第１１回　データ解析
（連立方程式，逆行列，多変量解析１）

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 20141211
% cd 20141211
```

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.

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

上記の連立方程式は以下のように表わすことができる.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \cdots \\ b_n \end{pmatrix} \tag{2}$$

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.

$$\begin{pmatrix} 1 & a_{12}/a_{11} & a_{13}/a_{11} & \cdots & a_{1n}/a_{11} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1/a_{11} \\ b_2 \\ b_3 \\ \cdots \\ b_n \end{pmatrix} \tag{3}$$

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．

$$\begin{pmatrix} 1 & a_{12}/a_{11} & a_{13}/a_{11} & \cdots & a_{1n}/a_{11} \\ 0 & a_{22}-a_{21}\,a_{12}/a_{11} & a_{23}-a_{21}\,a_{13}/a_{11} & \cdots & a_{2n}-a_{21}\,a_{1n}/a_{11} \\ 0 & a_{32}-a_{31}\,a_{12}/a_{11} & a_{33}-a_{31}\,a_{13}/a_{11} & \cdots & a_{3n}-a_{31}\,a_{1n}/a_{11} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & a_{n2}-a_{n1}\,a_{12}/a_{11} & a_{n3}-a_{n1}\,a_{13}/a_{11} & \cdots & a_{nn}-a_{n1}\,a_{1n}/a_{11} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1/a_{11} \\ b_2-a_{21}\,b_1/a_{11} \\ b_3-a_{31}\,b_1/a_{11} \\ \cdots \\ b_n-a_{n1}\,b_1/a_{11} \end{pmatrix} \tag{4}$$

第２列〜第ｎ列に関しても同様な操作を繰り返し

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_n \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \\ B_3 \\ \cdots \\ B_n \end{pmatrix} \tag{5}$$

となったときの上辺が解である．

＜注意点＞

この方法では消去の各段階で$a_{11}, a_{22}, \cdots, a_{nn}$（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．

```
      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．

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

サブルーチンは

http://www.eps.nagoya-u.ac.jp/~morota/NumericalAnalysis/example01.f

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．

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

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，

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

となり，行列 B を分けて表わすと，

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

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

$$\cdots$$

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

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，

$$
\begin{pmatrix}
a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & 1 & 0 & 0 & \cdots & 0 \\
a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & 0 & 1 & 0 & \cdots & 0 \\
a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & 0 & 0 & 1 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} & 0 & 0 & 0 & \cdots & 1
\end{pmatrix} \tag{9}
$$

を掃き出し法により以下のかたちに変換したとき

$$
\begin{pmatrix}
1 & 0 & 0 & \cdots & 0 & A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\
0 & 1 & 0 & \cdots & 0 & A_{21} & A_{22} & A_{23} & \cdots & A_{2n} \\
0 & 0 & 1 & \cdots & 0 & A_{31} & A_{32} & A_{33} & \cdots & A_{3n} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & 1 & A_{n1} & A_{n2} & A_{n3} & \cdots & A_{nn}
\end{pmatrix} \tag{10}
$$

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．

$$
A = \begin{pmatrix} -1 & 2 & 1 \\ 2 & -1 & -2 \\ 3 & 5 & -1 \end{pmatrix} \tag{11}
$$

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 x（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 x（テスト直前の勉強時間）+ a_2 x（講義の出席回数）
+ a_3 x（課題の提出回数）+ a_4 x（IQ）＋ …

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）を複数の説明変数で説明・予測できると仮定し，a_0〜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．

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = a_0 + \sum_{i=1}^{n} a_i x_i \quad (3)$$

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

$$\begin{aligned} Q &= \sum_{j=1}^{p} \{ y_j - (a_0 + \sum_{i=1}^{n} a_i x_{ij}) \}^2 \\ &= \sum_{j=1}^{p} \{ y_j^2 - 2 y_j (a_0 + \sum_{i=1}^{n} a_i x_{ij}) + (a_0 + \sum_{i=1}^{n} a_i x_{ij})^2 \} \end{aligned} \quad (4)$$

