Lecture #19

02/24/2023

Example (continued):

Rewrite minimizing the MSE over $\begin{pmatrix} v \\ \underline{\beta} \end{pmatrix}$ as a standard/ordinary least squares problem:

$$\min_{\underline{\theta} := \begin{pmatrix} v \\ \underline{\beta} \end{pmatrix} \in \mathbb{R}^{n+1}} \| A\underline{\theta} - \underline{y} \|_2^2$$

unknown parameter vector of size $(n+1) \times 1$

next pg.

where $$\underbrace{\underline{y} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{pmatrix}}_{\text{known } N\times 1 \text{ vector}}, \quad \underbrace{A = \begin{bmatrix} 1 & (\underline{x}^{(1)})^T \\ 1 & (\underline{x}^{(2)})^T \\ \vdots & \vdots \\ 1 & (\underline{x}^{(N)})^T \end{bmatrix}}_{N \times (n+1)}$$

$$\Updownarrow \qquad\qquad\qquad \Updownarrow$$

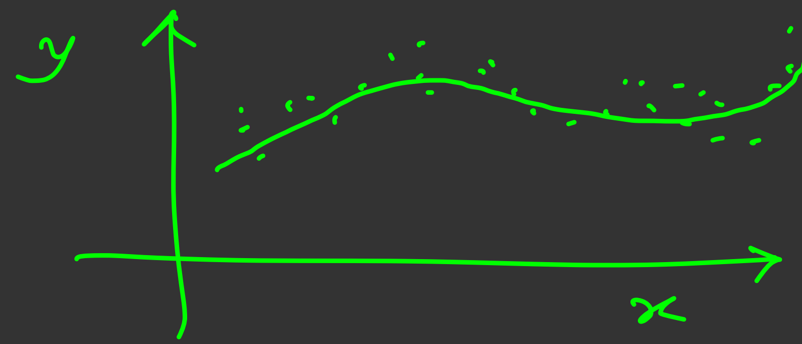
$$\underline{y} \in \mathbb{R}^N \qquad\qquad A \in \mathbb{R}^{N \times (n+1)}$$

Then: $\underline{\theta} = A \setminus \underline{y}$

Example: Polynomial regression in 1D

$$\boxed{\hat{f}(x) = \theta_1 + \theta_2 x + \dots + \theta_p x^{p-1}}$$

still linear in the parameters $\theta_1, \theta_2, \dots, \theta_p$



Problem:

Compute the $p \times 1$ vector $\underline{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}$ that minimizes the mean square error:

$$MSE = \frac{1}{N} \left\{ \left(y^{(1)} - \hat{f}(x^{(1)})\right)^2 + \left(y^{(2)} - \hat{f}(x^{(2)})\right)^2 + \dots + \left(y^{(N)} - \hat{f}(x^{(N)})\right)^2 \right\}$$

Now do the same thing as before: rewrite this minimization of MSE over $\underline{\theta} \in \mathbb{R}^p$ in the standard/ordinary least squares form:

$$\min_{\underline{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}} \| A \underline{\theta} - \underline{y} \|_2^2$$

where $\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \in \mathbb{R}^N$ and $A = \underbrace{\begin{bmatrix} 1 & x^{(1)} & (x^{(1)})^2 & \cdots & (x^{(1)})^{p-1} \\ 1 & x^{(2)} & (x^{(2)})^2 & \cdots & (x^{(2)})^{p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x^{(N)} & (x^{(N)})^2 & \cdots & (x^{(N)})^{p-1} \end{bmatrix}}_{N \times p \text{ matrix}}$

Vandermonde-like
tall rectangular matrix

<u>Example:</u> Regression/function approximation by a weighted sum of arbitrary functions:

$$\hat{f}(x) = \theta_1 f_1(x) + \theta_2 f_2(x) + \dots + \theta_p f_p(x)$$

So the previous example is a special case of above:

$$f_1(x) \equiv 1, \quad f_2(x) \equiv x, \quad f_3(x) \equiv x^2, \dots, f_p(x) \equiv x^{p-1}.$$

Still linear in $\underline{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix} \in \mathbb{R}^p$

Again, $r^{(i)} = y^{(i)} - \hat{f}(x^{(i)})$

$$\underset{\underline{\theta}}{\text{minimize}} \quad \| A\underline{\theta} - \underline{y} \|_2^2$$

$$\underline{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}$$

$$\underline{y} = \underbrace{\begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{pmatrix}}_{N \times 1}, \quad A = \underbrace{\begin{bmatrix} f_1(x^{(1)}) & \cdots & f_p(x^{(1)}) \\ \vdots & & \vdots \\ f_1(x^{(N)}) & \cdots & f_p(x^{(N)}) \end{bmatrix}}_{N \times p}$$

Again standard least squares: $\underline{\theta} = A \backslash \underline{y}$

- Solving underdetermined / wide rectangular system of linear equations :

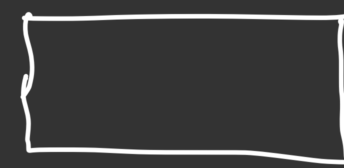
$$\underbrace{A}_{m \times n} \, \underline{x} = \underline{b} \,, \qquad m < n$$

A: $m \times n$ known

$\underline{x}$: $n \times 1$ unknown

$\underline{b}$: $m \times 1$ known

where matrix A is "wide" (more columns than rows)



$\iff$ # of unknowns $>$ # of equations

Assume: $\operatorname{rank}(A) = m$

$\iff$ A has full row rank

$\iff$ All rows of matrix A are linearly independent

In this case, the system $A\underline{x} = \underline{b}$ has non-unique / multiple solution vectors $\underline{x}$.

All solutions are of the form:

$$\{\underline{x} \in \mathbb{R}^n \mid A\underline{x} = \underline{b}\} = \{\underline{x}_{\text{particular}} + \underline{z} \mid \underline{z} \in \operatorname{nullspace}(A)\}$$

Recall: $\operatorname{nullspace}(A) := \{\underline{z} \in \mathbb{R}^n \mid A\underline{z} = \underline{0}\}$.