Linear Algebra

10.11.2023

Afonso Bandeira

* Please take a look at the typed lecture notes

Question: What do we do when we have a linear system without solutions?

linear system $Ax = b$ but there is no x s.t. $Ax = b$.

e.g. too many equations.

Idea: try to find x as close as possible to being a solution.

→ find x s.t. Ax is as close to b as possible.

$\hat{x}$ solution of $\min_{x \in \mathbb{R}^n} \|Ax - b\|$

$$\hat{x} = \operatorname{argmin}_{x \in \mathbb{R}^n} \|Ax - b\|$$

(A in $m \times n$)

what is the set of possible vectors $y \in \mathbb{R}^m$ s.t. $y = A\hat{x}$ for some $\hat{x} \in \mathbb{R}^m$?

It's $C(A)$

$\hookrightarrow = a_1 \hat{x}_1 + \cdots + a_n \hat{x}_n$

$A = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix}$

So if $\hat{x}$ is a solution of $\min_{x \in \mathbb{R}^n} \|Ax - b\|$

then $p = A\hat{x}$ is the projection of b on $C(A)$.

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$$
$$\|x\|^2 = x^T x = x \cdot x = \langle x, x \rangle$$

then $\hat{x}$ is a solution of the "normal equations"

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

$$A^T A \hat{x} = A^T b \qquad \text{(normal equations)}$$

* if $A^T A$ is invertible then $\hat{x} = (A^T A)^{-1} A^T b$

Do the normal equations always have solutions?
(even when $A^T A$ is not invertible)

Yes, because $C(A^T A) = C(A^T)$
(prove as HW)

Theorem: for A $m \times n$ and $b \in \mathbb{R}^m$ a solution $\hat{x}$ of $\min_{x \in \mathbb{R}^n} \|Ax - b\|$ is also a solution of

$$A^T A \hat{x} = A^T b \qquad \text{(and vice-versa)}$$

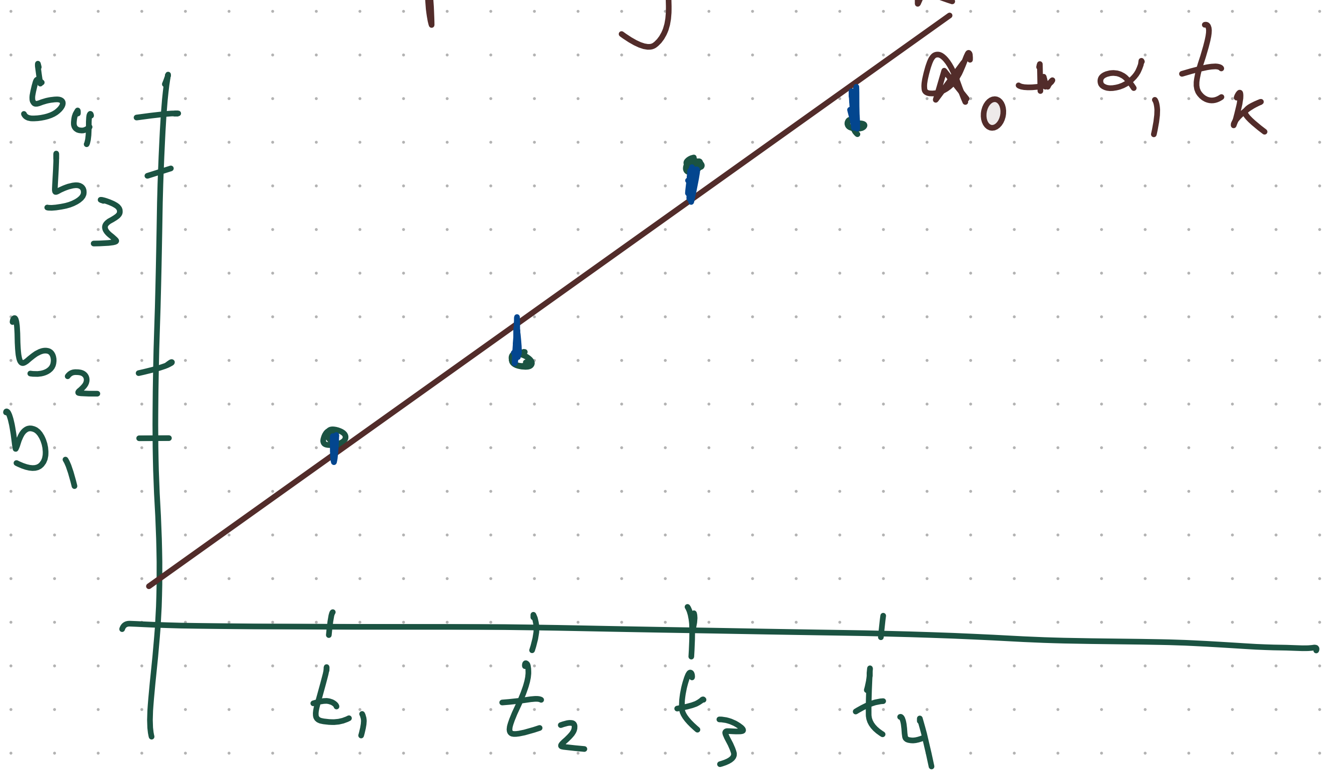
and if A has ind col. then $\hat{x} = (A^TA)^{-1}A^Tb$.

