

2024 年度 4 ターム「モデリングとシミュレーション」

レポートその 3

最小二乗法、散布図・相関解析

教育学部 第二類 技術・情報系コース 3 年
B220052 長田 麗生

提出期限：2025 年 1 月 29 日

提出日：2025 年 2 月 13 日

担当教員：田中秀幸 先生

1. はじめに

筆者はこのレポートを自らの学習状況を示すことを目的として書く。定義、証明、定式化、等の厳密さについては、本講義で求められていると筆者が考える水準に照らして、問題ないと思える範囲において放棄する。また線型代数はごく一般的な分野であるため、広く知られているであろうことについては参考文献を記載しない。

このレポートを書くにあたって、授業を受けていない人が読んで学習できる程度に、流れよく書くことを心がけた。自らも線型代数や統計についての知識が浅いため、式の解釈など理解の補助に LLM を活用したこともここで述べておく。

2. 一般の次元のベクトル・行列

(余裕がないので省略)

3. 最小二乗法

3.1. 導入

システムに対して x を与えると y という値が得られるとする。 n 個のデータ、すなわち x_1, \dots, x_n とそれに対応する y_1, \dots, y_n があるとき、これらを材料にシステムの挙動を予測したい。システムを近似するモデルとして、2 つの定数 a, b で表される 1 次関数 $\hat{y} = ax + b$ を考えてみる。いま、 a, b は自由に決めることができるが近似の良さはその取り方次第である。

モデルが良いモデルであるとは、もしくは、モデルがシステムを良く近似するとは、より具体的に言い換えるとどういうことだろうか。誤差という言葉を使えば、ざっくり「モデルとシステムの誤差が小さい」と言ってよいだろう。つまり、誤差を小さくするような a, b を求めることが目標となる。データは複数あるので全体を加味して誤差を評価するのが妥当だろう。

誤差が正規分布に従う場合、「誤差がどれだけ小さいか」の評価には、誤差 $e = \hat{y} - y$ の2乗和を指標にし、その最小化を目指すという良いことが知られている※1。一般に、何かを最適化するとき、指標となるような関数を評価関数と呼び、judge の頭文字を取ってよく J で表す。今は a, b に対する誤差の二乗和を指標とするので、評価関数 J は次のようになる。

$$J(a, b) = \sum_{k=1}^n e_k^2 = \sum_{k=1}^n (\hat{y} - y_k)^2 = \sum_{k=1}^n (ax_k + b - y_k)^2$$

さて、議論を簡単にするために行列を使った書き方に直しておこう。まず n 個のデータ x_1, \dots, x_n をまとめて

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

のように書く。また、 \mathbf{x} の成分をシステム、モデルに与えて得られるものについてもそれぞれ同様に、

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \hat{\mathbf{y}} = \begin{pmatrix} ax_1 + b \\ \vdots \\ ax_n + b \end{pmatrix}$$

とおく。さらに、

$$\mathbf{X} = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} a \\ b \end{pmatrix}$$

とおくと n 個の関係式 $y_k = ax_k + b \quad (k = 1, \dots, n)$ を

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{X}\boldsymbol{\theta} \\ \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} &= \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \end{aligned}$$

