THE LEAST SQUARES METHOD

Robert Joan Arinyo

Grup d’Informàtica a l’Enginyeria
Escola Tècnica Superior d’Enginyeria Industrial
Universitat Politècnica de Catalunya

Màster en Computació
Overview

1. Introduction
2. Fitting problem formulation
3. The general linear problem
4. Intersecting n lines in 2D
5. Intersecting n planes in 3D
6. Fitting a plane to n given points in 3D
The method of least squares is a standard approach to the approximate solution of overdetermined systems, i.e. sets of equations in which there are more equations than unknowns.

Least squares means that the overall solution minimizes the sum of the squares of the errors made in solving every single equation.

The most important application is in data fitting.

The best fit in the least-squares sense minimizes the sum of squared residuals, a residual being the difference between an observed value and the fitted value provided by a model.
Depending on whether or not the residuals are linear in all unknowns, Least squares problems fall into two categories:

1. linear least squares
2. nonlinear least squares
Linear least squares occurs in statistical regression analysis. It has a closed-form solution.

The approach is called **linear** least squares since the solution depends **linearly** on the data.
The nonlinear least squares problem has no closed solution and is usually solved by iterative refinement.

At each iteration the system is approximated by a linear one, thus the core calculation is similar in both cases.
Given $N$ points located at positions $\mathbf{x}_i$ in $\mathbb{R}^d$ with $i \in [1..N]$. We wish to obtain a globally defined function $f(\mathbf{x})$ that approximates the given scalar values $f_i$ at points $\mathbf{x}_i$ in such a way that minimizes the error functional

$$J_{LS} = \sum_i \| f(\mathbf{x}_i) - f_i \|^2$$
Illustrative example:

- Experimental data. Points shown in red in the picture

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

- It is desired to find a line $y = ax + b$ that fits "best" these four points. In other words, we would like to find the numbers $a$ and $b$ that approximately solve the overdetermined linear system

$$
\begin{align*}
  a + b &= 6 \\
  2a + b &= 5 \\
  3a + b &= 7 \\
  4a + b &= 10
\end{align*}
$$
The least squares approach minimizes the sum of squares of errors or residual values, that is, to find the minimum of the function

\[ R(a, b) = (6-(a+b))^2 + (5-(2a+b))^2 + (7-(3a+b))^2 + (10-(4a+b))^2 \]

The minimum is determined by calculating the partial derivatives of \( R(a, b) \) with respect to \( a \) and \( b \) and setting them to zero. This results in a system of two equations in two unknowns, called the normal equations.

When solved, we have \( a = 1.4 \) and \( b = 3.5 \). Therefore, the line \( y = 1.4x + 3.5 \) is the best least squares fit.
Solution to the first degree problem

The common computational procedure to find a first-degree polynomial function approximation over $n$ data points is as follows.

- The slope is given by
  \[ a = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \]

- The Y-intercept is given by
  \[ b = \frac{\sum y \sum x^2 - \sum x \sum xy}{n \sum x^2 - (\sum x)^2} \]
Consider an overdetermined system of $m$ linear equations each with $n$ unknowns such that $m > n$,

$$\sum_{j=1}^{n} a_j x_{ij} = b_j, \quad (i = 1, 2, \ldots, m)$$

Written in matrix form

$$XA = Y$$
General Problem (2/10)

Where

\[
\begin{align*}
X &= \begin{pmatrix}
  x_{11} & x_{12} & \cdots & x_{1n} \\
  x_{21} & x_{22} & \cdots & x_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{m1} & x_{m2} & \cdots & x_{mn}
\end{pmatrix} \\
A &= \begin{pmatrix}
a_1 \\
a_2 \\
\vdots \\
a_n
\end{pmatrix} \\
Y &= \begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{pmatrix}
\end{align*}
\]
Such a system usually has no solution, and the goal is then to find the coefficients $A$ which fit the equations "best", in the sense of minimizing the residuals:

$$\| Y - XA \|^2$$
This minimization problem has a unique solution, provided that the $n$ columns of the matrix $X$ are linearly independent.

The solution is given by solving the normal equations:

$$(X^T X)A = X^T Y$$
General Problem (5/10)

- Solving the normal equations

\[(X^\top X)A = X^\top Y\]

entails inverting \((X^\top X)\).

- However, for large values of \(m\), matrix \((X^\top X)\) is ill conditioned and thus the computation is numerically unstable.
• It is preferable to apply orthogonal decomposition methods.
• The residuals are \( r = Y - XA \)
• Apply QR decomposition to get \( X = QR \)
• \( Q \) is an \( m \times m \) orthogonal matrix and \( R \) is an \( m \times n \) which is partitioned into an \( n \times n \) upper triangular matrix block, say \( R_n \), and a \( (m - n) \times n \) zero block, say \( O \).

\[
\begin{pmatrix}
R_n \\
O
\end{pmatrix}
\]
General Problem (7/10)
Therefore, residuals $r = Y - QRA$ can be written as

$$Q^T r = Q^T Y - (Q^T Q) RA$$

$$= \begin{pmatrix} (Q^T Y)_n - R_n A \\ (Q^T Y)_{m-n} \end{pmatrix}$$

$$= \begin{pmatrix} u \\ v \end{pmatrix}$$

$v$ doesn't depend on $A$. Then the minimum residual value is attained when the upper block, $u$, is zero.
• Therefore the parameters are found by solving

\[ R_n A = \left( Q^T Y \right)_n \]

• These equations are easily solved as \( R_n \) is upper triangular.
General Problem (10/10)

HOMEWORK

Search for numerical libraries to perform the computations so far discussed
Intersecting n lines in 2D

- Set of given line equations

\[
\begin{align*}
    a_1 x + b_1 y + c_1 &= 0 \\
    a_2 x + b_2 y + c_2 &= 0 \\
    \vdots \\
    a_n x + b_n y + c_n &= 0
\end{align*}
\]

- Residuals are \( r_i = a_i x + b_i y - c_i \), \( 1 \leq i \leq n \)

- Apply what has been said in the solution to the first degree problem considering as unknowns \( x \) and \( y \).
Intersecting n planes in 3D

• Set of given plane equations

\[
\begin{align*}
  a_1 x + b_1 y + c_1 z + d_1 &= 0 \\
  a_2 x + b_2 y + c_2 z + d_2 &= 0 \\
  \vdots \\
  a_n x + b_n y + c_n z + d_n &= 0
\end{align*}
\]

• Residuals are \( r_i = a_i x + b_i y + c_i z - d_i \), \( 1 \leq i \leq n \)

• Apply what has been said in the solution to the general problem considering as unknowns \( x, y \) and \( z \).
Fitting a plane to n points in 3D

• Set of given 3D points

\[
\begin{array}{ccc}
  x_1 & y_1 & z_1 \\
  x_2 & y_2 & z_2 \\
  \ldots \\
  x_n & y_n & z_n
\end{array}
\]

• Plane equation to be fit \( ax + by + cz + d = 0 \)

• We want to find values for \( a, b, c \) and \( d \) that approximately solve the system of equations

\[
\begin{aligned}
  x_1a + y_1b + z_1c + d &= 0 \\
  x_2a + y_2b + z_2c + d &= 0 \\
  \ldots \\
  x_na + y_nb + z_nc + d &= 0
\end{aligned}
\]

• Apply what has been said in the solution to the general problem considering as unknowns \( a, b, c \) and \( d \).
This is it concerning Least Squares fitting