Linear Least Squares

CS/SE 4X03

Ned Nedialkov

McMaster University

November 2, 2021

Outline

Formulation

In linear least squares, we have $n+1$ basis functions and $m+1$ data points (x_k, y_k), $k = 1, \ldots, m$, where $m > n$

$$v(x) = \sum_{j=0}^{n} c_j \phi_j(x), \quad v(x_k) \approx y_k, \qquad k = 0, \ldots, m$$

Find the c_j such that the sum

$$\sum_{k=0}^{m} (v(x_k) - y_k)^2 = \sum_{k=0}^{m} \left(\sum_{j=0}^{n} c_j \phi_j(x_k) - y_k \right)^2$$

is minimized

Least squares vs. interpolation

In interpolation, given $(n+1)$ data points (x_k, y_k), we find a function $v(x)$ such that

$$v(x) = \sum_{j=0}^{n} c_j \phi_j(x), \quad v(x_k) = y_k, \qquad k = 0, \ldots, n$$

In real-life applications, the data points may not be accurate, e.g. may come from measurements

May not make sense to interpolate inaccurate data

With least squares, may want to pick up a trend in the data, e.g. average temperature over last 10 years, is it warming or cooling down?

Linear fit

Suppose we search for a linear fit: $y = ax + b$, i.e. find a and b

Error or residual

$$r_k = ax_k + b - y_k$$

Find a and b such that

$$\phi(a,b) = \sum_{k=0}^{m} r_k^2 = \sum_{k=0}^{m} (ax_k + b - y_k)^2$$

is minimized

Necessary conditions for minimum:

$$\frac{\partial \phi}{\partial a} = 0, \quad \frac{\partial \phi}{\partial b} = 0$$

$$0 = \frac{\partial \phi}{\partial a} = 2\sum_{k=0}^{m}(ax_k + b - y_k)x_k$$

$$0 = a\sum_{k=0}^{m} x_k^2 + b\sum_{k=0}^{m} x_k - \sum_{k=0}^{m} y_k x_k$$

from which

$$\left(\sum_{k=0}^{m} x_k^2\right) a + \left(\sum_{k=0}^{m} x_k\right) b = \sum_{k=0}^{m} x_k y_k \tag{1}$$

$$0 = \frac{\partial \phi}{\partial b} = 2\sum_{k=0}^{m}(ax_k + b - y_k)$$

$$0 = a\sum_{k=0}^{m} x_k + b\sum_{k=0}^{m} 1 - \sum_{k=0}^{m} y_k$$

from which

$$\left(\sum_{k=0}^{m} x_k\right) a + (m+1)b = \sum_{k=0}^{m} y_k \qquad (2)$$

From (1) and (2) we have the linear system

$$\begin{bmatrix} \sum_{k=0}^{m} x_k^2 & \sum_{k=0}^{m} x_k \\ \sum_{k=0}^{m} x_k & m+1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{m} x_k y_k \\ \sum_{k=0}^{m} y_k \end{bmatrix}$$

Denote

$$p = \sum_{k=0}^{m} x_k, \qquad q = \sum_{k=0}^{m} y_k$$

$$r = \sum_{k=0}^{m} x_k y_k, \qquad s = \sum_{k=0}^{m} x_k^2$$

Then the system is

$$\begin{bmatrix} s & p \\ p & m+1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} r \\ q \end{bmatrix}$$

Solve for a and b

This system can be also obtained as follows.

Write $ax_k + b = y_k$, $k = 1, \ldots, m$ as

$$Az = \begin{bmatrix} x_0 & 1 \\ x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix} = f$$

Multiply both sided by A^T, $A^T A z = A^T f$

$$A^T A = \begin{bmatrix} x_0 & x_1 & \cdots & x_m \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_0 & 1 \\ x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^m x_k^2 & \sum_{k=0}^m x_k \\ \sum_{k=0}^m x_k & m+1 \end{bmatrix}$$

$$= \begin{bmatrix} s & p \\ p & m+1 \end{bmatrix}$$

$$A^T f = \begin{bmatrix} x_0 & x_1 & \cdots & x_m \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^m x_k y_k \\ \sum_{k=0}^m y_k \end{bmatrix}$$

$$= \begin{bmatrix} r \\ q \end{bmatrix}$$

$Az = f$ is overdetermined, more equations than unknowns

In MATLAB, find z by `A\f`

Example

- Assume a program runs in αn^β, where α and β are real constants we don't know
- How to determine them?
- Run the program with sizes $n_1, n_2, \ldots, n_m$ and measure the corresponding CPU times $t_1, t_2, \ldots, t_m$, $m > 2$
- Write $\alpha n_i^\beta = t_i$, $i = 1, \ldots, m$
- Then