※1 参考：<https://oroshi.me/2021/01/lsm>

と表すことができる。誤差についても同様に $e = \hat{y} - y$ と直す。すると上の式はこの e を用いて次のように表せる。

$$e = X\theta - y$$

したがって、最小二乗評価関数も次のように書き換えることができる。

$$J(\theta) = \sum_{k=1}^n e_k^2 = e^\top e = (X\theta - y)^\top (X\theta - y)$$

「目標」も行列のことばで書き直せば「 $J(\theta)$ を最小化する θ を求めること」になる。そのような θ を θ_* と書くことにする。

定義 3.1.1: $\theta_* := \arg \min J(\theta)$

3.2. 仮に誤差がなければ

$y = X\theta$ が成り立っているということである。この式を順に変形し θ を既知の値で表したい。

$$\begin{aligned} y &= X\theta \\ X^\top y &= X^\top X\theta \end{aligned}$$

$X^\top X$ が正則である($X^\top X$ に逆行列が存在する)ならば、両辺に左から $(X^\top X)^{-1}$ をかけることで

$$\theta = (X^\top X)^{-1} X^\top y$$

と θ を計算することができる。誤差が一切ない状況を考えているので $J(\theta) = 0$ である。

3.3. 誤差がある場合

実は誤差がある場合も $\theta = (X^\top X)^{-1} X^\top y$ のときに $J(\theta)$ は最小となる。。つまり、

$$\theta_* = (X^\top X)^{-1} X^\top y$$

である。誤差があっても $\theta = \theta_*$ のときに $J(\theta)$ が最小となることを、続く問題と解答で説明する。

問題の前に、今一度記号を整理しておく。

- \mathbf{y} の上に「 $\hat{}$ 」が付いていればそれはモデルから得られる値を、付いていなければそれはシステムから得られる値を表す。
- 「 $*$ 」が右下に付く記号は、最適なときの値を表す。
- $\boldsymbol{\theta}$ はモデルのパラメータを表し、最適なモデルのパラメータ $\boldsymbol{\theta}_*$ を求めることが目標である。
- \mathbf{e} は誤差を表し、 $\mathbf{e} := \mathbf{y} - \hat{\mathbf{y}}$ で定義される

問題 3.3.1: 任意の $\boldsymbol{\theta}$ について、以下が成り立つことを示せ。

$$(\boldsymbol{\theta} - \boldsymbol{\theta}_*)^\top (X^\top X) (\boldsymbol{\theta} - \boldsymbol{\theta}_*) \geq 0$$

証明: $X(\boldsymbol{\theta} - \boldsymbol{\theta}_*)$ がベクトルであることに注意すると、単に $X(\boldsymbol{\theta} - \boldsymbol{\theta}_*)$ のノルムの2乗を表す式であるとわかる。

$$\begin{aligned} & (\boldsymbol{\theta} - \boldsymbol{\theta}_*)^\top (X^\top X) (\boldsymbol{\theta} - \boldsymbol{\theta}_*) \\ &= ((\boldsymbol{\theta} - \boldsymbol{\theta}_*)^\top X^\top) (X(\boldsymbol{\theta} - \boldsymbol{\theta}_*)) \\ &= (X(\boldsymbol{\theta} - \boldsymbol{\theta}_*))^\top (X(\boldsymbol{\theta} - \boldsymbol{\theta}_*)) \\ &= \|X(\boldsymbol{\theta} - \boldsymbol{\theta}_*)\|^2 \geq 0 \end{aligned}$$

ちなみに $\boldsymbol{\theta} = \boldsymbol{\theta}_*$ のとき、 $X(\boldsymbol{\theta} - \boldsymbol{\theta}_*) = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ であるから等号が成り立つ。 ■

問題 3.3.2: 次の等式を示せ。

$$J(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \boldsymbol{\theta}_*)^\top X^\top X (\boldsymbol{\theta} - \boldsymbol{\theta}_*) + \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top X (X^\top X)^{-1} X^\top \mathbf{y}$$

証明: まず $\boldsymbol{\theta}_*$ に関する式を整理しておく。 $\boldsymbol{\theta}_* = (X^\top X)^{-1} X^\top \mathbf{y}$ の両辺に左側から $X^\top X$ をかけて次の式を得る。

$$X^\top X \boldsymbol{\theta}_* = X^\top \mathbf{y}$$

また $\boldsymbol{\theta}_*$ の転置について、 $(X^\top X)^{-1}$ が対称行列であることに注意すると

$$\begin{aligned}
\boldsymbol{\theta}_*^\top &= \left((X^\top X)^{-1} X^\top \mathbf{y} \right)^\top \\
&= \mathbf{y}^\top X \left((X^\top X)^{-1} \right)^\top \\
&= \mathbf{y}^\top X (X^\top X)^{-1}
\end{aligned}$$

