

第12回 データ解析

(連立方程式, 逆行列, 多変量解析)

目標

- ・連立1次方程式の解法を学ぶ
- ・逆行列の算出法を学ぶ
- ・多変量解析の基礎を学ぶ

0. 準備

今日の作業をするディレクトリを作成しなさい。

```
% mkdir 20190110
% cd 20190110
```

1. 連立1次方程式と逆行列

1.1. 掃き出し法

以下のような連立方程式を考える。

$$\begin{aligned}
 a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\
 a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\
 a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n &= b_3 \\
 &\vdots \\
 a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \cdots + a_{nn}x_n &= b_n
 \end{aligned} \quad \text{--- (1)}$$

上記の連立方程式は以下のように表わすことができる。

$$\left(\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & x_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & x_2 \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & x_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} & x_n \end{array} \right) = \left(\begin{array}{c} b_1 \\ b_2 \\ b_3 \\ \cdots \\ b_n \end{array} \right) \quad \text{--- (2)}$$

この解を求める為には、以下のような行列の基本変形を行う。まず第1行を a_{11} で割る。

$$\left(\begin{array}{ccccc|c} 1 & a_{12}/a_{11} & a_{13}/a_{11} & \cdots & a_{1n}/a_{11} & x_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & x_2 \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & x_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} & x_n \end{array} \right) = \left(\begin{array}{c} b_1/a_{11} \\ b_2 \\ b_3 \\ \cdots \\ b_n \end{array} \right) \quad \text{--- (3)}$$

次に第2行から第n行に対して第1行に a_{21} から a_{n1} をそれぞれ掛けたものを引く。

$$\left(\begin{array}{cccccc} 1 & a_{12}/a_{11} & a_{13}/a_{11} & \cdots & a_{1n}/a_{11} & \\ 0 & a_{22} - a_{21}a_{12}/a_{11} & a_{23} - a_{21}a_{13}/a_{11} & \cdots & a_{2n} - a_{21}a_{1n}/a_{11} & \\ 0 & a_{32} - a_{31}a_{12}/a_{11} & a_{33} - a_{31}a_{13}/a_{11} & \cdots & a_{3n} - a_{31}a_{1n}/a_{11} & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & a_{n2} - a_{n1}a_{12}/a_{11} & a_{n3} - a_{n1}a_{13}/a_{11} & \cdots & a_{nn} - a_{n1}a_{1n}/a_{11} & \end{array} \right) \left(\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_n \end{array} \right) = \left(\begin{array}{c} b_1/a_{11} \\ b_2 - a_{21}b_1/a_{11} \\ b_3 - a_{31}b_1/a_{11} \\ \cdots \\ b_n - a_{n1}b_1/a_{11} \end{array} \right)$$

--- (4)

第2列～第n列に関しても同様な操作を繰り返し

$$\left(\begin{array}{ccccc} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{array} \right) \left(\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_n \end{array} \right) = \left(\begin{array}{c} B_1 \\ B_2 \\ B_3 \\ \cdots \\ B_n \end{array} \right)$$

--- (5)

となったときの右辺が解である。

＜注意点＞

この方法では消去の各段階で $a_{11}, a_{22}, \dots, a_{nn}$ (これらを軸要素と呼ぶ) で除算を行う。軸要素の値が 0 に近いと解の精度が軸要素の精度以下となってしまう。それを避けるため、消去の各段階で a_{ij} の中から絶対値の最大要素 a_{pq} を見つけ出し、 a_{pq} が a_{kk} の位置にくるように方程式と変数の順序の入れ替えを行う必要がある。

掃き出し法のサブルーチン例を以下に示す。

サブルーチンは

<http://www.eps.nagoya-u.ac.jp/~morota/20161215/subroutine01.html>

からダウンロードできる。

```
*****
subroutine GJM(A,n,m)
implicit none
real*8 A(n,m),max,W
integer Work(n),i,j,n,m,k,p,q,iw
do i=1,n
  Work(i)=i
enddo
do k=1,n
  max=abs(A(k,k))
  p=k
  q=k
  do j=k,n
    do i=k,n
      if(max.lt.abs(A(i,j))) then
        max=abs(A(i,j))
        p=i
        q=j
      endif
    enddo
  enddo
  do i=1,n
    W=A(i,k)
    A(i,k)=A(i,q)
    A(i,q)=W
  enddo
  do j=k,m
    W=A(k,j)
    A(k,j)=A(p,j)
    A(p,j)=W
  enddo
  i=Work(k)
  Work(k)=Work(q)
  Work(q)=i
  do j=k+1,m
    A(k,j)=A(k,j)/A(k,k)
  enddo
  do i=1,n
    if(i.ne.k) then
      do j=k+1,m
        do j=k+1,m
          A(i,j)=A(i,j)-A(i,k)*A(k,j)
        enddo
      endif
    enddo
  enddo
  do j=n+1,M
    do i=1,n
      iw=Work(i)
      A(iw,n)=A(i,j)
    enddo
    do i=1,n
      A(i,j)=A(i,n)
    enddo
  enddo
  return
end
*****
```

