最小二乗法

2022 年 4 月 29 日

目次

1 概要

ある観測データの組 (x, y) は、モデル関数 $f(x)$ と誤差 ε の和で

$$y = f(x) + \varepsilon \tag{1}$$

と表せるとする。

N 組の観測データ (x_i, y_i) が与えられたとき、各 x_i についてモデルから得られる予測値 $f(x_i)$ と、実際の測定値 y_i の誤差 ε_i つまり

$$\varepsilon_i := |f(x_i) - y_i| \tag{2}$$

の和を最小にすることを考えたい。しかし、絶対値があると後々の計算が大変なので、実際には、誤差の二乗和

$$\Delta := \frac{1}{2}\sum_{i=1}^{N}\varepsilon_i^2 = \frac{1}{2}\sum_{i=1}^{N}\left(f\left(x_i\right) - y_i\right)^2 \tag{3}$$

を最小にするようなモデル $f\left(x\right)$ を求める。なお、係数 $\frac{1}{2}$ は、後の計算の時に、多少計算が楽になるのを見越して付けているだけで、本質ではない。

以下のセクションでは、簡単なものから複雑なものまで、具体的に求めてみる。

2　線形回帰 (直線) の場合

モデル関数は

$$f\left(x\right) = ax + b \tag{4}$$

であるから、式 (3) は

$$\Delta = \frac{1}{2}\sum_{i=1}^{N}\varepsilon_i = \frac{1}{2}\sum_{i=1}^{N}\left(ax_i + b - y_i\right)^2 \tag{5}$$

となる。これを最小化する a, b を求めればよい。これを a, b に関する二次関数と見れば、この関数が最小値を取る点では a および b についての偏微分がゼロであるから、

$$\frac{\partial}{\partial a}\Delta = 0,$$
$$\frac{\partial}{\partial b}\Delta = 0$$

を連立して解けばよい。実際に計算してみると

$$\frac{\partial}{\partial a}\Delta = \sum_{i=1}^{N}\left(ax_i + b - y_i\right)x_i = a\sum_{i=1}^{N}x_i^2 + b\sum_{i=1}^{N}x_i - \sum_{i=1}^{N}x_i y_i = 0,$$
$$\frac{\partial}{\partial b}\Delta = \sum_{i=1}^{N}\left(ax_i + b - y_i\right)1 = a\sum_{i=1}^{N}x_i + b\sum_{i=1}^{N}1 - \sum_{i=1}^{N}y_i, = 0,$$

すなわち

$$
\begin{cases}
\left(\sum_{i=1}^{N} x_i^2\right) a + \left(\sum_{i=1}^{N} x_i\right) b & = \sum_{i=1}^{N} x_i y_i, \\
\left(\sum_{i=1}^{N} x_i\right) a + \left(\sum_{i=1}^{N} 1\right) b & = \sum_{i=1}^{N} y_i,
\end{cases}
\tag{6}
$$

を解けばよい。行列を使って書けば

$$
\begin{bmatrix} \sum_{i=1}^{N} x_i^2, & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i, & \sum_{i=1}^{N} 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} y_i \end{bmatrix}
\tag{7}
$$

よって、

$$
Det := \left(\sum_{i=1}^{N} x_i^2\right)\left(\sum_{i=1}^{N} 1\right) - \left(\sum_{i=1}^{N} x_i\right)^2
\tag{8}
$$

とおけば、

$$
\begin{aligned}
\begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} \sum_{i=1}^{N} x_i^2, & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i, & \sum_{i=1}^{N} 1 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} y_i \end{bmatrix} \\
&= \frac{1}{Det} \begin{bmatrix} \sum_{i=1}^{N} 1, & -\sum_{i=1}^{N} x_i, \\ -\sum_{i=1}^{N} x_i, & \sum_{i=1}^{N} x_i^2 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} y_i \end{bmatrix}
\end{aligned}
\tag{9}
$$

さらに $\sum_{i=1}^{N} 1 = N$ であるから、最終的に

$$
\begin{aligned}
a &= \frac{N \sum_{i=1}^{N} x_i y_i - \left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}, \\
b &= \frac{\left(\sum_{i=1}^{N} x_i^2\right)\left(\sum_{i=1}^{N} y_i\right) - \left(\sum_{i=1}^{N} x_i y_i\right)\left(\sum_{i=1}^{N} x_i\right)}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2},
\end{aligned}
\tag{10}
$$

