$

Best Fit Line: $y = 1x - \frac{1}{3}$.



In what sense is this "best"?

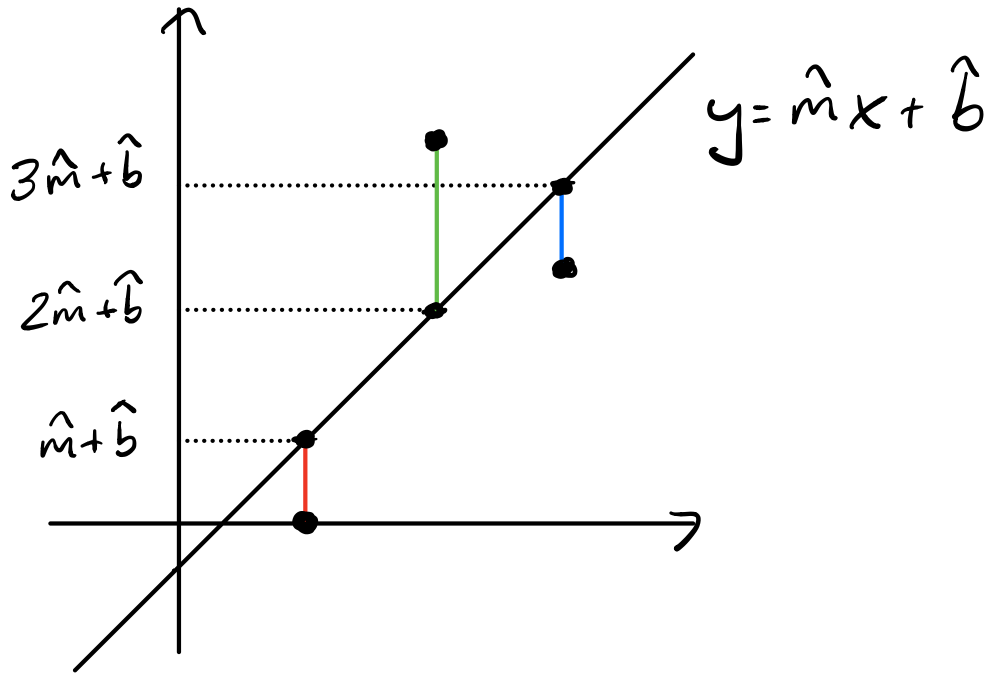
Meaning of OLS: (squared) length is minimized:

$$\left\| \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} \hat{m} \\ \hat{b} \end{pmatrix} - \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix} \right\|^2$$

$$= \left\| \begin{pmatrix} \hat{m} + \hat{b} \\ 2\hat{m} + \hat{b} \\ 3\hat{m} + \hat{b} \end{pmatrix} - \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix} \right\|^2$$

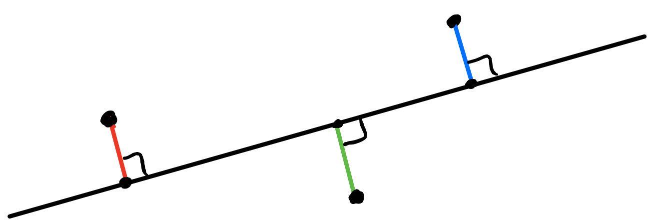
$$= (\hat{m} + \hat{b} - 0)^2 + (2\hat{m} + \hat{b} - 3)^2 + (3\hat{m} + \hat{b} - 2)^2$$

Picture:



Geometric Meaning: The OLS best fit line minimizes the sum of the squares of the vertical errors.

Why do we do it this way?
Why don't we try to minimize the sum of the orthogonal distances?



[This is called TLS (Total Least squares).] Well, maybe this would be more reasonable, but the math is much harder.

The OLS is:

- good enough
- easy to compute

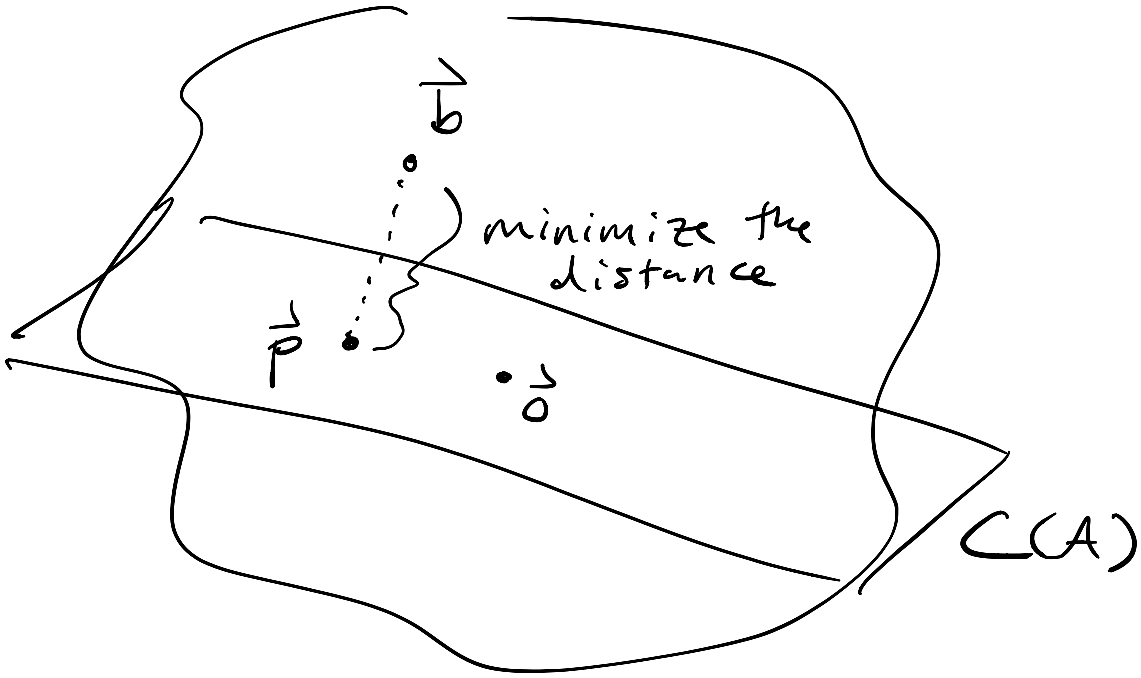
Next time I'll show you <u>why</u> it works.

Preview: Suppose $A\vec{x} = \vec{b}$ has no solution. This means that the point $\vec{b}$ is not in the so-called "column space" of A.

$$C(A) = \{\text{all vectors } A\vec{x}\}$$

$= \{$ all linear combinations of the columns of A $\}$.

Picture:



We will find the point $\vec{p}$ in the column space $C(A)$ that is <u>closest</u> to the point $\vec{b}$.

$\vec{p}$ is for "projection"