

応用システム工学

第五回 回帰分析

平成24年06月15日
重回帰分析

2012/06/15

1

線形重回帰

- 1つの目的変数 y に対する複数(p 個)の説明変数 x_1, x_2, \dots, x_p

– 線形重回帰モデル

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi} + e_i \quad (i = 1, 2, \dots, n)$$

	目的変数:y	説明変数:x	予測誤差
1	y_1	$x_{11}, x_{21}, \dots, x_{p1}$	$e_1 = y_1 - (a_0 + a_1 x_{11} + a_2 x_{21} + \dots + a_p x_{p1})$
2	y_2	$x_{12}, x_{22}, \dots, x_{p2}$	$e_2 = y_2 - (a_0 + a_1 x_{12} + a_2 x_{22} + \dots + a_p x_{p2})$
...			
i	y_i	$x_{1i}, x_{2i}, \dots, x_{pi}$	$e_i = y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})$
...			
n	y_n	$x_{1n}, x_{2n}, \dots, x_{pn}$	$e_n = y_n - (a_0 + a_1 x_{1n} + a_2 x_{2n} + \dots + a_p x_{pn})$

2012/06/15

2

線形重回帰

- 単回帰 (x,y)平面
- 重回帰 (x₁,x₂,...,x_p,y) (p+1)次元空間
 - P次元超平面

$$y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_px_p$$

– 説明変数x_{1i},x_{2i},...,x_{pi}から目的変数y_iを予測

- 予測誤差

$$e_i = y_i - (a_0 + a_1x_{1i} + a_2x_{2i} + \cdots + a_px_{pi})$$

- 予測誤差の平方和

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left\{ y_i - (a_0 + a_1x_{1i} + a_2x_{2i} + \cdots + a_px_{pi}) \right\}^2$$

2012/06/15=1

3

線形重回帰

- 予測誤差の平方和の最小化

$$\begin{aligned} F(a_0, a_1, \dots, a_p) &= \sum_{i=1}^n e_i^2 \\ &= \sum_{i=1}^n \left\{ y_i - (a_0 + a_1x_{1i} + a_2x_{2i} + \cdots + a_px_{pi}) \right\}^2 \end{aligned}$$

$$\left\{ \begin{array}{l} \frac{\partial}{\partial a_0} F(a_0, a_1, \dots, a_p) = 0 \\ \frac{\partial}{\partial a_1} F(a_0, a_1, \dots, a_p) = 0 \\ \vdots \\ \frac{\partial}{\partial a_p} F(a_0, a_1, \dots, a_p) = 0 \end{array} \right.$$

2012/06/15

4

線形重回帰

$$\frac{\partial}{\partial a_0} F(a_0, a_1, \dots, a_p) = \frac{\partial}{\partial a_0} \sum_{i=1}^n \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\}^2$$

$$= -2 \sum_{i=1}^n \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\} = 0$$

$$\frac{\partial}{\partial a_1} F(a_0, a_1, \dots, a_p) = \frac{\partial}{\partial a_1} \sum_{i=1}^n \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\}^2$$

$$= -2 \sum_{i=1}^n x_{1i} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\} = 0$$

$$\vdots \quad \frac{\partial}{\partial a_p} F(a_0, a_1, \dots, a_p) = \frac{\partial}{\partial a_p} \sum_{i=1}^n \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\}^2$$

$$= -2 \sum_{i=1}^n x_{pi} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\} = 0$$

2012/06/15

5

線形重回帰

$$\left\{ \begin{array}{l} \sum_{i=1}^n \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\} = 0 \\ \sum_{i=1}^n x_{1i} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\} = 0 \\ \vdots \\ \sum_{i=1}^n x_{pi} \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\} = 0 \end{array} \right.$$

2012/06/15

6

線形重回帰

- 正規方程式: 残差二乗和を最小にする推定値を与える方程式 → a_0, a_1, \dots, a_p の連立方程式

$$\left\{ \begin{array}{l} a_0 n + a_1 \sum_{i=1}^n x_{1i} + a_2 \sum_{i=1}^n x_{2i} + \dots + a_p \sum_{i=1}^n x_{pi} = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n x_{1i} + a_1 \sum_{i=1}^n x_{1i}^2 + a_2 \sum_{i=1}^n x_{1i} x_{2i} + \dots + a_p \sum_{i=1}^n x_{1i} x_{pi} = \sum_{i=1}^n x_{1i} y_i \\ \vdots \\ a_0 \sum_{i=1}^n x_{pi} + a_1 \sum_{i=1}^n x_{pi} x_{1i} + a_2 \sum_{i=1}^n x_{pi} x_{2i} + \dots + a_p \sum_{i=1}^n x_{pi} x_{pi} = \sum_{i=1}^n x_{pi} y_i \end{array} \right.$$

