Least Squares: Part 1

Lecture 3-1 - CMSE 382

Prof. Elizabeth Munch

Michigan State University
::
Dept of Computational Mathematics, Science & Engineering

Weds, Jan 28, 2026

Topics covered

- Ordinary Least squares (OLS)
- Linear fitting
- Polynomial fitting

Section 1

Least squares

Overdetermined systems

Recall Matrix Rank

The rank of matrix $A \in \mathbb{R}^{m \times n}$ is a non-negative integer $\text{rank}(A) \leq \max(m, n)$ given by any of the following:

- the maximum number of its linearly independent rows or columns.
- the number of non-zero rows in its row-echelon form
- the number of non-zero singular values in its SVD

A linear system of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is of full rank and $\mathbf{b} \in \mathbb{R}^{m}$ is

- **overdetermined** if $m > n$.
  - Almost always inconsistent (has no solution)
- **determined** if $m = n$, and the equations are independent
  - Has a unique solution
- **underdetermined** if $m < n$
  - Either inconsistent (no solution), or infinitely many solutions

Goal: Obtain an approximate solution for overdetermined systems

Ordinary Least Squares Approximation

Definition

The **residual sum of squares error** (or total sum of square errors) is

$$\mathrm{RSS} = \|A\mathbf{x}_{\mathrm{LS}} - \mathbf{b}\|^2$$

The minimizer of this equation is the **least squares estimate**

$$\mathbf{x}_{\mathrm{LS}} = (\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{A}^\top \mathbf{b}$$

- RSS measures how well $\mathbf{x}_{\mathrm{LS}}$ approximates the data $\mathbf{b}$
- The Least squares estimate minimizes RSS
$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$$
- You did this in CMSE381, but that textbook used the notation β_i for the coefficients. In this notation, the entries of $\mathbf{x}$ are the coefficients of the linear function we're trying to learn.

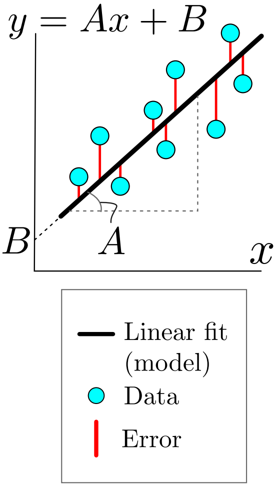
Linear Data Fitting

- Given a set of data points $(\mathbf{s}_i, t_i)$, $i = 1, 2, \ldots, m$, where $\mathbf{s}_i \in \mathbb{R}^n$ and $t_i \in \mathbb{R}$
- Assume that the target t_i can be approximated by the linear transformation of the data $\mathbf{s}_i$

$$t_i \approx \mathbf{s}_i^\top \mathbf{x}$$

- Then the least squares method is applied to find the parameter $\mathbf{x}$ that solves the following optimization problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \sum_{i=1}^{m} (\mathbf{s}_i^\top \mathbf{x} - t_i)^2 = \min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{S}\mathbf{x} - \mathbf{t}\|^2$$



Polynomial fitting

If we know that the points are approximately related to a polynomial of degree at most d, i.e., there exists $a_0, \ldots, a_d$ such that

$$\sum_{j=0}^{d} a_j u_i^j \approx y_i, \qquad i = 1, \ldots, m$$

we can use OLS $\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{Sx} - \mathbf{t}\|^2$ with

$$\begin{bmatrix} 1 & u_1 & u_1^2 & \ldots & u_1^d \\ 1 & u_2 & u_2^2 & \ldots & u_2^d \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & u_m & u_m^2 & \ldots & u_m^d \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$S = \begin{bmatrix} 1 & u_1 & u_1^2 & \ldots & u_1^d \\ 1 & u_2 & u_2^2 & \ldots & u_2^d \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & u_m & u_m^2 & \ldots & u_m^d \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{bmatrix} \qquad \mathbf{t} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix}$$

Groups - Round 2

Group 1
Abigail, Shanze, Jack, Quang Minh,

Group 2
Igor, Atticus, K M Tausif, Long,

Group 3
Yousif, Zheng, Jake, Purvi,

Group 4
Maye, Alice, Arjun, Kyle,

Group 5
Monirul Amin, Jay, Brandon, Luis,

Group 6
Scott, Ha, Lora, Tianjian,

Group 7
Braedon, Sai, Joseph, Noah,

Group 8
Michal, Aidan, Jonid, Dev,

Group 9
Vinod, Saitej, Anthony, Breena,

Group 10
Karen, Dori, Lowell, Aaron,

Group 11
Jamie, Sanskaar, Dominic, Lauryn,

Group 12
Andrew, Arya, Daniel, Morgan,

Next time

- Check course webpage for videos and reading for next class
- Office hours posted on the course webpage
- Homework posted, due Friday Jan 30 at 11:59pm
- Quiz 2 on Weds 2/4