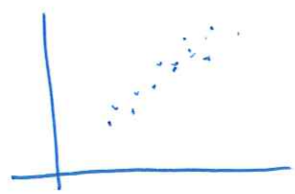
MATH 312
LECTURE 13 : Least-Squares

⌈ First, exercise on pages 104-106 ⌋

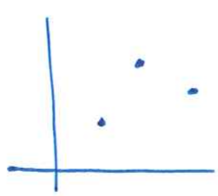
III.) Least-Squares

Let's say we want to find a model for some data, for example a line:



Of course, in most situations it is impossible to find a line that goes through all the points, but still the data looks like a line: and some lines seem better that other.

Ex:



$(x_1, y_1) = (1,1)$

$(x_2, y_2) = (2,3)$

$(x_3, y_3) = (3,2)$

$\to$ $y = mx + n$, m, n ?

Let's try to impose that the line goes through all three points:

$$\left.\begin{array}{l} y_1 = m x_1 + n \\ y_2 = m x_2 + n \\ y_3 = m x_3 + n \end{array}\right\} \Rightarrow \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \begin{bmatrix} n \\ m \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix};$$

In the example, $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} n \\ m \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$.

We can try to solve this system, but of course it doesn't have a solution! (just look at the picture: finding a solution (m,n) is finding a line through all three points).

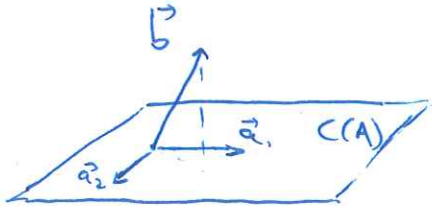
That is, in general, if we had k points,

$$A\vec{v} = \vec{b} \quad \text{with} \quad A = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_k \end{bmatrix}, \ \vec{v} = \begin{bmatrix} n \\ m \end{bmatrix}, \ \vec{b} = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix},$$

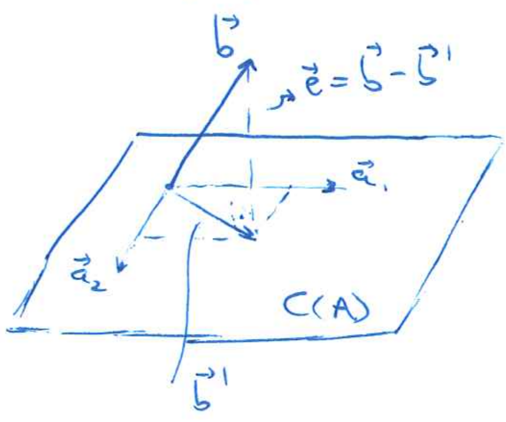
this system $A\vec{v} = \vec{b}$ has no solution.

$\Rightarrow$ That is, $\vec{b}$ is not in $C(A)$ $\|$.

Intuitively, the situation looks like the following picture (abstract)



We cannot obtain $\vec{b}$ as a linear combination of $\vec{a}_1, \vec{a}_2$ (more generally, the columns of A). The "best" we can do is to obtain a linear combination of $\vec{a}_1, \vec{a}_2$ given a vector $\vec{b}'$ as close as possible to $\vec{b}$: that is, to the projection of $\vec{b}$ onto $C(A)$.



• So now, we want to solve

$$\underbrace{\| A\vec{v} = \vec{b}' \|}_{\substack{\text{Always has a} \\ \text{solution} \\ (\text{since } \vec{b}' \in C(A))}} \text{, instead of } \underbrace{A\vec{v} = \vec{b}}_{\text{No solution.}}$$

The "error" will be $\vec{e} = \vec{b} - \vec{b}'$. What is the main characteristic in the picture? This "error" $\vec{e}$ is perpendicular to the "plane" (the subspace $C(A)$ in general).

• Thus we are saying that $\vec{e} \in C(A)^{\perp}$!

And we know that $C(A)^{\perp} = N(A^T)$ ‖

→ So, $\vec{e} \in N(A^T)$, i.e.,

$$A^T \vec{e} = \vec{0} \Rightarrow A^T(\vec{b} - \vec{b}') = \vec{0} \Rightarrow A^T \vec{b} = A^T \vec{b}',$$

where we said $\vec{b}' = A\vec{v}$.

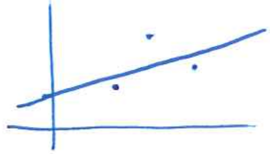
In summary, $\boxed{A^T A \vec{v} = A^T \vec{b}}$ Normal equations.