Linear Regression:

data points $(t_1, b_1), (t_2, b_2), \cdots, (t_m, b_m)$

(e.g. b_k is temp at time t_k in some experiment)

Let's say we suspect b_k has close to a linear dependecy on t_k



(#) $\quad b_k \approx \alpha_0 + \alpha_1 t_k$ for some $\alpha_0, \alpha_1 \in \mathbb{R}$

idea: find the best fit

$$(!!)\quad \min_{\alpha_0,\alpha_1} \sqrt{\sum_{k=1}^{m} \Big(\underbrace{b_k}_{\text{"target"}} - \underbrace{(\alpha_0+\alpha_1 t_k)}_{\text{"prediction"}}\Big)^2}$$

$$(\#)\quad \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \approx \begin{bmatrix} 1 & t_1 \\ \vdots & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \left(= \begin{bmatrix} \alpha_0+\alpha_1 t_1 \\ \vdots \\ \alpha_0+\alpha_1 t_m \end{bmatrix} \right)$$

$$\mathbb{R}^m \ni b \approx A \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \quad \text{when } A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}$$

$$(!!)\quad \min_{\alpha_0,\alpha_1} \left\| b - A \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \right\|^2$$

We can solve (!!) by solving

$$A^T A \begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{bmatrix} = A^T b$$

if A^TA is inv then

$$\begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{bmatrix} = (A^TA)^{-1} A^T b$$

"

$$A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \qquad A^TA = \begin{bmatrix} m & \sum_{k=1}^{m} t_k \\ \sum_{k=1}^{m} t_k & \sum_{k=1}^{m} t_k^2 \end{bmatrix}$$

$$[1 \; \cdots \; 1]\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = m \qquad [t_1 \; \cdots \; t_m]\begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix} = \sum t_k^2$$

when would it be that A doesn't have ind. col.

it $t_k = t \;\; \forall_k$ (all t_k's are equal)

degenerate situation in Linear regression.

$$A^TA = \begin{bmatrix} m & mt \\ mt & mt^2 \end{bmatrix}$$

So lin dep.

HW: Is this the only situation in which A^TA is not inv ?

if A^TA is inv

$$\begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{bmatrix} = \begin{bmatrix} m & \sum t_k \\ \sum t_k & \sum t_k^2 \end{bmatrix}^{-1} A^T b$$

if $\sum t_k = 0$ (via change of variables $t^{new} = t - \frac{1}{m}\sum t_k$

$$\begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{bmatrix} = \begin{bmatrix} m & 0 \\ 0 & \sum t_k^2 \end{bmatrix}^{-1} A^T b$$

$$= \begin{bmatrix} 1/m & 0 \\ 0 & \frac{1}{\sum t_k^2} \end{bmatrix} \begin{bmatrix} \sum b_k \\ \sum t_k b_k \end{bmatrix}$$

(HW: inv diag matrices is easy)

$$= \begin{bmatrix} \frac{1}{m} \sum b_k \\ \frac{\sum t_k b_k}{\sum t_k^2} \end{bmatrix}.$$

Fit a Parabola

$$b_k \approx \alpha_0 + \alpha_1 t_k + \alpha_2 t_k^2$$

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \quad \alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

$$b \approx \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ & \vdots & \\ 1 & t_m & t_m^2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

Now do least squares.

HW: → when is it that $A^T A$ is not invertible for this A?

→ how could we make $A^T A$ a diag. matrix?

CS Lens:

$$b_k \approx 3\alpha_0 + (t_k - 5)\alpha_1 + (t_k^2 + t_k)\alpha_2$$

CS Lens: Kernel Methods