回帰分析 ～線形代数の視点から～

Ver.2022.09.03
松浦 真也

1 重回帰分析の定式化

ある大学の m 人の学生の n 科目の成績のデータを扱う．i 人目の j 番目の科目の点数を u_{ij} で表す．例えば, 学生の名前が一郎君, 二郎君, 三郎君, $\ldots$ で, 科目名が代数 1, 代数 2, 代数 3, $\ldots$ だとすると, u_{32} は「三郎君の代数 2 の点数」を表す．

さらに, i 人目の学生がこの大学を受験した際の, 入試の点数を t_i とする．

氏名	入試	代数 1	代数 2	代数 3	$\cdots$	代数 n
一郎君	t_1	u_{11}	u_{12}	u_{13}	$\cdots$	u_{1n}
二郎君	t_2	u_{21}	u_{22}	u_{23}	$\cdots$	u_{2n}
三郎君	t_3	u_{31}	u_{32}	u_{33}	$\cdots$	u_{3n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m 郎君	t_m	u_{m1}	u_{m2}	u_{m3}	$\cdots$	u_{mn}

ここで, 学生の成績をもとに, 入試の点数を推定することを考える．つまり, u_{ij} $(j = 1, 2, \ldots, n)$ から t_i を推測したい．例えば, 三郎君の入試の成績 t_3 を, 彼の代数 1 の成績 u_{31}, 代数 2 の成績 $u_{32}, \ldots$, 代数 n の成績 u_{3n} から推定する．これを数学的な言葉で言うと, n 変数関数 $g(x_1, x_2, \ldots, x_n)$ を上手く選んで,

$$t_i = g(u_{i1}, u_{i2}, \ldots, u_{in}) + \varepsilon_i \qquad (i = 1, 2, \ldots, m) \tag{1}$$

と表現するということである．ただし, ε_i は推定の際の誤差であり, この誤差をなるべく小さくするように, 関数 g を選びたい．

関数 g として, あまり複雑なものを用いると, 理論的な解析が困難になる．そのため, g が 1 次関数の場合を扱う．つまり,

$$g(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} = a_j x_j + b \qquad (a_j, b \in \mathbb{R})$$

のときについて考える．この場合, (1) は

$$t_i = \sum_{j=1}^{n} a_j u_{ij} + b + \varepsilon_i \qquad (i = 1, 2, \ldots, m) \tag{2}$$

となる．この式の a_j と b を上手く選んで, 誤差 ε_i を最小に抑えたい．

(2) を行列とベクトルを用いて表しておくと, 後の計算がスマートに行える. そこで,

$$U = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n, \boldsymbol{1}) = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} & 1 \\ u_{21} & u_{22} & \cdots & u_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mn} & 1 \end{pmatrix}, \tag{3}$$

$$\boldsymbol{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{pmatrix}, \qquad \boldsymbol{c} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ b \end{pmatrix}, \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} \tag{4}$$

とおく. すると (2) は

$$\boldsymbol{t} = U\boldsymbol{c} + \boldsymbol{\varepsilon} \tag{5}$$

あるいは

$$\boldsymbol{\varepsilon} = \boldsymbol{t} - U\boldsymbol{c} \tag{6}$$

と書ける. 以上より, 我々の考えるべき問題は, 次のようになる.

問題 P1 次の量を最小化するように, $\boldsymbol{c} \in \mathbb{R}^{n+1}$ を定めよ.

$$\|\boldsymbol{\varepsilon}\|^2 = \|\boldsymbol{t} - U\boldsymbol{c}\|^2 = {}^t\left(\boldsymbol{t} - U\boldsymbol{c}\right)\left(\boldsymbol{t} - U\boldsymbol{c}\right).$$

ただし, 一般にベクトル $\boldsymbol{v}$ や行列 V に対して, $\|\boldsymbol{v}\|$ は $\boldsymbol{v}$ のユークリッドノルムを表し, ${}^tV$ は V の転置行列を表す.

2 ベクトルの近似

この節では, 前節の問題 P1 を解くための一般論を述べる (一般論のため, ここで用いる記号は, 前節での記号と直接は関係ない).

