→ last time
→ FTLA and linear least squares; derived normal equations.

10/20/2021

Again, our setup is $\underline{\underline{A}} \in \mathbb{R}^{m \times n}$, $m > n$,

and we want to minimize

$$\| \underline{\underline{A}}\, \underline{x} - \underline{b} \|_2^2 .$$

First step... decompose $\underline{b}$ as

$$\underline{b} = \underline{b}_R + \underline{b}_N ,$$

where $\underline{b}_R \in R(\underline{\underline{A}})$ and $b_N \in N(\underline{\underline{A}}^T)$.

Now, let's expand:

$$\| \underline{\underline{A}}\, \underline{x} - \underline{b} \|_2^2 = \| \underline{\underline{A}}\, \underline{x} - \underline{b}_R - \underline{b}_N \|_2^2$$

$$= \| \underline{\underline{A}}\, \underline{x} - \underline{b}_R \|_2^2 - (\underline{\underline{A}}\, \underline{x} - \underline{b}_R)^T \underline{b}_N$$

$$- \underline{b}_N^T (\underline{\underline{A}}\, \underline{x} - \underline{b}_R) + \| b_N \|_2^2 .$$

Note that $\underline{b}_N^T(\underline{\underline{A}}\, \underline{x} - \underline{b}_R) = (\underline{\underline{A}}\, \underline{x} - \underline{b}_R)^T \underline{b}_N = 0$

since $\underline{b}_N \in N(\underline{\underline{A}}^T)$ and $\underline{\underline{A}}\, \underline{x} - \underline{b}_R \in R(\underline{\underline{A}})$,

and $R(\underline{\underline{A}}) \perp N(\underline{\underline{A}}^T)$.

So we have $\|\underline{\underline{A}}\,\underline{x}-\underline{b}\|_2^2 = \|\underline{\underline{A}}\,\underline{x}-\underline{b}_R\|_2^2 + \|b_N\|_2^2$.

$\|\underline{b}_N\|_2^2$ is independent of $\underline{x}$, so minimizing $\|\underline{\underline{A}}\,\underline{x}-\underline{b}\|_2^2$ is the same as minimizing $\|\underline{\underline{A}}\,\underline{x}-\underline{b}_R\|_2^2$.

Note that $\|\underline{\underline{A}}\,\underline{x}-\underline{b}_R\|_2^2 \geq 0$, and there actually must be some $\underline{x}^*$ so that $\|\underline{\underline{A}}\,\underline{x}^*-\underline{b}_R\|_2^2 = 0$ since $\underline{b}_R \in R(\underline{\underline{A}})$.

It is not clear how to compute $\underline{b}_R$, however, let's show that:

$\underline{\underline{A}}^T\underline{\underline{A}}\,\underline{x}^* = \underline{\underline{A}}^T\underline{b} \iff \underline{\underline{A}}\,\underline{x}^* = \underline{b}_R$.

($\Rightarrow$) If $\underline{\underline{A}}^T\underline{\underline{A}}\,\underline{x}^* = \underline{\underline{A}}^T\underline{b}$, we plug in $\underline{b} = \underline{b}_R + \underline{b}_N$ and get

$\underline{\underline{A}}^T\underline{\underline{A}}\,\underline{x}^* = \underline{\underline{A}}^T\underline{b}_R$, implying:

$$\underline{\underline{A}}^T(\underline{\underline{A}}\,\underline{x}^*-\underline{b}_R) = \underline{0}.$$

This implies $\underline{\underline{A}}\,\underline{x}^* - \underline{b}_R \in N(\underline{\underline{A}}^T)$.

Since $\underline{\underline{A}}\,\underline{x}^* - \underline{b}_R$ is also in $R(\underline{\underline{A}})$,

the only possibility is that $\underline{\underline{A}}\,\underline{x}^* - \underline{b}_R = \underline{0}$.