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．

$$\begin{cases} \dfrac{\partial Q}{\partial a_0} = -2\sum_{j=1}^{p}\left\{y_j - \left(a_0 + a_1 x_{1j} + \ldots + a_n x_{nj}\right)\right\} = 0 \\ \dfrac{\partial Q}{\partial a_1} = -2\sum_{j=1}^{p} x_{1j}\left\{y_j - \left(a_0 + a_1 x_{1j} + \ldots + a_n x_{nj}\right)\right\} = 0 \\ \qquad \ldots \\ \dfrac{\partial Q}{\partial a_n} = -2\sum_{j=1}^{p} x_{nj}\left\{y_j - \left(a_0 + a_1 x_{1j} + \ldots + a_n x_{nj}\right)\right\} = 0 \end{cases} \tag{5}$$

これを整理すると，

$$\begin{cases} a_0 m + a_1\sum_{j=1}^{p} x_{1j} + \cdots + a_n \sum_{j=1}^{p} x_{nj} = \sum_{j=1}^{p} y_j \\ a_0 \sum_{j=1}^{p} x_{1j} + a_1 \sum_{j=1}^{p} x_{1j}^2 + \cdots + a_n \sum_{j=1}^{p} x_{1j} x_{nj} = \sum_{j=1}^{p} x_{1j} y_j \\ \qquad \ldots \\ a_0 \sum_{j=1}^{p} x_{nj} + a_1 \sum_{j=1}^{p} x_{1j} x_{nj} + \cdots + a_n \sum_{j=1}^{p} x_{nj}^2 = \sum_{j=1}^{p} x_{nj} y_j \end{cases} \tag{6}$$

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

$$\begin{aligned} S_{ik} &= \sum_{j=1}^{p}(x_{ij} - \overline{x}_i)^2 && (i = k = 1,2,\ldots,n) \\ S_{ik} &= \sum_{j=1}^{p}(x_{ij} - \overline{x}_i)(x_{kj} - \overline{x}_k) && (i \neq k, i = 1,2,\ldots,n, k = 1,2,\ldots,n) \\ S_{iy} &= \sum_{j=1}^{p}(x_{ij} - \overline{x}_i)(y_j - \overline{y}) && (i = 1,2,\ldots,n) \end{aligned} \tag{7}$$

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．

$$\begin{cases} S_{11}a_1 + S_{12}a_2 + \cdots + S_{1n}a_n = S_{1y} \\ S_{21}a_1 + S_{22}a_2 + \cdots + S_{2n}a_n = S_{2y} \\ \qquad \cdots \\ S_{i1}a_1 + S_{i2}a_2 + \cdots + S_{in}a_n = S_{iy} \\ \qquad \cdots \\ S_{n1}a_1 + S_{n2}a_2 + \cdots + S_{nn}a_n = S_{ny} \end{cases} \tag{8}$$

これを行列で表わすと以下のようになる．

$$
\begin{pmatrix}
S_{11} & S_{12} & \cdots & S_{1k} & \cdots & S_{1n} \\
S_{21} & S_{22} & \cdots & S_{2k} & \cdots & S_{2n} \\
\vdots & \vdots & \ddots & & & \vdots \\
S_{i1} & S_{i2} & & S_{ik} & & S_{in} \\
\vdots & \vdots & & & \ddots & \vdots \\
S_{n1} & S_{n2} & \cdots & S_{nk} & \cdots & S_{nn}
\end{pmatrix}
\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix}
=
\begin{pmatrix} S_{1y} \\ S_{2y} \\ \vdots \\ S_{iy} \\ \vdots \\ S_{ny} \end{pmatrix}
\quad (9)
$$

よって，データ x_{ij}， y_j から S_{ik}， S_{iy} を計算し，(7)式の連立方程式を解けば，求めたい(1)式の係数 $a_1 \sim a_n$ が計算できる.

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

<u>練習問題３</u>

以下のようなデータがある．これを(3)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， $Z_j = a_1x_{1j} + a_2x_{2j} + a_3x_{3j} + a_4x_{4j}$ 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.

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

http://www.eps.nagoya-u.ac.jp/~morota/NumericalAnalysis/data03.dat

からダウンロードできる.

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.

宿題の提出先: 城野 (sirono@eps.nagoya-u.ac.jp)

諸田 (morota@eps.nagoya-u.ac.jp)

野上 (nogami.tatsuhiko@g.mbox.nagoya-u.ac.jp)

宿題のしめきり： 12 月 16 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」