

Linear Least Squares

CS/SE 4X03

Ned Nedialkov

McMaster University

October 28, 2021

Outline

Example: linear least squares fit

Solving overdetermined systems

Normal equations

Example: linear least squares fit

- 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, \dots, n_m and measure the corresponding CPU times t_1, t_2, \dots, t_m , $m > 2$
- Write $\alpha n_i^\beta = t_i$, $i = 1, \dots, m$

- Then

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

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

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

- Write

$$1 \cdot x + \ln n_1 \cdot \beta = \ln t_1$$

$$1 \cdot x + \ln n_2 \cdot \beta = \ln t_2$$

$$\vdots$$

$$1 \cdot x + \ln n_m \cdot \beta = \ln t_m$$

- 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 $y = A \backslash b$

- $y(1) = \ln \alpha, \quad \alpha = \exp(y(1))$
- $\beta = y(2)$
- Find these constants when solving linear systems using Matlab's \

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

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

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

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

Solve the system

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

Normal equations

- The above system is the same as

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

- These 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