3　一般的な多項式の場合

モデル関数は

$$
f(x) = a_0 + a_1 x^1 + a_2 x_i^2 \cdots + a_{n-1} x^{n-1} + a_n x^n
\tag{11}
$$

であるから、式 (3) は

$$\Delta = \frac{1}{2}\sum_{i=1}^{N}\left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_{n-1}x_i^{n-1} + a_n x_i^n - y_i\right)^2 \tag{12}$$

となる。先ほどと同様に $a_0, a_a, \ldots, a_n$ で偏微分すると

$$\begin{aligned}
\frac{\partial}{\partial a_0}\Delta &= \sum_{i=1}^{N}\left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots a_n x_i^n - y_i\right)1,\\
\frac{\partial}{\partial a_1}\Delta &= \sum_{i=1}^{N}\left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots a_n x_i^n - y_i\right)x_i,\\
\frac{\partial}{\partial a_2}\Delta &= \sum_{i=1}^{N}\left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots a_n x_i^n - y_i\right)x_i^2,\\
&\vdots\\
\frac{\partial}{\partial a_n}\Delta &= \sum_{i=1}^{N}\left(a_0 + a_1 x_i + a_2 x_i^2 + \cdots a_n x_i^n - y_i\right)x_n^2
\end{aligned} \tag{13}$$

この各式がゼロとおいて、移項して行列の形に纏めると

$$\begin{bmatrix}
\sum_{i=1}^{N} x_i^0, & \sum_{i=1}^{N} x_i^1, & \sum_{i=1}^{N} x_i^2, & \cdots, & \sum_{i=1}^{N} x_i^n\\
\sum_{i=1}^{N} x_i^1, & \sum_{i=1}^{N} x_i^2, & \sum_{i=1}^{N} x_i^3, & \cdots, & \sum_{i=1}^{N} x_i^{n+1}\\
\sum_{i=1}^{N} x_i^2, & \sum_{i=1}^{N} x_i^3, & \sum_{i=1}^{N} x_i^4, & \cdots, & \sum_{i=1}^{N} x_i^{n+2}\\
\vdots, & \vdots, & \vdots, & \ddots, & \vdots,\\
\sum_{i=1}^{N} x_i^n, & \sum_{i=1}^{N} x_i^{n+1} & \sum_{i=1}^{N} x_i^{n+2}, & \cdots, & \sum_{i=1}^{N} x_i^{2n}
\end{bmatrix}
\begin{bmatrix} a_0\\ a_1\\ a_2\\ \vdots\\ a_n \end{bmatrix}
=
\begin{bmatrix}
\sum_{i=1}^{N} x_i^0 y_i\\
\sum_{i=1}^{N} x_i^1 y_i\\
\sum_{i=1}^{N} x_i^2 y_i\\
\vdots\\
\sum_{i=1}^{N} x_i^n y_i
\end{bmatrix} \tag{14}$$

後はこれに掃き出し法や LU 分解など、連立一次方程式の解法アルゴリズムを適用すれば良い。

4　発展編

4.1　円の場合

円の場合は $y = f(x)$ の形で書けない (書けなくはないが面倒) ので工夫が必要である。中心 (a, b) 半径 r の円の場合 (x, y) は

$$(x-a)^2 + (y-b)^2 = r^2 \tag{15}$$

を満たす。

今までの方法では $y = f(x)$ という形式だったが、これを変形すると 点 (x_i, y_i) がこのモデル関数のグラフ上に乗っているならば、

$$f(x_i) - y_i = 0 \tag{16}$$

を満たすべきであって、ε_i は、この左辺の値 (と 0 との誤差) である、とも言える。これと同様に考えると、

$$\varepsilon_i = (x_i - a)^2 + (y - b)^2 - r^2 \tag{17}$$

で同様の手続きをすればよいが、そのままでは上の式を二乗したときに a, b, r に関する四次式になってしまうので、最小値を求めることが難しい (偏微分がゼロの点が最小値とは限らない) 。そこで、以下のように式変形を行う。

$$\begin{aligned}
\varepsilon_i &= (x_i - a)^2 + (y_i - b)^2 - r^2 \\
&= x_i^2 - 2ax_i + a^2 + y_i^2 - 2by_i + b^2 - r^2 \\
&= -2ax_i - 2by_i + (a^2 + b^2 - r^2) + x_i^2 + y_i^2 \\
&= Ax_i + By_i + C + x_i^2 + y_i^2
\end{aligned} \tag{18}$$