となる。これらを使って $J(\boldsymbol{\theta}) = \mathbf{e}^\top \mathbf{e} = (X\boldsymbol{\theta} - \mathbf{y})^\top (X\boldsymbol{\theta} - \mathbf{y})$ を展開した式を整理する。

$$\begin{aligned}
J(\boldsymbol{\theta}) &= (X\boldsymbol{\theta} - \mathbf{y})^\top (X\boldsymbol{\theta} - \mathbf{y}) \\
&= (X\boldsymbol{\theta})^\top (X\boldsymbol{\theta}) - (X\boldsymbol{\theta})^\top \mathbf{y} - \mathbf{y}^\top (X\boldsymbol{\theta}) + \mathbf{y}^\top \mathbf{y} \\
&= (X\boldsymbol{\theta})^\top (X\boldsymbol{\theta}) - 2(X\boldsymbol{\theta})^\top \mathbf{y} + \mathbf{y}^\top \mathbf{y} \quad (\because \text{ベクトルの内積の交換法則}) \\
&= (X\boldsymbol{\theta})^\top (X\boldsymbol{\theta}) - 2\boldsymbol{\theta}^\top X^\top \mathbf{y} + \mathbf{y}^\top \mathbf{y} \\
&= (X\boldsymbol{\theta})^\top (X\boldsymbol{\theta}) - 2\boldsymbol{\theta}^\top X^\top X\boldsymbol{\theta}_* + \mathbf{y}^\top \mathbf{y} \quad (\because X^\top X\boldsymbol{\theta}_* = X^\top \mathbf{y}) \\
&= (X\boldsymbol{\theta})^\top (X\boldsymbol{\theta}) - 2(X\boldsymbol{\theta})^\top (X\boldsymbol{\theta}_*) + \mathbf{y}^\top \mathbf{y} \\
&= \left\{ (X\boldsymbol{\theta})^\top (X\boldsymbol{\theta}) - 2(X\boldsymbol{\theta})^\top (X\boldsymbol{\theta}_*) + (X\boldsymbol{\theta}_*)^\top (X\boldsymbol{\theta}_*) \right\} + \mathbf{y}^\top \mathbf{y} - (X\boldsymbol{\theta}_*)^\top (X\boldsymbol{\theta}_*) \\
&= (X\boldsymbol{\theta} - X\boldsymbol{\theta}_*)^\top (X\boldsymbol{\theta} - X\boldsymbol{\theta}_*) + \mathbf{y}^\top \mathbf{y} - (X\boldsymbol{\theta}_*)^\top (X\boldsymbol{\theta}_*) \\
&= (X(\boldsymbol{\theta} - \boldsymbol{\theta}_*))^\top (X(\boldsymbol{\theta} - \boldsymbol{\theta}_*)) + \mathbf{y}^\top \mathbf{y} - (X\boldsymbol{\theta}_*)^\top (X\boldsymbol{\theta}_*) \\
&= (\boldsymbol{\theta} - \boldsymbol{\theta}_*)^\top X^\top X(\boldsymbol{\theta} - \boldsymbol{\theta}_*) + \mathbf{y}^\top \mathbf{y} - \boldsymbol{\theta}_*^\top X^\top X\boldsymbol{\theta}_* \\
&= (\boldsymbol{\theta} - \boldsymbol{\theta}_*)^\top X^\top X(\boldsymbol{\theta} - \boldsymbol{\theta}_*) + \mathbf{y}^\top \mathbf{y} - (\boldsymbol{\theta}_*^\top)(X^\top X\boldsymbol{\theta}_*) \\
&= (\boldsymbol{\theta} - \boldsymbol{\theta}_*)^\top X^\top X(\boldsymbol{\theta} - \boldsymbol{\theta}_*) + \mathbf{y}^\top \mathbf{y} - \left(\mathbf{y}^\top X (X^\top X)^{-1} \right) (X^\top \mathbf{y}) \\
&= (\boldsymbol{\theta} - \boldsymbol{\theta}_*)^\top X^\top X(\boldsymbol{\theta} - \boldsymbol{\theta}_*) + \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top X (X^\top X)^{-1} X^\top \mathbf{y}
\end{aligned}$$

■

この等式を自分の思うきれいな形で書くと次のような感じだろうか。

$$\begin{aligned}
J(\boldsymbol{\theta}) &= \|X(\boldsymbol{\theta} - \boldsymbol{\theta}_*)\|^2 + \|\mathbf{y}\|^2 - \|X\boldsymbol{\theta}_*\|^2 \\
&= \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_*\|^2 + \|\mathbf{y}\|^2 - \|\hat{\mathbf{y}}_*\|^2
\end{aligned}$$