$$\ln \alpha + \beta \ln n_i = \ln t_i, \quad i = 1, \ldots, m$$

- Let $x = \ln \alpha$
- Then

$$1 \cdot x + \ln n_i \cdot \beta = \ln t_i, \quad i = 1, \ldots, m$$

Write

$$\begin{aligned}
1 \cdot x + \ln n_1 \cdot \beta &= \ln t_1 \\
1 \cdot x + \ln n_2 \cdot \beta &= \ln t_2 \\
&\vdots \\
1 \cdot x + \ln n_m \cdot \beta &= \ln t_m
\end{aligned}$$

Then

$$Ay = \begin{bmatrix} 1 & \ln n_1 \\ 1 & \ln n_2 \\ \vdots & \vdots \\ 1 & \ln n_m \end{bmatrix} \begin{bmatrix} x \\ \beta \end{bmatrix} = \begin{bmatrix} \ln t_1 \\ \ln t_2 \\ \vdots \\ \ln t_m \end{bmatrix} = b$$

Solve in Matlab as `y = A\b`; $\alpha =$ `exp(y(1))` $\beta =$ `y(2)`

Solving overdetermined systems

- $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$
  $m > n$
- $Ax = b$ is an overdetermined system: more equations than variables
- Find x that minimizes $\|b - Ax\|_2$
- $r = b - Ax$
- $\|r\|_2^2 = \sum_{i=1}^m r_i^2 = \sum_{i=1}^m \left(b_i - \sum_{j=1}^n a_{ij}x_j\right)^2$
- Let
$$\phi(x) = \frac{1}{2}\|r\|_2^2 = \frac{1}{2}\sum_{i=1}^m \left(b_i - \sum_{j=1}^n a_{ij}x_j\right)^2$$

- We want to find the minimum of $\phi(x)$
- Necessary conditions are

$$\frac{\partial \phi}{\partial x_k} = 0, \quad \text{for } k = 1, \dots, n$$

$$\begin{aligned}
0 = \frac{\partial \phi}{\partial x_k} &= \frac{\partial}{\partial x_k} \left(\frac{1}{2} \sum_{i=1}^{m} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right)^2 \right) \\
&= \frac{1}{2} \sum_{i=1}^{m} \frac{\partial}{\partial x_k} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right)^2 \\
&= \sum_{i=1}^{m} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right) (-a_{ik})
\end{aligned}$$

$$
\begin{aligned}
0 &= \sum_{i=1}^{m} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right) (-a_{ik}) \\
&= -\sum_{i=1}^{m} a_{ik} b_i + \sum_{i=1}^{m} a_{ik} \sum_{j=1}^{n} a_{ij} x_j
\end{aligned}
$$

We have

$$
\sum_{i=1}^{m} a_{ik} \sum_{j=1}^{n} a_{ij} x_j = \sum_{i=1}^{m} a_{ik} b_i, \quad k = 1, \ldots, n
$$

This is the same as $A^T A x = A^T b$, as explained below

A is $m \times n$. A^T is $n \times m$.

Let $y = Ax$. $y_i = \sum_{j=1}^{n} a_{ij} x_j$, $i = 1, \ldots, m$.

The kth component of $A^T A x = A^T y$ is

$$(A^T A x)_k = (A^T y)_k = \sum_{i=1}^{m} (A^T)_{ki} y_i = \sum_{i=1}^{m} a_{ik} y_i = \sum_{i=1}^{m} a_{ik} \sum_{j=1}^{n} a_{ij} x_j$$

The kth component of $A^T b$ is

$$(A^T b)_k = \sum_{i=1}^{m} (A^T)_{ki} b_i = \sum_{i=1}^{m} a_{ik} b_i$$

$(A^T A x)_k = (A^T b)_k$, $k = 1, \ldots, n$ is

$$\sum_{i=1}^{m} a_{ik} \sum_{j=1}^{n} a_{ij} x_j = \sum_{i=1}^{m} a_{ik} b_i$$

Normal equations

- $A^T A x = A^T b$ are called *normal equations*
- If A has a full-column rank (all columns are linearly independent),
$$\min_x \|b - Ax\|_2$$
has a unique solution which is the solution to $(A^T A)x = A^T b$:
$$x = (A^T A)^{-1} A^T b = A^\dagger b$$
- $A^\dagger = (A^T A)^{-1} A^T$ is the *pseudo inverse* of A