ベクトル $\boldsymbol{w} \in \mathbb{R}^m$ を, n 個のベクトル $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \in \mathbb{R}^m$ の 1 次結合で近似したい. ここで, 「近似の良さ」として, $\boldsymbol{w}$ からのズレのユークリッドノルム (の 2 乗) を用いる. つまり,

$$f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n) = \left\| \boldsymbol{w} - \sum_{i=1}^{n} x_i \boldsymbol{v}_i \right\|^2$$

を最小にする $\boldsymbol{x} = {}^t(x_1, x_2, \ldots, x_n)$ を求めたい. ここで, $m \times n$ 行列 V を

$$V = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)$$

と定義する. このとき,

$$f(\boldsymbol{x}) = \|\boldsymbol{w} - V\boldsymbol{x}\|^2 = {}^t\left(\boldsymbol{w} - V\boldsymbol{x}\right)\left(\boldsymbol{w} - V\boldsymbol{x}\right) = {}^t\boldsymbol{x}\,{}^tVV\boldsymbol{x} - 2\,{}^t\boldsymbol{w}V\boldsymbol{x} + {}^t\boldsymbol{w}\boldsymbol{w} \tag{7}$$

となる (${}^t\boldsymbol{w}V\boldsymbol{x}$ はベクトルや行列ではなく, スカラーになるので, ${}^t\boldsymbol{w}V\boldsymbol{x} = {}^t({}^t\boldsymbol{w}V\boldsymbol{x})$, つまり ${}^t\boldsymbol{w}V\boldsymbol{x} = {}^t\boldsymbol{x}\,{}^tV\boldsymbol{w}$ であることを用いた).

ところで, $f(\boldsymbol{x})$ を最小とする $\boldsymbol{x}$ は

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = 0 \qquad (i = 1, 2, \ldots, n) \tag{8}$$

の解になっていないといけない. ここで,

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{pmatrix} \dfrac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_n} \end{pmatrix}$$

と書くことにすれば, (8) は,

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = 0 \tag{9}$$

とまとめて表現できる. (9) を解くために, 補題を 1 つ用意する.

補題 1 A を n 次対称行列, $\boldsymbol{b}$ を n 次元縦ベクトルとし, ともに各成分は定数とする. このとき, 次が成り立つ.

(i) $\dfrac{\partial}{\partial \boldsymbol{x}}({}^t\boldsymbol{x}A\boldsymbol{x}) = 2A\boldsymbol{x}$,

(ii) $\dfrac{\partial}{\partial \boldsymbol{x}}({}^t\boldsymbol{b}\boldsymbol{x}) = \boldsymbol{b}$,

(iii) $\dfrac{\partial}{\partial \boldsymbol{x}}({}^t\boldsymbol{b}\boldsymbol{b}) = 0$.

練習問題 1 補題 1 を証明せよ.

補題 1 を用いて (9) の微分を具体的に計算すると, (7) より

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}}\left({}^t\boldsymbol{x}\,{}^tVV\boldsymbol{x} - 2\,{}^t\boldsymbol{w}V\boldsymbol{x} + {}^t\boldsymbol{w}\boldsymbol{w}\right) = 2\,{}^tVV\boldsymbol{x} - 2\,{}^tV\boldsymbol{w} \tag{10}$$

となる. したがって, (9) が満たされるのは

$${}^tVV\boldsymbol{x} = {}^tV\boldsymbol{w}$$

のときである. 特に, ${}^tVV$ が正則ならば

$$\boldsymbol{x} = \left({}^tVV\right)^{-1}\,{}^tV\boldsymbol{w}$$

を得る.
　以上より, 次の定理が示された.

定理 1　${}^tVV$ が正則のとき,

$$f(\boldsymbol{x}) = \|\boldsymbol{w} - V\boldsymbol{x}\|^2 = \left\|\boldsymbol{w} - \sum_{i=1}^{n} x_i \boldsymbol{v}_i\right\|^2$$

を最小にする $\boldsymbol{x}$ は

$$\boldsymbol{x} = \left({}^tVV\right)^{-1} {}^tV\boldsymbol{w}$$

で与えられる.

練習問題 2　${}^tVV$ が正則となるための必要十分条件を, ベクトル $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \in \mathbb{R}^m$ を用いて表せ. 結論だけではなく，導出過程も記すこと．特に, $n \leq m$ でないといけないことを示せ.

3　重回帰分析の計算

　この節では, 再び第 1 節の問題 P1 に戻る. 以下では, ${}^tUU$ は正則であるとする. 第 2 節の定理 1 を用いれば, この問題の答えは, 次のように与えられる.