$J(\boldsymbol{\theta})$ は $\boldsymbol{\theta} = \boldsymbol{\theta}_*$ のとき最小となるが、このとき右辺の第1項が0となるので改めて式を書くと

$$\begin{aligned}
J(\boldsymbol{\theta}_*) &= \|\mathbf{e}_*\|^2 = \|\mathbf{y}\|^2 - \|X\boldsymbol{\theta}_*\|^2 \\
&= \|\mathbf{y}\|^2 - \|\hat{\mathbf{y}}_*\|^2
\end{aligned}$$

のようになる。三平方の定理の逆から、 $\mathbf{y}, \hat{\mathbf{y}}_*, \mathbf{e}_*$ は直角三角形をなすと言える。

3.4. 最小二乗法の幾何学的解釈

次の図1のように n 次元空間の絵を書いてみる。3次元っぽい表現しかできないが、都合の良いことにこれで十分説明がつく。この絵は見やすさのために $x, 1$ がなす平面が「水平」になるように書いている。さて、ベクトル $x, y, 1, e, \hat{y}$ の関係を見てみよう。

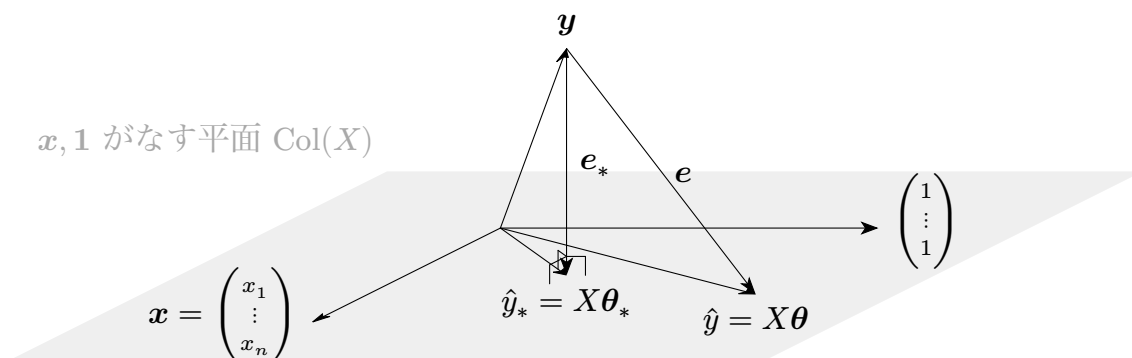


図 1: n 次元空間中の各ベクトル

モデルから得られる \hat{y} は $x, 1$ がなす平面上にある。 $\hat{y} = ax + b1$ と書けることから明らかである。 $\theta = \begin{pmatrix} a \\ b \end{pmatrix}$ が動けば、それに合わせて \hat{y} はこの平面上を動く。

$J(\theta)$ を最小化するとは、 $\|e\|$ を最小化することに等しい。 y が偶然 $x, 1$ がなす平面の上であれば、 $J(\theta) = 0$ となるように θ を取ってこれる。これが第3.2節の話である。一般の場合についても、この図を見れば $\|e\|$ が最小になるのは e と \hat{y} が直交するときだとわかる。

$J(\theta)$ を平方完成したときの $X(\theta - \theta_*)$ の部分も見て取れる。確かに、そこが離れている分だけ e は最小値 e_* から三平方の定理的に大きくなる。

3.5. 「 θ_* 」はどこから来たのか

はじめはある種、天下りの θ_* が与えられたが、構成的に求めることもできる。次のように、幾何学的な表現を使うとわかりやすく説明できるだろう。

e は y の終点から出て $x, 1$ がなす平面 $\text{Col}(X)^{\text{※2}}$ 上の \hat{y} に落ちるベクトルである。 e の大きさを最小化したければ、「 e と平面 $\text{Col}(X)$ が直交する」ようにすればよい。この条件は「 e が $x, 1$ とそれぞれ直交する」、すなわち「 $(x^T e = 0) \wedge (1^T e = 0)$ 」と言い換えることができ、表現を変えれば「 $X^T e = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ 」となる。 $e = X\theta - y$ だったので結局のところ $X^T(X\theta - y) = 0$ となるような θ を求めればよく、それを θ_* としていたのだ。ここから式を変形することで $\theta_* = (X^T X)^{-1} X^T y$ が得られる。

※2 X の列空間 という。

4. 散布図・相関解析

2変数の関係を見たい場合、一方を横軸に、もう一方を縦軸に取ってデータをプロットする散布図を使用すると良い。相関係数 r は、2変数の関係性を表す指標である。

5. 実習の記録

5.1. 第9回 実習(1)–(3)

