Understanding Linear Least Squares

Eldad Haber

Data Science 100

Motivation

- Why do we need least squares?
- Examples:
  - Predicting trends from data (e.g., temperature over time)
  - Fitting a straight line to noisy measurements
  - Solving systems of equations that have no exact solution

Problem Setup

- Given m equations with n unknowns, written as a matrix equation:

$$Ax = b$$

- When $m > n$, the system is overdetermined (more equations than unknowns)
- Usually, no exact solution exists, so we seek the "best" approximate solution

What Does "Best" Mean?

- We define "best" as minimizing the error $\|Ax - b\|$
- The goal: find x that minimizes the sum of squared differences:
$$\min_x \|Ax - b\|^2$$
- This ensures the closest possible fit in a least-squares sense

Deriving the Solution

- The error is minimized when the residual $r = b - Ax$ is orthogonal to the column space of A
- Mathematically, this leads to the normal equations:

$$A^T Ax = A^T b$$

- If A has full column rank, we solve for x:

$$x = (A^T A)^{-1} A^T b$$

More math

- Given an overdetermined system $Ax = b$, we seek the best approximation by minimizing:

$$\min_x \|Ax - b\|^2$$

- Leads to the **normal equations**:

$$A^T Ax = A^T b$$

Expanding the Squared Norm

- The least squares error is given by:

$$\|Ax - b\|^2 = (Ax - b)^T(Ax - b)$$

- Expanding:

$$\underbrace{x^T A^T Ax}_{\text{Quadratic}} - \underbrace{2b^T Ax}_{\text{linear}} + \underbrace{b^T b}_{\text{constant}}$$

Taking the Gradient

- Differentiate with respect to x:

$$2A^T Ax - 2A^T b = 0$$

- Solve for x:

$$A^T Ax = A^T b$$

Geometric Interpretation

- The residual $r = b - Ax$ is **orthogonal** to the column space of A
- This ensures the best projection of b onto the space spanned by the columns of A

Solution for x

- If $A^T A$ is invertible:

$$x = (A^T A)^{-1} A^T b$$

- This gives the best least-squares solution.

Summary

- Minimizing $\|Ax - b\|^2$ leads to the normal equations.
- The normal equations ensure that the error is minimized in a least-squares sense.
- Solution: $x = (A^T A)^{-1} A^T b$ when $A^T A$ is invertible.

Questions?

Open floor for questions and discussion!

Example Problem

- Suppose we want to fit a line $y = mx + c$ through three points: $(1, 2), (2, 3), (3, 5)$
- Set up the system:

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$$

- Solve using normal equations

Geometric Interpretation

- The column space of A is a subspace of $\mathbb{R}^m$
- The best solution projects b onto this subspace
- The residual r is perpendicular to the column space

Applications

- Data fitting (regression in statistics)
- Engineering (signal processing, control systems)
- Machine learning (least squares classifiers)
- Scientific computing (approximating solutions to equations)

Summary

- Least squares is a method to approximate solutions to overdetermined systems
- The normal equations provide a way to compute the best fit
- Applications span many fields, from data science to physics

Questions & Discussion

- Open floor for student questions and interactive discussion
- Possible exercises:
  - Solve a small least-squares problem by hand
  - Interpret the geometric meaning of a given least-squares solution