2012/06/15

7

線形重回帰

$$\begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \dots & \sum_{i=1}^n x_{pi} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i} x_{2i} & & \sum_{i=1}^n x_{1i} x_{pi} \\ \vdots & & & & \vdots \\ \sum_{i=1}^n x_{pi} & \sum_{i=1}^n x_{pi} x_{1i} & \sum_{i=1}^n x_{pi} x_{2i} & & \sum_{i=1}^n x_{pi} x_{pi} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1i} y_i \\ \vdots \\ \sum_{i=1}^n x_{pi} y_i \end{bmatrix}$$

2012/06/15

8

線形重回帰

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \cdots & \sum_{i=1}^n x_{pi} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i}x_{2i} & \cdots & \sum_{i=1}^n x_{1i}x_{pi} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{pi} & \sum_{i=1}^n x_{pi}x_{1i} & \sum_{i=1}^n x_{pi}x_{2i} & \cdots & \sum_{i=1}^n x_{pi}x_{pi} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1i}y_i \\ \vdots \\ \sum_{i=1}^n x_{pi}y_i \end{bmatrix}$$

$$a_0, a_1, \dots, a_p \Rightarrow \hat{a}_0, \hat{a}_1, \dots, \hat{a}_p$$

分散共分散で表すことを考える

2012/06/15

9

線形重回帰

- x_1, \dots, x_p の分散共分散行列

$$V = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1l} & \cdots & s_{1p} \\ s_{21} & s_{22} & & & & s_{2p} \\ \vdots & & & & & \vdots \\ s_{j1} & s_{j2} & \cdots & s_{jl} & \cdots & s_{jp} \\ \vdots & & & & & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pl} & \cdots & s_{pp} \end{bmatrix} \quad s_{jl} = \frac{1}{n} \sum_{i=1}^n (x_{ji} - \bar{x}_j)(x_{li} - \bar{x}_l)$$

$$j, l = 1, 2, \dots, p$$

- y と x_1, \dots, x_p の共分散

$$s_{yj} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(x_{ji} - \bar{x}_j) \quad j = 1, 2, \dots, p$$

2012/06/15

10

線形重回帰

分散共分散行列であらわす

- 正規方程式の一行目

$$\hat{a}_0 n + \hat{a}_1 \sum_{i=1}^n x_{1i} + \hat{a}_2 \sum_{i=1}^n x_{2i} + \cdots + \hat{a}_p \sum_{i=1}^n x_{pi} = \sum_{i=1}^n y_i$$

$$\hat{a}_0 = \frac{1}{n} \sum_{i=1}^n y_i - \frac{1}{n} \left(\hat{a}_1 \sum_{i=1}^n x_{1i} + \hat{a}_2 \sum_{i=1}^n x_{2i} + \cdots + \hat{a}_p \sum_{i=1}^n x_{pi} \right)$$

$$= \bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p)$$

$$\hat{a}_0 = \bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p)$$

– 正規方程式の a_0 に代入

線形重回帰

$$\left\{ \begin{array}{l} \sum_{i=1}^n \{y_i - [\bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p)] + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \cdots + \hat{a}_p x_{pi}\} = 0 \\ \sum_{i=1}^n x_{1i} \{y_i - [\bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p)] + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \cdots + \hat{a}_p x_{pi}\} = 0 \\ \vdots \\ \sum_{i=1}^n x_{ji} \{y_i - [\bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p)] + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \cdots + \hat{a}_p x_{pi}\} = 0 \\ \vdots \\ \sum_{i=1}^n x_{pi} \{y_i - [\bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p)] + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \cdots + \hat{a}_p x_{pi}\} = 0 \end{array} \right.$$

線形重回帰

$$\left\{ \begin{array}{l} \sum_{i=1}^n \{(y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \dots - \hat{a}_p(x_{pi} - \bar{x}_p)\} = 0 \\ \sum_{i=1}^n x_{1i} \{(y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \dots - \hat{a}_p(x_{pi} - \bar{x}_p)\} = 0 \\ \vdots \\ \sum_{i=1}^n x_{ji} \{(y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \dots - \hat{a}_p(x_{pi} - \bar{x}_p)\} = 0 \\ \vdots \\ \sum_{i=1}^n x_{pi} \{(y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \dots - \hat{a}_p(x_{pi} - \bar{x}_p)\} = 0 \end{array} \right.$$

