1

重回帰分析入門

京都大学　加納 学

Division of Process Control & Process Systems Engineering
Department of Chemical Engineering, Kyoto University

manabu@cheme.kyoto-u.ac.jp
http://www-pse.cheme.kyoto-u.ac.jp/~kano/

標準化

2

各変数を平均0，分散1の変数に変換する．

$$x_{nm} = \frac{x^*_{nm} - \bar{x}_m}{\sigma_m}$$

変数　m
サンプル　n

平均　$\bar{x}_m = \frac{1}{N}\sum_{n=1}^{N} x^*_{nm}$

分散　$\sigma_m^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x^*_{nm} - \bar{x}_m)^2$

準備

3

入力変数　　出力変数

測定データ

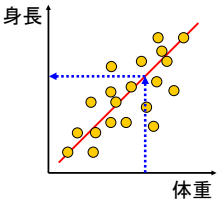
$$X^* = \begin{bmatrix} x^*_{11} & x^*_{12} & \cdots & x^*_{1M} \\ x^*_{21} & x^*_{22} & \cdots & x^*_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ x^*_{N1} & x^*_{N2} & \cdots & x^*_{NM} \end{bmatrix} \quad Y^* = \begin{bmatrix} y^*_{11} \\ y^*_{21} \\ \vdots \\ y^*_{N1} \end{bmatrix}$$

標準化データ

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix} \quad Y = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{N1} \end{bmatrix}$$

最小二乗法（OLS）

4



体重から身長を推定したい！

身長 = 定数 × 体重 + 誤差
y　a　x　e

$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - ax_i)^2$　を最小にする係数 a を求める．

重回帰分析： 設定

5

いま，出力変数 y の推定値を入力変数の線形結合

$$\hat{y} = \sum_{m=1}^{M} a_m x_m$$

で与えたい．このとき，出力変数の測定値と推定値との差

$$e = y - \hat{y} = y - \sum_{m=1}^{M} a_m x_m$$

の二乗和が最小となるように，偏回帰係数 a_m を決定せよ．

重回帰分析： 必要条件

6

誤差の二乗和

$$J = \sum_{n=1}^{N} e^2 = \sum_{n=1}^{N}(y_n - \hat{y}_n)^2$$
$$= (Y - Xa)^T (Y - Xa)$$

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix}$$

必要条件（Jが最小となるための）

$$\frac{\partial J}{\partial a} = 0$$

重回帰分析： 正規方程式

7

$$\frac{\partial J}{\partial a}=\begin{bmatrix}\frac{\partial J}{\partial a_1}\\ \frac{\partial J}{\partial a_2}\\ \vdots\\ \frac{\partial J}{\partial a_M}\end{bmatrix}=2(X^TXa-X^TY)=0$$

正規方程式　$X^TXa-X^TY=0$

偏回帰係数　$a=(X^TX)^{-1}X^TY$

重回帰分析： 重回帰式

8

$$\hat{y}=\sum_{m=1}^{M}a_m x_m$$

a_m　標準偏回帰係数

$\frac{a_m\sigma_y}{\sigma_m}$　偏回帰係数

$$\frac{\hat{y}^*-\bar{y}}{\sigma_y}=\sum_{m=1}^{M}a_m\frac{x_m^*-\bar{x}_m}{\sigma_m}$$

$$\hat{y}^*=\sum_{m=1}^{M}\frac{a_m\sigma_y}{\sigma_m}x_m^*+\left(\bar{y}-\sum_{m=1}^{M}\frac{a_m\sigma_y}{\sigma_m}\bar{x}_m\right)$$

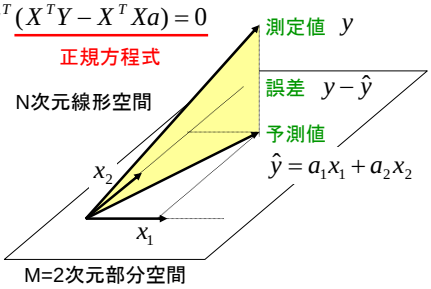
重回帰分析： 幾何学的解釈

9

誤差が最小となるためには，誤差と予測値が直交すればよい．

$$\langle\hat{y},y-\hat{y}\rangle=\langle Xa,Y-Xa\rangle$$
$$=a^T(X^TY-X^TXa)=0$$