問題 P1 の解　$\|\boldsymbol{\varepsilon}\|^2 = \|\boldsymbol{t} - U\boldsymbol{c}\|^2$ を最小にする $\boldsymbol{c} \in \mathbb{R}^{n+1}$ は

$$\boldsymbol{c} = \left({}^tUU\right)^{-1} {}^tU\boldsymbol{t} \tag{11}$$

である.

　これで一応, 問題は解決したが, もう少し整理しなおした方が, 統計学的な意味が分かりやすくなる. まずは, 次の事実に注意する.

定理 2　$\|\boldsymbol{\varepsilon}\|^2$ が (11) により最小化されているとき, $\mathbf{1}$ と $\boldsymbol{\varepsilon}$ は直交する. つまり,

$$(\mathbf{1}, \boldsymbol{\varepsilon}) = (1, 1, \ldots, 1)\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} = \sum_{i=1}^{m} \varepsilon_i = 0 \tag{12}$$

である. 言い換えれば, 誤差の標本平均 $\dfrac{1}{m}\displaystyle\sum_{i=1}^{m} \varepsilon_i$ が 0 となる.

証明　(11) より,

$$\boldsymbol{\varepsilon} = \boldsymbol{t} - U\boldsymbol{c} = \boldsymbol{t} - U\left({}^tUU\right)^{-1} {}^tU\boldsymbol{t}$$

となる．したがって，

$$ {}^tU\boldsymbol{\varepsilon} = {}^tU\boldsymbol{t} - {}^tUU\left({}^tUU\right)^{-1}{}^tU\boldsymbol{t} = {}^tU\boldsymbol{t} - {}^tU\boldsymbol{t} = 0 \tag{13} $$

である．ここで, (3) での定義を思い出すと, ${}^tU$ の $n+1$ 行目は $(1,1,\ldots,1)$ である．よって, (13) の両辺の $n+1$ 行目を比較すれば, (12) が得られる．　□

ここで, t_i および u_{ij} (j は固定) の標本平均を, それぞれ $\bar{t}$, $\overline{u}_j$ と記す．すなわち,

$$ \bar{t} = \frac{1}{m}\sum_{i=1}^{m} t_i, \qquad \overline{u}_j = \frac{1}{m}\sum_{i=1}^{m} u_{ij} \qquad (j=1,2,\ldots,n) \tag{14} $$

とする．さらに,

$$ v_{ij} = u_{ij} - \overline{u}_j, \qquad w_i = t_i - \bar{t} \qquad (i=1,2,\ldots,m,\ j=1,2,\ldots,n) \tag{15} $$

$$ V = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{pmatrix}, \qquad \boldsymbol{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix}, \qquad \boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \tag{16} $$

とおく．このとき, 以下の関係が成り立つ．

補題 2

(i) $\dfrac{(\mathbf{1},\boldsymbol{t})}{m} = \bar{t}, \qquad \dfrac{1}{m}\,{}^t\mathbf{1}\,U = (\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_n, 1),$

(ii) $\boldsymbol{t} - \dfrac{(\mathbf{1},\boldsymbol{t})}{m}\mathbf{1} = \boldsymbol{w}, \qquad \left(U - \dfrac{1}{m}\mathbf{1}\,{}^t\mathbf{1}U\right)\boldsymbol{c} = V\boldsymbol{a}.$

練習問題 3　補題 2 を証明せよ．

定理 3　(5) の式

$$ \boldsymbol{t} = U\boldsymbol{c} + \boldsymbol{\varepsilon} $$

において, $(\mathbf{1},\boldsymbol{\varepsilon}) = 0$ となるための必要十分条件は

$$ \boldsymbol{w} = V\boldsymbol{a} + \boldsymbol{\varepsilon} \tag{17} $$

が成り立つことである．

証明　(5) の両辺と $\mathbf{1}$ との内積をとると,

$$ (\mathbf{1},\boldsymbol{t}) = {}^t\mathbf{1}U\boldsymbol{c} + (\mathbf{1},\boldsymbol{\varepsilon}). $$

よって,

$$\frac{(\mathbf{1}, \boldsymbol{t})}{m}\mathbf{1} = \frac{1}{m}\mathbf{1}\,{}^t\mathbf{1}U\boldsymbol{c} + \frac{(\mathbf{1}, \boldsymbol{\varepsilon})}{m}\mathbf{1} \tag{18}$$