ただし

$$\begin{aligned}
A &= -2a, \\
B &= -2b, \\
C &= a^2 + b^2 - r^2
\end{aligned} \tag{19}$$

である。すると式 (3) は

$$\Delta := \frac{1}{2}\sum_{i=1}^{N} \varepsilon_i^2 = \frac{1}{2}\sum_{i=1}^{N} \left(Ax_i + By_i + C + x_i^2 + y_i^2\right)^2 \tag{20}$$

となって A, B, C の二次式となる。x_i, y_i は測定データで既知つまり定数であるから、これが四乗になっていても問題ないことに注意する。

これを A, B, C で偏微分すると

$$
\begin{aligned}
\frac{\partial \Delta}{\partial A} &= \sum_{i=1}^{N} \left(Ax_i + By_i + C + x_i^2 + y_i^2 \right) x_i = 0 \\
\frac{\partial \Delta}{\partial B} &= \sum_{i=1}^{N} \left(Ax_i + By_i + C + x_i^2 + y_i^2 \right) y_i = 0 \\
\frac{\partial \Delta}{\partial C} &= \sum_{i=1}^{N} \left(Ax_i + By_i + C + x_i^2 + y_i^2 \right) 1 = 0 \qquad (21)
\end{aligned}
$$

従って

$$
\begin{bmatrix} \sum_{i=1}^{N} x_i x_i, & \sum_{i=1}^{N} x_i y_i, & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i y_i, & \sum_{i=1}^{N} y_i y_i, & \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i, & \sum_{i=1}^{N} y_i, & \sum_{i=1}^{N} 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = - \begin{bmatrix} \sum_{i=1}^{N} \left(x_i^2 + y_i^2 \right) x_i \\ \sum_{i=1}^{N} \left(x_i^2 + y_i^2 \right) y_i \\ \sum_{i=1}^{N} \left(x_i^2 + y_i^2 \right) 1 \end{bmatrix} \qquad (22)
$$

後は、この連立方程式を解いて A, B, C を決定し、式 (19) に代入して

$$
a = -\frac{A}{2}, \qquad (23)
$$

$$
b = -\frac{B}{2}, \qquad (24)
$$

$$
r = \sqrt{a^2 + b^2 - C} = \frac{\sqrt{A^2 + B^2 - 4C}}{2} \qquad (25)
$$

が得られる。

4.2 球の場合

球の場合は、変数が増えるだけで、円の場合と全く同じである。
中心 (a, b, c) 半径 r の球の場合、点 (x_i, y_i, z_i) がこの球の表面上にある時

$$
(x_i - a)^2 + (y_i - b)^2 + (z_i - c) - r^2 = 0 \qquad (26)
$$

を満たす筈である。よって誤差 ε_i は

$$
\begin{aligned}
\varepsilon_i &= (x_i - a)^2 + (y_i - b)^2 + (z_i - c)^2 - r^2 \\
&= x_i^2 - 2ax_i + a^2 + y_i^2 - 2by_i + b^2 + z_i^2 - 2cz_i + c^2 - r^2 \\
&= -2ax_i - 2by_i - 2cz_i + (a^2 + b^2 + c^2 - r^2) + x_i^2 + y_i^2 + z_i^2 \\
&= Ax_i + By_i + Cz_i + D + x_i^2 + y_i^2 + z_i^2 \qquad (27)
\end{aligned}
$$

となる。ただし

$$
\begin{aligned}
A &= -2a, \\
B &= -2b, \\
C &= -2c, \\
D &= a^2 + b^2 + c^2 - r^2
\end{aligned}
\tag{28}
$$

である。すると式 (3) は

$$
\Delta := \frac{1}{2}\sum_{i=1}^{N} \varepsilon_i^2 = \frac{1}{2}\sum_{i=1}^{N} \left(Ax_i + By_i + Cz_i + D + x_i^2 + y_i^2 + z_i^2\right)^2 \tag{29}
$$

となって、これを A, B, C, D で偏微分すると

$$
\begin{aligned}
\frac{\partial \Delta}{\partial A} &= \sum_{i=1}^{N} \left(Ax_i + By_i + Cz_i + D + x_i^2 + y_i^2 + z_i^2\right) x_i = 0 \\
\frac{\partial \Delta}{\partial B} &= \sum_{i=1}^{N} \left(Ax_i + By_i + Cz_i + D + x_i^2 + y_i^2 + z_i^2\right) y_i = 0 \\
\frac{\partial \Delta}{\partial C} &= \sum_{i=1}^{N} \left(Ax_i + By_i + Cz_i + D + x_i^2 + y_i^2 + z_i^2\right) z_i = 0 \\
\frac{\partial \Delta}{\partial D} &= \sum_{i=1}^{N} \left(Ax_i + By_i + Cz_i + D + x_i^2 + y_i^2 + z_i^2\right) 1 = 0
\end{aligned}
\tag{30}
$$