正規方程式



重回帰分析： 相関係数

10

誤差が最小となるためには，誤差と予測値が直交すればよい．

誤差が最小となるためには，
測定値と予測値がなす角 θ が最小になればよい．

誤差が最小となるためには，
測定値と予測値の相関係数が最大になればよい．

重相関係数　$r_{y\hat{y}}=\frac{\sigma_{y\hat{y}}^2}{\sigma_y\sigma_{\hat{y}}}=\frac{y^T\hat{y}}{\|y\|\|\hat{y}\|}=\cos\theta$

重回帰分析： 重相関係数

11

重相関係数　$R\equiv r_{y\hat{y}}=\frac{\sigma_{y\hat{y}}^2}{\sigma_y\sigma_{\hat{y}}}=\frac{y^T\hat{y}}{\|y\|\|\hat{y}\|}=\cos\theta$

$$R^2=\frac{(\sigma_{y\hat{y}}^2)^2}{\sigma_y^2\sigma_{\hat{y}}^2}=\frac{\left[(y-\bar{y})^T(\hat{y}-\bar{y})\right]^2}{\|y-\bar{y}\|^2\|\hat{y}-\bar{y}\|^2}=\frac{\|\hat{y}-\bar{y}\|^2}{\|y-\bar{y}\|^2}$$

$$(y-\bar{y})^T(\hat{y}-\bar{y})=\left[(y-\hat{y})-(\hat{y}-\bar{y})\right]^T(\hat{y}-\bar{y})$$
$$=(y-\hat{y})^T(\hat{y}-\bar{y})-(\hat{y}-\bar{y})^T(\hat{y}-\bar{y})$$
$$=\|\hat{y}-\bar{y}\|^2$$

例題： 重回帰分析（10人）

12

	身長(y)	胸囲(x1)	体重(x2)
1	167.0	84.0	61.0
2	167.5	87.0	55.5
3	168.4	86.0	57.0
4	172.0	85.0	57.0
5	155.3	82.0	50.0
6	151.4	87.0	50.0
7	163.0	92.0	66.5
8	174.0	94.0	65.0
9	168.0	88.0	60.5
10	160.4	84.9	49.5

例題： 重回帰分析（10人） 13

	身長(y)	胸囲(x1)	体重(x2)
平均	164.7	87.0	57.2
標準偏差	7.18	3.63	6.13
偏回帰係数	–	-0.427	0.969
標準偏回帰係数	–	-0.216	0.828
重相関係数（R）	0.687	–	–
決定係数（R2）	0.472	–	–

分散分析 14

変動要因	平方和	自由度	不偏分散	分散比
全変動	SS_y	$N-1$	–	–
回帰による変動	SS_r	p	$V_r = \frac{SS_r}{p}$	$F = \frac{V_r}{V_e}$
残差の変動	SS_e	$N-p-1$	$V_e = \frac{SS_e}{N-p-1}$	

ある仮定（重回帰分析の前提条件）の下で，
F は自由度 p, N-p-1 の F 分布に従う．

重要な式 15

$$y^* - \bar{y} = \sum_{i=1}^{p} a_i (x_i^* - \bar{x}_i)$$

$$F = \frac{V_r}{V_e} = \frac{R^2 / p}{(1-R^2)/(N-p-1)}$$

$$SS_y = \sum_{i=1}^{N} (y_i^* - \bar{y})^2$$

$$SS_r = \sum_{i=1}^{N} (\hat{y}_i^* - \bar{y})^2$$

$$SS_e = \sum_{i=1}^{N} (y_i^* - \hat{y}_i)^2$$

$$SS_y = SS_r + SS_e$$

例題： 分散分析（10人） 16

変動要因	平方和	自由度	不偏分散	分散比
全変動	464.1	9	–	–
回帰による変動	219.0	2	109.5	3.13
残差の変動	245.1	7	35.0	

$F(p, N-p-1; \alpha)$ 自由度 p, N-p-1 の F 分布，危険率α

$F(2,7;0.05) = 4.737 > 3.13$

例題： 重回帰分析（20人） 17

	身長(y)	胸囲(x1)	体重(x2)
1	167.0	84.0	61.0
2	167.5	87.0	55.5
3	168.4	86.0	57.0
4	172.0	85.0	57.0
5	155.3	82.0	50.0
6	151.4	87.0	50.0
7	163.0	92.0	66.5
8	174.0	94.0	65.0
9	168.0	88.0	60.5
10	160.4	84.9	49.5