となる. (5) の各辺から, (18) の各辺を引き, 補題 2 (ii) を用いると,

$$\boldsymbol{w} = V\boldsymbol{a} + \boldsymbol{\varepsilon} - \frac{(\mathbf{1}, \boldsymbol{\varepsilon})}{m}\mathbf{1}$$

であることがわかる. したがって, (17) が成り立つことと, $(\mathbf{1}, \boldsymbol{\varepsilon}) = 0$ であることは, 同値である. □

定理 2 と定理 3 により, 第 1 節の問題 P1 は, 次の問題で置き換え可能である.

問題 P2 $\|\boldsymbol{w} - V\boldsymbol{a}\|^2$ を最小化するように, $\boldsymbol{a} \in \mathbb{R}^n$ を定めよ.

なぜなら, 定理 2 より, 問題 P1 において, $\|\boldsymbol{\varepsilon}\|$ を最小化したら, $(\mathbf{1}, \boldsymbol{\varepsilon}) = 0$ が成り立つので, 最初から $(\mathbf{1}, \boldsymbol{\varepsilon}) = 0$ の成り立つ範囲で, $\|\boldsymbol{\varepsilon}\|$ を最小化すれば良い. さらに, 定理 3 より, $(\mathbf{1}, \boldsymbol{\varepsilon}) = 0$ は (17) 式を意味するので, (17) 式の成り立つ範囲で $\|\boldsymbol{\varepsilon}\|$ を最小化すれば良い. さらに, (17) 式が成り立つということは, $\boldsymbol{\varepsilon} = \boldsymbol{w} - V\boldsymbol{a}$ なので, 結局, $\|\boldsymbol{w} - V\boldsymbol{a}\|^2$ を最小化することになる.

ところで. 問題 P2 の解は定理 1 より,

$$\boldsymbol{a} = \left({}^tVV\right)^{-1}\,{}^tV\boldsymbol{w} \tag{19}$$

と求まる.

練習問題 4 問題 P1 の解に対して

$$b = \bar{t} - \sum_{j=1}^{m} a_j \bar{u}_j \tag{20}$$

となることを示せ (つまり, $\boldsymbol{a}$ を決めれば, b は自動的に決まる).

4 回帰分析の統計学的な解釈

この節では, (19) を統計学的な観点から見てみる. まず, $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ の分散共分散行列 S, それに $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ と $\boldsymbol{t}$ の共分散ベクトル $\boldsymbol{s}$ を次のように定義する.

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{pmatrix}, \qquad \boldsymbol{s} = \begin{pmatrix} s_{1t} \\ s_{2t} \\ \vdots \\ s_{nt} \end{pmatrix}. \tag{21}$$

ただし, s_{ij} は $\boldsymbol{u}_i$ と $\boldsymbol{u}_j$ の標本共分散, s_{jt} は $\boldsymbol{u}_j$ と $\boldsymbol{t}$ の標本共分散である. つまり,

$$s_{ij} = \frac{1}{m-1}\sum_{k=1}^{m}(u_{ki}-\overline{u}_i)(u_{kj}-\overline{u}_j) = \frac{1}{m-1}\sum_{k=1}^{m} v_{ki}v_{kj} \tag{22}$$

$$s_{jt} = \frac{1}{m-1}\sum_{k=1}^{m}(u_{kj}-\overline{u}_j)(t_k-\overline{t}) = \frac{1}{m-1}\sum_{k=1}^{m} v_{kj}w_k \tag{23}$$

である. ここで, 定義 (16) を思い出すと,

$$\frac{1}{m-1}\,{}^tVV = S, \qquad \frac{1}{m-1}\,{}^tV\boldsymbol{w} = \boldsymbol{s} \tag{24}$$

であることがわかる. これを (19) に代入すると,

$$\boldsymbol{a} = S^{-1}\boldsymbol{s} \tag{25}$$

を得る. つまり, 重回帰分析の係数ベクトル $\boldsymbol{a}$ は, 分散共分散行列 S の逆行列を共分散ベクトル $\boldsymbol{s}$ にかければ求まる.

練習問題 5 分散共分散行列 S が非正則となるのは，データ u_{ij} がどのような性質をもつときか調べよ．また，そのとき，重回帰分析はどのように実行すれば良いか考察せよ．