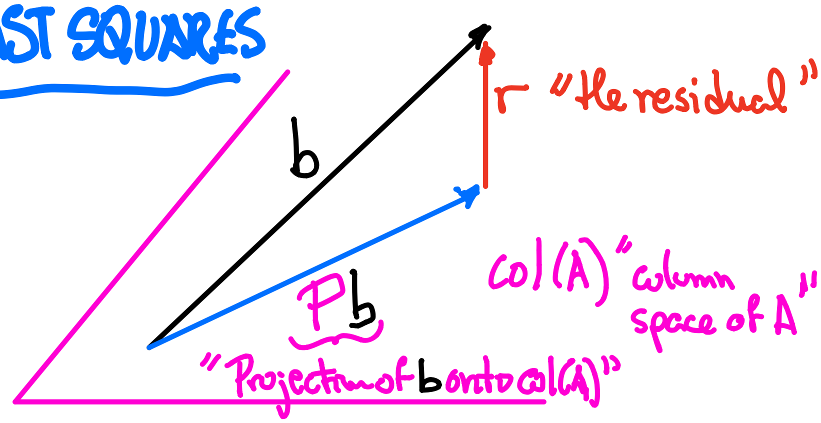
LEAST SQUARES



Suppose $A \in \mathbb{F}^{m \times n}$

$x \in \mathbb{F}^{n}$ $\quad b \in \mathbb{F}^{m}$

$r \equiv Ax - b \in \mathbb{F}^{m}$

if $r = 0 \Rightarrow b$ is in col(A), but what if it is not entirely in col(A) (see figure above)?

We choose a solution called an "estimate" $\hat{x}$ of x such that

$$\boxed{A\hat{x} = Pb}$$

$$\boxed{b = Pb + r}$$

So Pb is the projection of b onto $col(A)$ and $r \perp col(A)$, $r \perp Pb$.

Moreover, $\|r\|^2 = \|b - Pb\|^2 =$

$\|(I-P)b\|^2$ is minimized:

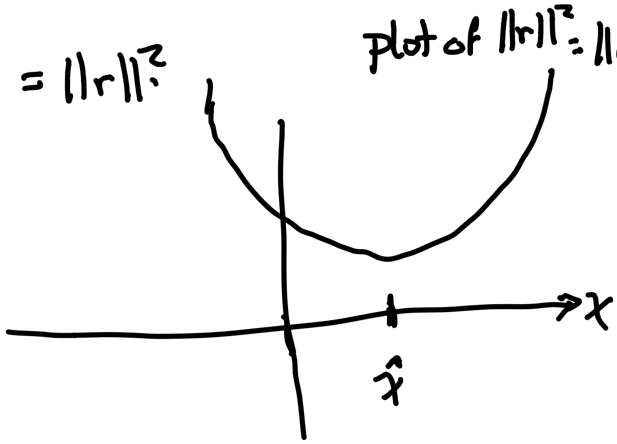
the smallest distance between b and $col(A)$.

Least Squares derives its name from the optimization point of view (this bit is optional)

Let $r = Ax - b$, so

$$r^T r = (Ax-b)^T(Ax-b)$$

$$= \|r\|^2.$$



the $\min_x \|r\|^2 \equiv \hat{x}$

we recognize $\|Ax-b\|^2 \equiv f(x)$

to be a quadratic (convex)

(norm) function of x. This function has a unique minimizer $\hat{x}$.

To find $\min_x \|r\|^2$, differentiate $f(x)$ with respect to x and set to 0:

$\hat{x}$ is given by $\min_x \nabla f(x) = 0$ explicitly:

$$\min_x A^*(Ax-b) = 0$$

$$\text{or} \quad A^*A\hat{x} = A^*b$$

Instead of optimization, let's consider the equivalent more geometric approach.

We said that $r = Ax-b$ and $r \perp \mathrm{col}(A)$

which is $a_1^* r = 0,\ a_2^* r = 0 \ldots\ a_n^* r = 0$ (\$)

where a_i are the columns of A.

So (\$) can be written compactly

$$A^*r = A^*(b-Ax) = 0 \quad (\#)$$

Call the solution of (#) $\hat{x}$:

$$(\exists) \quad \boxed{A^*A\hat{x} = A^*b}$$

The "normal" Equations

Solving from $\hat{x}$ from (∃):

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

satisfies

$$A\hat{x} = Pb = \underbrace{A(A^*A)^{-1}A^*}_{P}b$$

//

ex)

Let's work through a problem which can be easily (intuitively) understood:

for $a_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $a_2 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$ and $b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$

so $b \in \mathbb{R}^3$, so are a_1 & a_2, but b is not a linear combination of a_1 and a_2, i.e.,

$$b \neq x_1 a_1 + x_2 a_2$$

but maybe we could find

$$Pb = \hat{x}_1 a_1 + \hat{x}_2 a_2$$

so that $r = b - Pb = (I-P)b$. So,

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} \qquad b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \qquad \begin{matrix} m = 3 \\ n = 2 \\ x \in \mathbb{R}^2 \\ b \in \mathbb{R}^3 \end{matrix}$$

Find x s.t. $Ax = b$ (no solution)

$A^T A \hat{x} = A^T b$ the normal equations.

