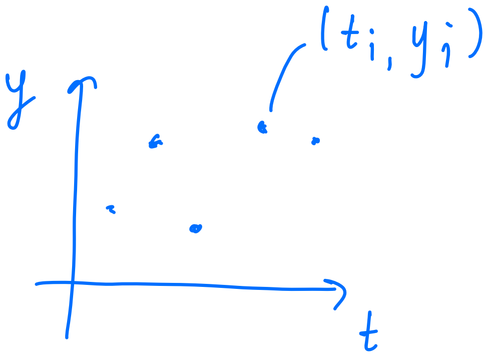
Least squares problems

Lecture 07

September 21, 2021

Beck, sections 3.1-3.2

Fitting data



Given data: $\{(t_i, y_i)\}_{i=1}^{M}$

Goal: fit a model to data

"input": t

"output": y

"learn" model $y = f(t)$ that fits data

t: generally in $\mathbb{R}^n$

y: scalar (for simplicity)

Idea: "linear" approximation

- prescribe a "dictionary" of plausible functions/features/models.

  $\{ \phi_1(t) \ldots \phi_N(t) \}$

- enforce agreement on data using a <u>linear</u> combination of dictionary elements

I.e.: for some scalars $x_1 \ldots x_N$, enforce

$$x_1 \phi_1(t_m) + x_2 \phi_2(t_m) + \ldots + x_N \phi_N(t_m) = y_m$$

$$m = 1 \ldots M.$$

"find" x_j

All conditions:

$$\left.\begin{aligned} x_1 \phi_1(t_1) + \ldots + x_N \phi_N(t_1) &= y_1 \\ x_1 \phi_1(t_2) + \ldots + x_N \phi_N(t_2) &= y_2 \\ &\vdots \\ x_1 \phi_1(t_M) + \ldots + x_N \phi_N(t_M) &= y_M \end{aligned}\right\} \quad \underline{\underline{A}}\, \underline{x} = \underline{b}$$

$$\underline{b} = \begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix} \quad \underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}, \quad (\underline{\underline{A}})_{i,j} = \phi_j(t_i)$$

$$i = 1 \ldots M$$
$$j = 1 \ldots N$$

"Solve" $\underline{\underline{A}}\,\underline{x} = \underline{b}$ for $\underline{x}$, $\underline{\underline{A}} \in \mathbb{R}^{M \times N}$

E.g., if $M < N$

$\Rightarrow$ $\ker(\underline{\underline{A}})$ non-empty and non-trivial

$\Rightarrow$ infinitely many solutions.

Because of this, assume $M \geq N$.

If $M \geq N$: it's usually not possible to find $\underline{x}$ s.t. $\underline{\underline{A}}\,\underline{x} = \underline{b}$

$$\begin{pmatrix} & & \\ & \underline{\underline{A}} & \\ & & \end{pmatrix} \begin{pmatrix} \underline{x} \end{pmatrix} = \begin{pmatrix} \\ \underline{b} \\ \\ \end{pmatrix}$$

So: instead of asking for $\underline{\underline{A}}\,\underline{x} = \underline{b}$, ask for "$\underline{\underline{A}}\,\underline{x} \simeq \underline{b}$".

The (linear) least squares problem

Instead of $\underline{\underline{A}}\,\underline{x} = \underline{b}$, request a small "residual",

i.e. try to minimize $\|\underline{\underline{A}}\,\underline{x} - \underline{b}\|$.

Choose $\|\cdot\| = \|\cdot\|_2$:

Find $\underline{x}$ s.t. $\|\underline{\underline{A}}\,\underline{x} - b\|_2$ is minimized.

$$\Updownarrow$$

Find $\underline{x}$ s.t. $\|\underline{\underline{A}}\,\underline{x} - \underline{b}\|_2^2$ is minimized.

$$\Updownarrow$$

$$\underset{\underline{x} \in \mathbb{R}^N}{\operatorname{argmin}} f(\underline{x}), \quad f(\underline{x}) = \|\underline{\underline{A}}\,\underline{x} - \underline{b}\|_2^2$$

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

Existence + uniqueness of solutions

$\underset{\underline{x} \in \mathbb{R}^N}{\operatorname{argmin}} f(\underline{x})$, $\quad f(\underline{x}) = \underline{x}^T \underline{\underline{A}}^T \underline{\underline{A}} x - 2\underline{b}^T \underline{\underline{A}}\,\underline{x} + \|b\|_2^2$

Recall: if $\nabla^2 f \succ \underline{\underline{0}} \rightarrow$ exactly 1 solution, global minimum

$\nabla f = 2\underline{\underline{A}}^T \underline{\underline{A}} x - 2\underline{\underline{A}}^T \underline{b}$

$\nabla^2 f = 2\underline{\underline{A}}^T \underline{\underline{A}}$

HW #2.

Recall: $\underline{\underline{A}}^T \underline{\underline{A}} \succ \underline{\underline{0}}$ iff $\underline{\underline{A}}^T$ has full row rank

$\Longrightarrow$ $\underline{\underline{A}}^T \underline{\underline{A}} \succ \underline{\underline{0}}$ iff $\underline{\underline{A}}$ has full column rank. $(= N)$

$\Longrightarrow$ if $\underline{\underline{A}} \in \mathbb{R}^{M \times N}$ has full column rank $(= N)$, then the least squares problem:

$$\operatorname*{argmin}_{\underline{x} \in \mathbb{R}^N} f(\underline{x})$$

has a unique (globally minimum) solution given by $\nabla f = 2\underline{\underline{A}}^T \underline{\underline{A}}\,\underline{x} - 2\underline{\underline{A}}^T \underline{b} = \underline{0}$.

$$\Downarrow$$

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

~~$\underline{\underline{A}}$ has full column rank~~

$$\underline{x} = (\underline{\underline{A}}^T \underline{\underline{A}})^{-1} \underline{\underline{A}}^T \underline{b} \quad (\text{since } \underline{\underline{A}}^T \underline{\underline{A}} \succ \underline{\underline{0}})$$

If $\underline{\underline{A}}$ has full column rank, the least squares problem has a single, unique solution $\underline{x} = (\underline{\underline{A}}^T \underline{\underline{A}})^{-1} \underline{\underline{A}}^T \underline{b}$, which is the solution to the <u>normal equations</u>: $\underline{\underline{A}}^T \underline{\underline{A}}\,\underline{x} = \underline{\underline{A}}^T \underline{b}$

Back to fitting data: compute

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

$$= \underset{\underline{x} \in \mathbb{R}^N}{\operatorname{argmin}} \| \underline{\underline{A}}\,\underline{x} - \underline{b} \|_2^2$$

model: $f(t) = \sum_{j=1}^{N} (\underline{x}_*)_j \, \phi_j(t)$