従って

$$
\begin{bmatrix}
\sum_{i=1}^{N} x_i x_i, & \sum_{i=1}^{N} y_i x_i, & \sum_{i=1}^{N} z_i x_i, & \sum_{i=1}^{N} x_i \\
\sum_{i=1}^{N} x_i y_i, & \sum_{i=1}^{N} y_i y_i, & \sum_{i=1}^{N} z_i y_i, & \sum_{i=1}^{N} y_i \\
\sum_{i=1}^{N} x_i z_i, & \sum_{i=1}^{N} y_i z_i, & \sum_{i=1}^{N} z_i z_i, & \sum_{i=1}^{N} z_i \\
\sum_{i=1}^{N} x_i, & \sum_{i=1}^{N} y_i, & \sum_{i=1}^{N} z_i, & \sum_{i=1}^{N} 1
\end{bmatrix}
\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix}
= -
\begin{bmatrix}
\sum_{i=1}^{N} \left(x_i^2 + y_i^2 + z_i^2\right) x_i \\
\sum_{i=1}^{N} \left(x_i^2 + y_i^2 + z_i^2\right) y_i \\
\sum_{i=1}^{N} \left(x_i^2 + y_i^2 + z_i^2\right) z_i \\
\sum_{i=1}^{N} \left(x_i^2 + y_i^2 + z_i^2\right) 1
\end{bmatrix}
\tag{31}
$$

後は、この連立方程式を解いて A, B, C, D を決定し、式 (28) に代入して

$$a = -\frac{A}{2}, \tag{32}$$
$$b = -\frac{B}{2}, \tag{33}$$
$$c = -\frac{C}{2}, \tag{34}$$
$$r = \sqrt{a^2 + b^2 + c^2 - D} = \frac{\sqrt{A^2 + B^2 + C^2 - 4D}}{2} \tag{35}$$

が得られる。

5 (応用) 疑似逆行列

行列 A が正方行列であって、逆行列 A^{-1} が存在する時、連立方程式 $A\boldsymbol{x} = \boldsymbol{b}$ は、$\boldsymbol{x} = A^{-1}\boldsymbol{b}$ と解くことができる。

ここでは A が正方行列でない場合を考える。今回は、行の方が多いつまり A が $m \times n$ 行列の時に $m > n$ である場合だけ考えよう。この場合、連立方程式で言うならば、変数の数 n よりも式の数 m の方が多いのだから、一般には解が存在しない。そこで、近似解を求めることで妥協する。

まず、連立方程式の右辺を移項して、右辺をゼロに変形しておく。

$$A\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0} \tag{36}$$

このような x は存在しないから、先ほどまでの最小二乗法と同じ考えで、誤差に相当する

$$\|A\boldsymbol{x} - \boldsymbol{b}\| \tag{37}$$

が最小になるように $\boldsymbol{x}$ を求めることにする。

$m \times n$ 行列 A に対して、以下の 4 条件を満たす $n \times m$ 行列 A^+ がただ一つ定まる：

$$AA^+A = A, \tag{38}$$
$$A^+AA^+ = A^+ \tag{39}$$
$$(AA^+)^* = AA^+ \tag{40}$$
$$(A^+A)^* = A^+A \tag{41}$$

なお、A^* は A の随伴行列 (各要素の複素共役を取りさらに転置した行列) を表す。

この行列 A^+ を A の疑似逆行列と呼ぶ。A が正則でなくても A^+ は定まる。A が正則ならば A の逆行列はこの性質を満たす (つまり $A^+ = A^{-1}$) 。

詳しくは省略するが A の各列が線形独立であれば、A^*A は可逆であり、

$$A^+ = (A^*A)^{-1} A^* \tag{42}$$

が成り立つ。この時 存在しない A^{-1} の代わりに A^+ を用いることで

$$\boldsymbol{x} = A^+\boldsymbol{b} \tag{43}$$

とすれば、この $\boldsymbol{x}$ は 式 (37) を最小にするベクトルである。