練習問題 1

上記のサブルーチンを用いて次の連立方程式を解け。この場合、 $n=4, m=5$ 。

$$\begin{aligned} 2x_1 - 2x_2 + x_3 - x_4 &= -3 \\ x_1 + 2x_2 + 4x_3 - x_4 &= -7 \\ -2x_1 + x_2 + 2x_3 + 5x_4 &= 8 \\ 3x_1 + 2x_2 - x_3 - 2x_4 &= 4 \end{aligned} \quad \text{--- (6)}$$

サブルーチンは

<http://www.eps.nagoya-u.ac.jp/~morota/20161215/subroutine01.html>

からダウンロードできる。

1.2. 逆行列

行列 A とその逆行列 B を以下のように定義する。

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} \quad A^{-1} = B = \begin{pmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2n} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & b_{nn} \end{pmatrix} \quad \text{--- (7)}$$

行列 A が与えられたとき、逆行列 B は掃き出し法のアルゴリズムで解くことができる。

定義より、

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2n} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \quad \text{--- (8)}$$

となり、行列 B を分けて表わすと、

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} \\ b_{21} \\ b_{31} \\ \cdots \\ b_{n1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{pmatrix}$$

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{array} \right) \left(\begin{array}{c} b_{12} \\ b_{22} \\ b_{32} \\ \cdots \\ b_{n2} \end{array} \right) = \left(\begin{array}{c} 0 \\ 1 \\ 0 \\ \cdots \\ 0 \end{array} \right)$$

...

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{array} \right) \left(\begin{array}{c} b_{1n} \\ b_{2n} \\ b_{3n} \\ \cdots \\ b_{nn} \end{array} \right) = \left(\begin{array}{c} 0 \\ 0 \\ 0 \\ \cdots \\ 1 \end{array} \right)$$

となり、これは n 個の連立方程式から b_{ij} を求めることであり、掃き出し法のアルゴリズムで解くことができる。つまり、

$$\left(\begin{array}{ccccc|cccc} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & 1 & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & 0 & 1 & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} & 0 & 0 & 0 & \cdots & 1 \end{array} \right) \quad \text{--- (9)}$$

を掃き出し法により以下のかたちに変換したとき

$$\left(\begin{array}{cccc|cccc} 1 & 0 & 0 & \cdots & 0 & A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\ 0 & 1 & 0 & \cdots & 0 & A_{21} & A_{22} & A_{23} & \cdots & A_{2n} \\ 0 & 0 & 1 & \cdots & 0 & A_{31} & A_{32} & A_{33} & \cdots & A_{3n} \\ \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & A_{n1} & A_{n2} & A_{n3} & \cdots & A_{nn} \end{array} \right) \quad \text{--- (10)}$$

行列 A_{ij} が逆行列 b_{ij} を表わしている。

練習問題 2

練習問題 1 のプログラムを書き換えて、逆行列を算出するプログラムをつくれ。このプログラムを用いて、以下の行列の逆行列を算出せよ。

$$A = \begin{pmatrix} -1 & 2 & 1 \\ 2 & -1 & -2 \\ 3 & 5 & -1 \end{pmatrix} \quad \text{--- (11)}$$

2. 多変量解析

多変量解析とは、多くの変数データの間に隠れた傾向を統計的に抽出する方法である。

ここでは多変量解析の導入を学ぶ。

2.1. 重回帰分析

以前、2変量（例えば、yとx）データに対して、目的変数（y）の値を説明変数（x）の関係を一次式で表わせると仮定した際の最小二乗法フィッティングを学んだ（第三回「DOループの応用」参照）。次は多変量データについて考える。