Col(A) is a plane, say the x-y plane

$$A^T A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 5 & 13 \end{bmatrix}$$

$$\hat{x} = (A^T A)^{-1} A^T b = \underbrace{\begin{bmatrix} 13 & -5 \\ -5 & 2 \end{bmatrix}}_{(A^T A)^{-1}} \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$Pb = A\hat{x} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}$$

$$b - Pb = r = \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix} = b - Pb = (I - P)\, b$$

$\|r\|_2 = 6$ "the error in the 2-norm"

REMARKS ON LEAST SQUARES

Rmk1: Spse $b \in \mathrm{Col}(A)$ i.e. $Ax = b$

Then $Pb = A(A^TA)^{-1}A^TAx = Ax = b$

Rmk2: Spse $b \perp \mathrm{col}(A)$, i.e. $A^Tb = 0$

then b has 0 as a projection.

i.e. $b \in \mathrm{Null}(A^T)$

$$Pb = A(A^TA)^{-1}A^Tb = A(A^TA)^{-1}0 = 0 \quad //$$

Rmk3: When $A \in \mathbb{R}^{n \times n}$ & invertible, $\mathrm{Col}(A)$ is the whole space (i.e. $\mathrm{rank}(A) = n$). Every vector projects to itself:

$$Pb = b$$

$$\hat{x} = x$$

Rmk4: $A \in \mathbb{R}^{n \times n}$

$$(A^TA)^{-1} = A^{-1}(A^T)^{-1}$$

will exist if A is invertible.

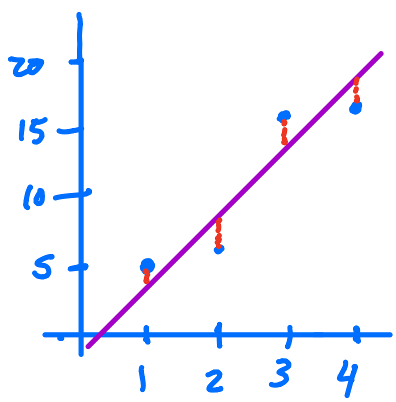
Rmk 5: Spse A has only 1 column, a.

Then $A^T A = a^T a$

and $\hat{x} = \dfrac{a^T b}{a^T a}$ //

Ex) Suppose you have data

x_i	y_i
1	5
2	7
3	17
4	18



You would like to "model" this data by a straight line. Hence

$$y = \alpha x + \beta$$

Set up $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}$

so that $y_i = \alpha x_i + \beta$ is the "model" in discrete form. The 2 parameters in the model are α, β.

Let $v = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

so $Av = w + r$

$$w = \begin{bmatrix} 5 \\ 7 \\ 17 \\ 18 \end{bmatrix}$$

An estimate of $\hat{v}$ is found by using the normal equations

$$A\hat{v} = Pw$$

$$\therefore \hat{v} = [A^T A]^{-1} A^T w$$

$$A^T A = \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix}; \quad [A^T A]^{-1} = \begin{bmatrix} 2/10 & 5/10 \\ 5/10 & 15/10 \end{bmatrix}$$

$$\hat{v} = \frac{1}{10}\begin{bmatrix} 2 & 5 \\ 5 & 15 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 5 \\ 7 \\ 17 \\ 18 \end{bmatrix} = \begin{bmatrix} 4.9 \\ -0.5 \end{bmatrix}$$

Hence a least squares fit is

$$y = 4.9x - \frac{1}{2}.$$

How good is the fit?

$$\|r\|_2^2 = \sum_{i=1}^{4} r_i^2 = \left(4.9 \cdot 1 - \tfrac{1}{2} - 5\right)^2 + \left(4.9 \cdot 2 - \tfrac{1}{2} - 7\right)^2 + \left(4.9 \cdot 3 - \tfrac{1}{2} - 17\right)^2 + \left(4.9 \cdot 4 - \tfrac{1}{2} - 18\right)^2 = 14.7$$

$\therefore$ error $= \|r\|_2 \approx 3.8$

//