- 対角行列: `diag(a_1, a_2, ..., a_n)`
- 逆行列: `inv(A)`
- 行列の積: `A * B`
- 行列の転置: `A'`

ランダムに生成したデータに対して、最小二乗法を適用して近似曲線を求めるプログラムを作成した。プロットした結果を図2に示す。

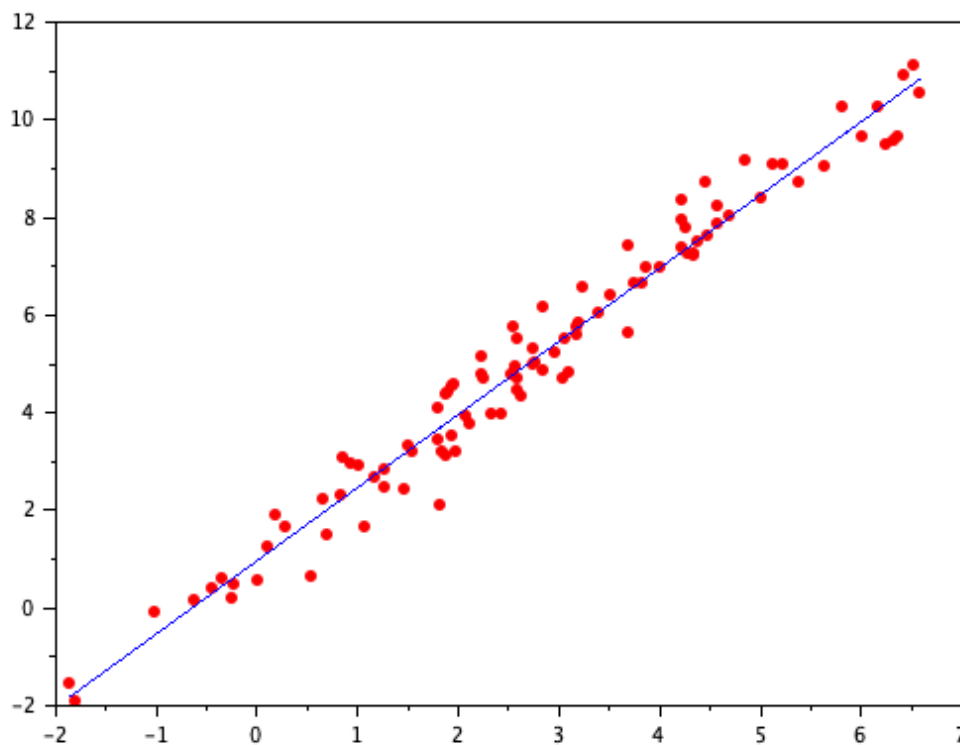


図 2: 散布図と近似曲線のプロット

```

1  //-----
2  // 最小二乗法
3  mode(0);
4
5  N = 100;
6
7  x = 2*rand(N, 1, 'normal') + 3;
8
9  a = 1.5;
10 b = 1;
11 e = 0.5*rand(N, 1, 'normal'); // 雑音
12 y = a*x + b + e;
13
14 scf(1);
15 plot(x, y, 'r. ');
16
17 // 最小二乗法により、解を求める。
18 X = [x, ones(N,1)];
19 theta = inv(X'*X) * X' * y; // a, b を推定する。
20 a_ = theta(1)
21 b_ = theta(2)
22
23 // 結果をプロット
24 y_ = zeros(N,1);
25 for k = 1:N,
26     y_(k) = a_ * x(k) + b_; // y_ は推定値
27 end
28 plot(x,y_);

```

コード 1: 実行したコード

5.2. 第 10 回 実習(1)–(4)

Scilab には統計分析に便利な関数が用意されている。

- 散布図: `plot(x, y)`
- 平均: `mean(x)`
- 分散: `variance(x)`
- 共分散: `cov(x, y)`
- 標準偏差: `stdev(x)`
- 相関係数: `correl(x, y)`

6. 参考文献

1. 「最小二乗法は誤差を正規分布と仮定した最尤推定である」, <https://oroshi.me/2021/01/lsm>, 2025 年 1 月 21 日閲覧