例えば、「テストの点数」と「テスト直前の勉強時間」との関係を考えると、おそらく両者には相関関係があると思われる。もし両者が下記のような一次式で関係付けられるとすると、

$$(テストの点数) = a + b \times (テスト直前の勉強時間)$$

複数人のデータからaとbの値を最小二乗法で決定することで、勉強時間からテストの点数を予測することができるようになる。しかし実際には、「テスト直前の勉強時間」だけでなく、「講義の出席回数」や「課題の提出回数」、「IQ」によっても点数は変わるだろう。その場合、よりよいテストの点数の予測は

$$(テストの点数) = a_0 + a_1 \times (テスト直前の勉強時間) + a_2 \times (講義の出席回数) + a_3 \times (課題の提出回数) + a_4 \times (IQ) + \dots$$

から得られる。このように1つの目的変数（この場合、テストの点数）を複数（n個）の説明変数で説明・予測できると仮定し、 $a_0 \sim a_n$ を決定する分析手法を重回帰分析という。

ここでは一次式でのフィッティングを考える。

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n = a_0 + \sum_{i=1}^n a_i x_i \quad \text{--- (12)}$$

関数のフィッティングの条件は、観測データ（ここでデータ数はp個、データの番号をjで表す）と関数との差の自乗和

$$\begin{aligned} Q &= \sum_{j=1}^p \{y_j - (a_0 + \sum_{i=1}^n a_i x_{ij})\}^2 \\ &= \sum_{j=1}^p \{y_j^2 - 2y_j(a_0 + \sum_{i=1}^n a_i x_{ij}) + (a_0 + \sum_{i=1}^n a_i x_{ij})^2\} \end{aligned} \quad \text{--- (13)}$$

が最小になることである。よって、それぞれの a_i でQを偏微分して0になる場合を考えればよい。

$$\left\{ \begin{array}{l} \frac{\partial Q}{\partial a_0} = -2 \sum_{j=1}^p \left\{ y_j - (a_0 + a_1 x_{1j} + a_2 x_{2j} + \dots + a_n x_{nj}) \right\} = 0 \\ \frac{\partial Q}{\partial a_1} = -2 \sum_{j=1}^p x_{1j} \left\{ y_j - (a_0 + a_1 x_{1j} + a_2 x_{2j} + \dots + a_n x_{nj}) \right\} = 0 \\ \dots \\ \frac{\partial Q}{\partial a_n} = -2 \sum_{j=1}^p x_{nj} \left\{ y_j - (a_0 + a_1 x_{1j} + a_2 x_{2j} + \dots + a_n x_{nj}) \right\} = 0 \end{array} \right. \quad \text{--- (14)}$$

これを整理すると,

$$\left\{ \begin{array}{l} a_0 p + a_1 \sum_{j=1}^p x_{1j} + a_2 \sum_{j=1}^p x_{2j} + \dots + a_n \sum_{j=1}^p x_{nj} = \sum_{j=1}^p y_j \\ a_0 \sum_{j=1}^p x_{1j} + a_1 \sum_{j=1}^p x_{1j}^2 + a_2 \sum_{j=1}^p x_{1j} x_{2j} + \dots + a_n \sum_{j=1}^p x_{1j} x_{nj} = \sum_{j=1}^p x_{1j} y_j \\ \dots \\ a_0 \sum_{j=1}^p x_{nj} + a_1 \sum_{j=1}^p x_{1j} x_{nj} + a_2 \sum_{j=1}^p x_{1j} x_{nj} + \dots + a_n \sum_{j=1}^p x_{nj}^2 = \sum_{j=1}^p x_{nj} y_j \end{array} \right. \quad \text{--- (15)}$$

となる. 第一式を $a_0 = \bar{y} - a_1 \bar{x}_1 - \dots - a_n \bar{x}_n$ としたものを第二式以降に代入し,

$$\begin{aligned} S_{ik} &= \sum_{j=1}^p (x_{ij} - \bar{x}_i)^2 & (i = k = 1, 2, \dots, n) \\ S_{ik} &= \sum_{j=1}^p (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k) & (i \neq k, i = 1, 2, \dots, n, k = 1, 2, \dots, n) \\ S_{iy} &= \sum_{j=1}^p (x_{ij} - \bar{x}_i)(y_j - \bar{y}) & (i = 1, 2, \dots, n) \end{aligned} \quad \text{--- (16)}$$

とすると, $\partial Q / \partial a_i = 0$ の連立方程式は以下のようになる.

$$\left\{ \begin{array}{l} S_{11}a_1 + S_{12}a_2 + \dots + S_{1n}a_n = S_{1y} \\ S_{21}a_1 + S_{22}a_2 + \dots + S_{2n}a_n = S_{2y} \\ \dots \\ S_{i1}a_1 + S_{i2}a_2 + \dots + S_{in}a_n = S_{iy} \\ \dots \\ S_{n1}a_1 + S_{n2}a_2 + \dots + S_{nn}a_n = S_{ny} \end{array} \right. \quad \text{--- (17)}$$