($\Leftarrow$) If $\underline{\underline{A}}\,\underline{x}^* = \underline{b}_R$, then $\underline{\underline{A}}^T\underline{\underline{A}}\,\underline{x}^* = \underline{\underline{A}}^T\underline{b}_R$.

Note that $\underline{\underline{A}}^T\underline{b}_N = \underline{0}$, so

$$\underline{\underline{A}}^T\underline{\underline{A}}\,\underline{x}^* = \underline{\underline{A}}^T\underline{b}_R + \underline{\underline{A}}^T\underline{b}_N = \underline{\underline{A}}^T\underline{b}.$$

Geometry of the Normal Equations

We have: $\underline{\underline{A}}^T\underline{\underline{A}}\,\underline{x}^* = \underline{\underline{A}}^T\underline{b}$,

which can be rewritten as:

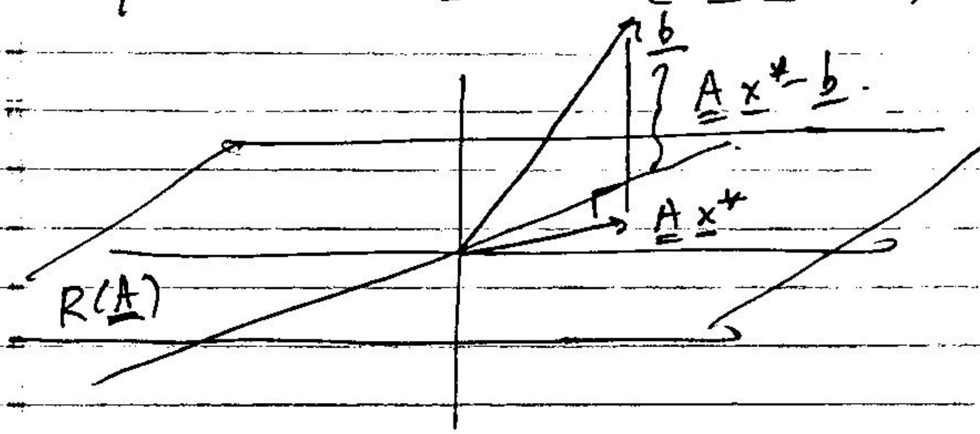
$$\underline{\underline{A}}^T(\underline{\underline{A}}\,\underline{x}^* - \underline{b}) = \underline{0}.$$

Take $\underline{y}$ arbitrary and consider

$$\underline{y}^T\underline{\underline{A}}^T(\underline{\underline{A}}\,\underline{x}^* - \underline{b}) = 0$$

$$\Leftrightarrow (\underline{\underline{A}}\,\underline{y})^T(\underline{\underline{A}}\,\underline{x}^* - \underline{b}) = 0.$$

Since y is arbitrary, this statement says $R(\underline{\underline{A}}) \perp (\underline{\underline{A}}\,\underline{x}^* - \underline{b})$.



<u>Note</u> that since $\underline{\underline{A}} \in \mathbb{R}^{m\times n}$, $m > n$, i.e. $\underline{\underline{A}}$ is "tall and skinny," the $R(\underline{\underline{A}})$ is a low dimensional subspace of $\mathbb{R}^m$, hence it is very likely $\underline{b}$ is not in the $R(\underline{\underline{A}})$.

<u>Observation:</u> Since $\underline{\underline{A}}^T\underline{\underline{A}}\,\underline{x}^* = \underline{\underline{A}}^T\underline{b}$, assuming $\underline{\underline{A}}^T\underline{\underline{A}}$ is invertible, we have

$$\underline{x}^* = (\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T\underline{b} \iff \underline{\underline{A}}\,\underline{x}^* = \underbrace{\underline{\underline{A}}(\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T}\underline{b}$$

what type of matrix is $\underline{\underline{A}}(\underline{\underline{A}}^T\underline{\underline{A}})^{-1}\underline{\underline{A}}^T$?