	身長(y)	胸囲(x1)	体重(x2)
11	164.7	78.0	49.5
12	171.0	90.0	61.0
13	162.6	88.0	59.5
14	164.8	87.0	58.4
15	163.3	82.0	53.5
16	167.6	84.0	54.0
17	169.2	86.0	60.0
18	168.0	83.0	58.8
19	167.4	85.2	54.0
20	172.0	82.0	56.0

例題： 重回帰分析（20人） 18

	身長(y)	胸囲(x1)	体重(x2)
平均	165.9	85.8	56.8
標準偏差	5.54	3.68	4.93
偏回帰係数	–	-0.663	0.986
標準偏回帰係数	–	-0.441	0.879
重相関係数（R）	0.636	–	–
決定係数（R^2）	0.405	–	–

例題：分散分析（20人）

19

変動要因	平方和	自由度	不偏分散	分散比
全変動	582.5	19	—	—
回帰による変動	235.7	2	117.9	5.78
残差の変動	346.7	17	20.4	

$F(p, N-p-1; \alpha)$　自由度 p, N-p-1 の F 分布，危険率α

$F(2,17;0.05) = 3.592 < 5.78$

重回帰分析の問題点

20

偏回帰係数　$a = (X^T X)^{-1} X^T Y$

$X^T X$ が逆行列を持たない場合，最小二乗法は使えない．

入力変数が線形従属である場合

サンプル数が入力変数の数より少ない場合もダメ．
以下では，サンプル数は十分にあるとする．

多重共線性

21

	Data “A”			Data “B”		
y	x1	x2	x3	x1	x2	x3
241	15.9	34.6	64.8	16.1	34.7	65.1
321	37.0	16.1	72.1	36.9	16.3	72.0
82	61.1	83.0	28.6	60.6	82.8	28.9
156	86.0	65.9	33.9	85.9	65.9	34.2
係数	1.36	-0.80	5.01	-4.28	-18.9	-26.0

入力変数が厳密に線形従属でなくても，入力変数間に強い相関関係が存在する場合には，係数推定値の分散が大きくなり，推定結果の信頼性が低下してしまう．

何が問題なのか？

22

推定値の分散が大きくなると，何が問題なのか？
推定ができれば良いのではないか？

＜重回帰分析で酷い目に遭う例＞

$y = a_1 x_1 + a_2 x_2$　　$y = x_1 = x_2$

測定データ　$y = 1.00,\ x_1 = 1.01,\ x_2 = 0.99$

Model 1	$\hat{y} = x_2$	0.99
Model 2	$\hat{y} = 0.5x_1 + 0.5x_2$	1.00
Model 3	$\hat{y} = 100x_1 - 99x_2$	2.99

係数が大きいほど，測定ノイズの影響を受けやすい．

最小二乗法の拡張

23

Ordinary Least Squares (OLS)

$a = (X^T X)^{-1} X^T Y$　　$\min \|Y - Xa\|^2$

Minimum Norm Solution

$a = X^+ Y$　　X^+ ：一般化逆行列

Ridge Regression (RR)

$a = (X^T X + \lambda I)^{-1} X^T Y$　　$\min \|Y - Xa\|^2 + \lambda \|a\|^2$

Principal Component Regression (PCR)
Partial Least Squares (PLS)

いずれの手法も係数を小さく抑えようとする．

リッジ回帰

24

重回帰　$\min \|Y - Xa\|^2$

リッジ回帰　$\min \|Y - Xa\|^2 + \lambda \|a\|^2$

必要条件（評価が最小となるための）

$\frac{\partial J}{\partial a} = 2(X^T X a - X^T Y + \lambda a) = 0$

$a = (X^T X + \lambda I)^{-1} X^T Y$

例題： リッジ回帰

25

		Data Set: A			Data Set: B		
	y	x1	x2	x3	x1	x2	x3
1	241	15.9	34.6	64.8	16.1	34.7	65.1
2	321	37.0	16.1	72.1	36.9	16.3	72.0
3	82	61.1	83.0	28.6	60.6	82.8	28.9
4	156	86.0	65.9	33.9	85.9	65.9	34.2
偏回帰係数		—	—	—	—	—	—
重回帰		1.36	-0.80	5.01	-4.28	-18.9	-26.0
リッジ回帰		0.86	-2.34	2.36	0.87	-2.38	2.34

おわり

26