The solution $\vec{v}$ to this system is called the "least-squares" solution to $A\vec{v} = \vec{b}$.

• In our example,

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} n \\ m \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \quad \text{so}$$

$$\begin{bmatrix} n \\ m \end{bmatrix} = (A^T A)^{-1} A^T \vec{b} = \begin{bmatrix} 1 \\ 1/2 \end{bmatrix} \Rightarrow$$ The line of best fit in the least-square sense was



$y = \frac{x}{2} + 1$

$y = \frac{x}{2} + 1$.

• What are the errors?

First point: $(x_1, y_1) = (1, 1)$

The line gives $\tilde{y}_1 = \frac{1}{2} \cdot 1 + 1 = \frac{3}{2}$ $\Big\} \Rightarrow e_1 = y_1 - \tilde{y}_1 =$

$= 1 - \frac{3}{2} = \frac{-1}{2}$.

Second point: $(x_2, y_2) = (2, 3)$

$$e_2 = y_2 - \tilde{y}_2 = 3 - \left(\frac{2}{2} + 1\right) = 1$$

Third point: ...

Faster: $$\vec{e} = \vec{b} - \underbrace{\vec{b}'}_{\substack{\parallel \\ A\vec{v}}} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1/2 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1 \\ -1/2 \end{bmatrix}$$

• <u>Remark</u>: Is it a coincidence that the errors add up to 1? (someone asked this in class).

↳ Not in this case (But it is not always true for other models different than a line) (check why).

→ In a similar way "least-squares solutions" can be use to fit a plane, a parabola, etc. (see homework).

- Why the name "least-squares"?

Recall that when we fit the line, we make the following errors:

$$\left.\begin{aligned} e_1 &= y_1 - m x_1 - n \\ e_2 &= y_2 - m x_2 - n \\ &\vdots \\ e_k &= y_k - m x_k - n \end{aligned}\right\}$$

We can measure the total error in different ways:

$\|\vec{e}\|_1 = |e_1| + |e_2| + \dots + |e_k|$ ("L^1 norm")

$\|\vec{e}\| = \left(e_1^2 + e_2^2 + \dots + e_k^2\right)^{1/2}$ ("L^2 norm")

(this is the one we usually use. It gives the length of $\vec{e}$.

Goal? $\rightarrow \min\limits_{n,m} \|\vec{e}\|$ } That is, find n, m such that $\|\vec{e}\|$ is minimum.

But minimizing $\|\vec{e}\|$ is the same as minimizing $\|\vec{e}\|^2$ (same result but easier computations).

Notice that $\|\vec{e}\|^2 = e_1^2 + \dots + e_k^2 \equiv$ sum of the errors "squared".

Indeed, we could have used calculus to solve this problem:

$$\min_{\{n,m\}} \|\vec{e}\|^2 = \min_{\{n,m\}} \sum_{i=1}^{k} e_i^2 = \min_{\{n,m\}} \sum_{i=1}^{k} (y_i - m x_i - n)^2$$

So we have to find (n,m) such that the derivatives of $\|\vec{e}\|^2$ are zero (gradient equal to zero):

1) $$\frac{\partial \|\vec{e}\|^2}{\partial n} = -\sum_{i=1}^{k} 2(y_i - m x_i - n) = 2\left(nk + m\sum_{i=1}^{k} x_i - \sum_{i=1}^{k} y_i\right) = 0$$

2) $$\frac{\partial \|\vec{e}\|^2}{\partial m} = -\sum_{i=1}^{k} 2x_i(y_i - m x_i - n) = 2\left(n\sum_{i=1}^{k} x_i + m\sum_{i=1}^{k} x_i^2 - \sum_{i=1}^{k} x_i y_i\right) = 0$$

Thus, we have two equations (1), 2)) with two variables (n,m):

$$1), 2) \equiv \underbrace{\begin{bmatrix} k & \sum_{i=1}^{k} x_i \\ \sum_{i=1}^{k} x_i & \sum_{i=1}^{k} x_i^2 \end{bmatrix}}_{A^T A} \begin{bmatrix} n \\ m \end{bmatrix} = \underbrace{\begin{bmatrix} \sum_{i=1}^{k} y_i \\ \sum_{i=1}^{k} x_i y_i \end{bmatrix}}_{A^T \vec{b}}$$

check.