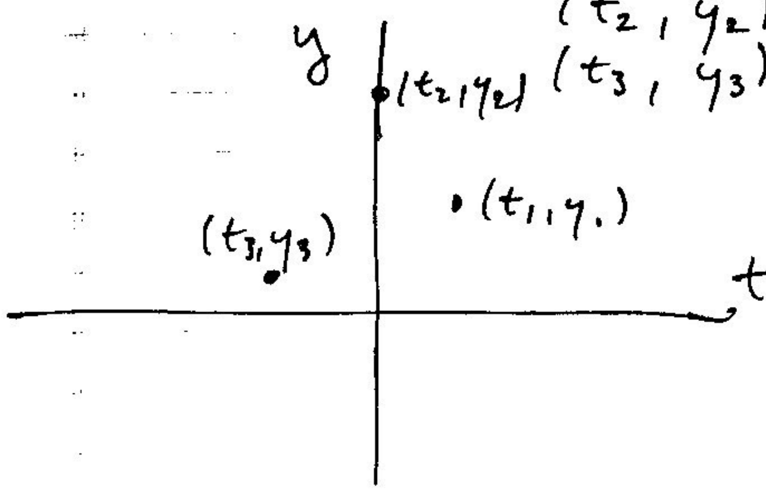
<u>Check:</u> $\underline{P} = \underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T$ is an <u>orthogonal</u> <u>projector</u> onto $R(\underline{A})$.

<u>Question:</u> Why is $\underline{A}^T\underline{A}$ likely invertible? <u>Hint:</u> see that

$$N(\underline{A}) = N(\underline{A}^T\underline{A}).$$

If $N(\underline{A})$ contains more than just the zero vector, then the very few number of columns of $\underline{A}$ are linearly dependent, which is "not likely."

<u>Ex:</u> <u>data:</u> $(t_1, y_1) = (1, 2)$
$(t_2, y_2) = (0, 4)$
$(t_3, y_3) = (-2, 1)$.



<u>goal:</u> fit $f(t) = \beta t + \gamma$ to the data

We can set up.

$$\underline{b} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \quad \underline{\underline{A}} = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ t_3 & 1 \end{pmatrix}, \quad \underline{x} = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}.$$

$$\underline{\underline{A}}^T \underline{\underline{A}} = \begin{pmatrix} 1 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & -1 \\ -1 & 3 \end{pmatrix}.$$

$$\underline{\underline{A}}^T \underline{b} = \begin{pmatrix} 1 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 7 \end{pmatrix}.$$

So, $\underline{x}^* = \begin{pmatrix} 5 & -1 \\ -1 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 7 \end{pmatrix}.$

We solve $\begin{pmatrix} 5 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 7 \end{pmatrix}$

$$\rightsquigarrow \begin{pmatrix} \beta^* \\ \gamma^* \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{5}{2} \end{pmatrix}.$$

So, the line of best fit is

$$f(t) = \frac{1}{2} x + \frac{5}{2}.$$

Ex: data: $(t_1, y_1) = (0, 2)$
$(t_2, y_2) = (-1, 5)$
$(t_3, y_3) = (-2, 4)$
$(t_4, y_4) = (1, 1)$

goal: fit a parabola $g(t) = \gamma t^2 + \alpha t + \beta$ to the data.

$$\underline{b} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 4 \\ 1 \end{pmatrix}.$$

$$\underline{A} = \begin{pmatrix} t_1^2 & t_1 & 1 \\ t_2^2 & t_2 & 1 \\ t_3^2 & t_3 & 1 \\ t_4^2 & t_4 & 1 \end{pmatrix} \qquad \underline{x} = \begin{pmatrix} \gamma \\ \alpha \\ \beta \end{pmatrix}.$$

$$\Rightarrow \quad \underline{A} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & -1 & 1 \\ 4 & -2 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Solve normal equations for $\underline{x}^* = \begin{pmatrix} \gamma^* \\ \alpha^* \\ \beta^* \end{pmatrix} \ldots$