Understanding and Applying the Least Squares Method: Estimation and Error Minimization

Sai Bharadwaj Sirigadi

Introduction

The least-squares method is a powerful mathematical technique used for estimating unknown parameters in various fields, ranging from statistics and engineering to data analysis and optimization. It provides a systematic way to find the best-fitting solution that minimizes the overall difference between observed data and a mathematical model's predictions. By leveraging the principle of minimizing the sum of squared residuals, the least-squares method helps us make optimal approximations and informed decisions in situations where direct solutions might be unavailable or inaccurate.
Least Square Method plays a prominent role in..

- Optimal Estimation.
- Handling Noisy Data.
- Error Minimization.

Sources of Errors

1 Limitations of Sensors

Sensors don't measure the exact Physical value. These are termed as Measurement errors.

- **Deterministic Error Sources:** These are due to Measurement bias & Misalignment, which can be corrected by calibration.
- **Non-Deterministic Error Sources:** These are termed as noise as they are random by nature, which needs to model stochastically.

2 Limitations in our Model

We don't understand all the forces acting as the physical phenomenon is so complex.

3 Limitations due to simplification in Implementation

More Approximations while Implementation to simplify the formulation of a problem which lead to errors.

Error Minimization using Least Squares Method

Consider $(e = y - x)$ is the equation of error.
To fit the model for more errors/more measurements, matrix formulation always a feasible way for ease of computation. Formulation of errors in matrix form is as follows:

$$
\begin{aligned}
e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} &= y - Hx \\
&= \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x
\end{aligned}
$$

H is a Jacobian with dimension $M \times N$
Where,

- M - Number of Measurements.
- N - Number of Parameters need to be estimated.

Minimizing the Squared Error Criterion

$$\begin{aligned}\hat{x} &= argmin_x(e_1^2 + e_2^2 + e_3^2 + e_4^2)\\ &= e^T e = (y - Hx)^T (y - Hx)\\ &\ldots\ldots\ldots\\ \hat{x} &= (H^T H)^{-1} H^T y\end{aligned}$$

Solution exists if $(H^T H)$ is not singular.
Example: Consider Temperatures of a Person is recorded, where each measurement is incorporated with noise and we wanted optimal temperature out of measurements:

Observation	**Temperature (C)**
1	36
2	35.5
3	36.8
4	36.5

Measurement Model

$$\begin{aligned}y_1 &= x + v_1\\ y_2 &= x + v_2\\ y_3 &= x + v_3\\ y_4 &= x + v_4\end{aligned}$$

where x is measurement & v is noise.
Squared Error

$$\begin{aligned}e_1^2 &= (y_1 - x)^2\\ e_2^2 &= (y_2 - x)^2\\ e_3^2 &= (y_3 - x)^2\\ e_4^2 &= (y_4 - x)^2\end{aligned}$$

$\hat{x} = argmin_x(e_1^2 + e_2^2 + e_3^2 + e_4^2)$
Thus,

$$\begin{aligned}\hat{x} &= (H^T H)^{-1} H^T y\\ &= (\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix})^{-1}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}\begin{bmatrix}36\\35.5\\36.8\\36.5\end{bmatrix}\\ &= \frac{1}{4}(36 + 35.5 + 36.8 + 36.5)\\ &= 36.2\end{aligned}$$

What if there are two or more different thermometers, and one is more accurate than the other.
To solve this, Weighted Least Square is used.
Linear Model

$$y = Hx + v$$

Noise v is IID-Independent and Identically Distributed.

$$E[v_i^2] = \sigma^2$$

$$R = E[vv^T] = \begin{bmatrix}\sigma^2 & \ldots & 0\\ \vdots & \ddots & \vdots\\ 0 & \ldots & \sigma^2\end{bmatrix}$$

Linear Model

$$y = Hx + v$$

Noise v is IID-Independent but different variance.

$$E[v_i^2] = \sigma_i^2 \qquad i = 1, 2, ..., m$$

$$R = E[vv^T] = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_m^2 \end{bmatrix}$$

Weighted Least Square Criterion

$$\hat{x} = e^T R^{-1} e$$

$$\dots\dots\dots$$

$$\hat{x} = (H^T R^{-1} H)^{-1} H^T R^{-1} y$$

For the same problem temperature of Person would be:

$$\hat{x} = (H^T R^{-1} H)^{-1} H^T R^{-1} y$$

$$= (\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0.04 & 0 \\ 0 & 0 & 0 & 0.04 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix})^{-1} \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 070.25 & 0 & 0 & \\ 0 & 0 & 0.04 & 0 \\ 0 & 0 & 0 & 0.04 \end{bmatrix}^{-1} \begin{bmatrix} 36 \\ 35.5 \\ 36.8 \\ 36.5 \end{bmatrix} = 36.5$$

References

[1] Skill-Lync, *Localization, Mapping & SLAM.*

[2] Gilbert Strang. *Introduction to Linear Algebra* (2016).