10/18/2021

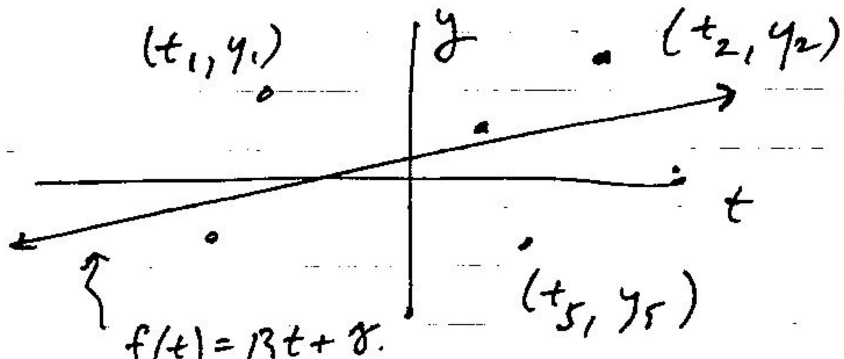
→ last time: finished FTLA proof and began linear least squares.

---

Linear least squares

1 setup: $(t_i, y_i)$, $i = 1, \dots, m$ ← "data".

goal: fit a line $f(t) = \beta t + \gamma$ to the data.




we want to find $\beta, \gamma$ so that $f(t_i) \approx y_i$

$$\Leftrightarrow \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}}_{= \underline{b}} \approx \underbrace{\begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{pmatrix}}_{= \underline{\underline{A}}} \underbrace{\begin{pmatrix} \beta \\ \gamma \end{pmatrix}}_{= \underline{x}}.$$

To make this statement more precise...

$$\min_{\underline{x}=\binom{\beta}{\gamma}} \| \underline{\underline{A}}\,\underline{x} - \underline{b} \|_2^2 = \min_{\underline{x}} r(\underline{x}).$$

How do we solve this? (let's assume $\underline{\underline{A}} \in \mathbb{R}^{m \times n}$, $m > n$)

At a minimizer $\underline{x}^*$, it is necessary that $\nabla r(\underline{x}^*) = \underline{0}$.

let's expand $r(\underline{x})$:

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

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

$$= \underline{x}^T \underline{\underline{A}}^T \underline{\underline{A}}\,\underline{x} - 2\,\underline{x}^T \underline{\underline{A}}^T \underline{b} + \underline{b}^T \underline{b}.$$

let's define $\underline{\underline{C}} = \underline{\underline{A}}^T \underline{\underline{A}}$ and $\underline{d} = \underline{\underline{A}}^T b$, so

$$r(\underline{x}) = \underline{x}^T \underline{\underline{C}}\,\underline{x} - 2\,\underline{x}^T \underline{d} + \underline{b}^T \underline{b}.$$

let's write out the quadratic term in component form...

$$(\underset{=}{C}\,\underline{x})_k = \sum_{i=1}^{n} C_{ki} x_i \text{ , and:}$$

$$\underline{x}^T \underset{=}{C}\,\underline{x} = \sum_{k=1}^{n} x_k (\underset{=}{C}\,\underline{x})_k = \sum_{k=1}^{n}\sum_{i=1}^{n} x_k C_{ki} x_i .$$

Now, let's compute the contribution of this quadratic term to the jth component of $\nabla r$.

$$\frac{\partial}{\partial x_j}\left(\underline{x}^T \underset{=}{C}\,\underline{x}\right) = \frac{\partial}{\partial x_j}\left(\sum_{k=1}^{n}\sum_{i=1}^{n} x_k C_{ki} x_i\right)$$

$$\begin{pmatrix}\text{product}\\ \text{rule}\end{pmatrix} = \sum_{k=1}^{n}\sum_{i=1}^{n}\left(\frac{\partial x_k}{\partial x_j} C_{ki} x_i + x_k C_{ki} \frac{\partial x_i}{\partial x_j}\right)$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{n}\left(\delta_{kj} C_{ki} x_i + x_k C_{ki} \delta_{ij}\right)$$

$$\left(\text{where } \delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i\neq j .\end{cases}\right)$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{n} \delta_{kj} C_{ki} x_i + \sum_{k=1}^{n}\sum_{i=1}^{n} \delta_{ij} C_{ki} x_k$$

$$= \sum_{i=1}^{n} C_{ji} x_i + \sum_{k=1}^{n} C_{kj} x_k$$

Note that $\underline{\underline{C}} = \underline{\underline{A}}^T \underline{\underline{A}}$ is symmetric, i.e.

$$C_{ij} = C_{ji}.$$

So, we get $\frac{\partial}{\partial x_j} (\underline{x}^T \underline{\underline{C}}\, \underline{x}) = 2 \sum_{i=1}^{n} C_{ji} x_i$

$$= 2 \left( \underline{\underline{A}}^T \underline{\underline{A}}\, \underline{x} \right)_j .$$

Similarly, one can show:

$$\frac{\partial}{\partial x_j} \left( 2\, \underline{x}^T \underline{d} \right) = 2 d_j = 2 \left( \underline{\underline{A}}^T \underline{b} \right)_j .$$

and for sure we have: $\frac{\partial}{\partial x_j} (\underline{b}^T \underline{b}) = 0$,

so: $\nabla r(\underline{x}^*) = \underline{0} \iff \underline{\underline{A}}^T \underline{\underline{A}}\, \underline{x}^* - \underline{\underline{A}}^T \underline{b} = \underline{0}.$

Normal Equations $\underline{\underline{A}}^T \underline{\underline{A}}\, \underline{x}^* = \underline{\underline{A}}^T \underline{b}$.

Deriving the normal equations in this way seems messy... let's instead use FTLA...