これを行列で表わすと以下のようになる.

$$\begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1k} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2k} & \cdots & S_{2n} \\ \vdots & \vdots & \ddots & & & \vdots \\ S_{i1} & S_{i2} & & S_{ik} & & S_{in} \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nk} & \cdots & S_{nn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} S_{1y} \\ S_{2y} \\ \vdots \\ S_{iy} \\ \vdots \\ S_{ny} \end{pmatrix} \quad \text{--- (18)}$$

よって、データ x_{ij} , y_j から S_{ik} , S_{iy} を計算し、(17)式の連立方程式を解けば、求めたい(12)式の係数 $a_1 \sim a_n$ が計算できる。

a_0 は $a_0 = \bar{y} - a_1 \bar{x}_1 - a_2 \bar{x}_2 - \cdots - a_n \bar{x}_n$ から計算できる。

!!!! 注意!!!!

資料と異なる文字を変数に使用すると必ず混乱します。同じ文字を使いましょう。

練習問題 3

以下のようなデータがある。これを(12)式でフィッティングせよ。答えが正しいかを確認する為に、 $Z_j = a_0 + a_1 x_{1j} + a_2 x_{2j} + a_3 x_{3j} + a_4 x_{4j}$ を算出し、gnuplot で Z_j を横軸、 y_j を縦軸とした図を描き、直線性を確認せよ。このケースではデータの個数は $p=6$ 、説明変数の個数は $n=4$ であることに注意。

y	x1	x2	x3	x4
-13.5D0	1.2D0	3.2D0	2.2D0	4.D0
10.3D0	3.5D0	4.D0	3.3D0	3.D0
-64.5D0	-5.D0	6.D0	5.5D0	6.D0
44.2D0	4.3D0	-3.D0	6.D0	2.D0
10.D0	2.D0	3.D0	2.D0	1.D0
4.8D0	5.5D0	4.3D0	5.D0	7.D0

上記のデータは下記からダウンロードできる。

<http://www.eps.nagoya-u.ac.jp/~morota/20161215/data03.html>

3. 宿題

練習問題 3 のプログラムと結果を下記アドレスに送ること。

宿題のしめきり：1月15日（火曜日）

4. ログアウト

作業が終了したら必ずログアウトしてください。

- ・ログアウト

「 マーク」 => 「…をログアウト」

5. 追加の宿題

次の問題の答え ((3)式の係数) とプログラムを送ること。

宿題のしめきり：1月15日（火曜日）

下記のような月の岩石試料があったとする。月の表面を構成している鉱物は主に斜長石（密度~2700 kg/m³），輝石（~3300 kg/m³），かんらん石（~3300 kg/m³），チタン鉄鉱（~4700 kg/m³）である。FeOは主に、輝石、かんらん石、チタン鉄鉱に、TiO₂は主にチタン鉄鉱に多く含まれており、それらの鉱物は相対的に密度が高い。このことから、月の岩石の密度とFeO量、TiO₂量の間には関係性が見出せるように思われる。目的変数（ y_j ）を密度、説明変数（ x ）をFeO量とTiO₂量として下記のデータに(3)式でフィッティングすることによって経験的な関係式を導出せよ。答えが正しいかを確認する為に、 $Z_j = a_0 + a_1 x_{1j} + a_2 x_{2j} + a_3 x_{3j} + a_4 x_{4j}$ を算出し、gnuplotで Z_j を横軸、 y_j を縦軸とした図を描き、直線性を確認せよ。

密度[kg/m3]	FeO量 [wt%]	TiO2量 [wt%]
2791.0	0.33	0.02
2840.6	2.62	0.13
2915.4	3.67	0.05
2987.8	5.48	0.18
2921.5	4.99	0.05
2957.2	6.40	0.25
2789.6	1.87	0.10
3004.2	11.20	2.00
3072.3	10.40	1.50
3303.3	17.20	4.60
2860.4	3.10	0.26
3230.3	14.00	0.30
3181.0	14.10	2.60
3261.1	15.50	1.70
3091.3	10.00	2.20
3391.5	18.90	12.20
3435.1	19.65	11.78
3322.5	16.51	9.61
3394.1	19.77	3.32

3388.4	20.23	1.80
3398.8	20.19	1.78
3358.7	21.66	2.60

上記のデータは

<http://www.eps.nagoya-u.ac.jp/~morota/20161215/data04.html>
からダウンロードできる。