各行j=1~pから、j=0行× \bar{x}_j を引く

共通

2012/06/15

13

線形重回帰

$$\left\{ \begin{array}{l} \sum_{i=1}^n (x_{1i} - \bar{x}_1) \{(y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \dots - \hat{a}_p(x_{pi} - \bar{x}_p)\} = 0 \\ \vdots \\ \sum_{i=1}^n (x_{ji} - \bar{x}_j) \{(y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \dots - \hat{a}_p(x_{pi} - \bar{x}_p)\} = 0 \\ \vdots \\ \sum_{i=1}^n (x_{pi} - \bar{x}_p) \{(y_i - \bar{y}) - \hat{a}_1(x_{1i} - \bar{x}_1) - \hat{a}_2(x_{2i} - \bar{x}_2) - \dots - \hat{a}_p(x_{pi} - \bar{x}_p)\} = 0 \end{array} \right.$$

各行の両辺をnで割ると

2012/06/15

14

線形重回帰

$$\begin{cases} s_{y1} - \hat{a}_1 s_{11} - \hat{a}_2 s_{12} - \cdots - \hat{a}_p s_{1p} = 0 \\ \vdots \\ s_{yj} - \hat{a}_1 s_{j1} - \hat{a}_2 s_{j2} - \cdots - \hat{a}_p s_{jp} = 0 \\ \vdots \\ s_{yp} - \hat{a}_1 s_{p1} - \hat{a}_2 s_{p2} - \cdots - \hat{a}_p s_{pp} = 0 \end{cases}$$

$$\begin{cases} \hat{a}_1 s_{11} + \hat{a}_2 s_{12} + \cdots + \hat{a}_p s_{1p} = s_{y1} \\ \vdots \\ \hat{a}_1 s_{j1} + \hat{a}_2 s_{j2} + \cdots + \hat{a}_p s_{jp} = s_{yj} \\ \vdots \\ \hat{a}_1 s_{p1} + \hat{a}_2 s_{p2} + \cdots + \hat{a}_p s_{pp} = s_{yp} \end{cases}$$

2012/06/15

15

線形重回帰

$$\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ \vdots & & & \\ s_{j1} & s_{j2} & & s_{jp} \\ \vdots & & & \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} s_{y1} \\ \vdots \\ s_{yj} \\ \vdots \\ s_{yp} \end{bmatrix}$$

分散共分散行列V yのxに対する共分散

別途 $\hat{a}_0 = \bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \cdots + \hat{a}_p \bar{x}_p)$ で \hat{a}_0 を求めればよい

目的変数yの, 説明変数 x_1, x_2, \dots, x_p に対する線形重回帰式

$$y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \cdots + \hat{a}_p x_p$$

重回帰係数 $\hat{a}_j \quad j = 1, 2, \dots, p$

2012/06/15

16

重回帰式の予測誤差の標準偏差

- 予測誤差の標準偏差

$$s_e = \sqrt{\frac{1}{n-(p+1)} \sum_{i=1}^n (e_i - \bar{e})^2} = \sqrt{\frac{1}{n-(p+1)} \sum_{i=1}^n e_i^2}$$

– ただし $\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i = \frac{1}{n} \sum_{i=1}^n \{y_i - (\hat{a}_0 + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \dots + \hat{a}_p x_{pi})\}$

$$\hat{a}_0 = \bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \dots + \hat{a}_p \bar{x}_p) \quad \text{より}$$

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n \{y_i - [\bar{y} - (\hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \dots + \hat{a}_p \bar{x}_p) + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \dots + \hat{a}_p x_{pi}]\}$$

$$= \frac{1}{n} \sum_{i=1}^n \{(y_i - \bar{y}) - \hat{a}_1 (x_{1i} - \bar{x}_1) - \dots - \hat{a}_p (x_{pi} - \bar{x}_p)\} = 0$$

2012/06/15

 $\frac{\partial}{\partial a_0} F(a_0, a_1, \dots, a_p) = 0$ に一致 17

重相関係数

- 重回帰式による予測値

$$Y_i = \hat{a}_0 + \hat{a}_1 \bar{x}_1 + \hat{a}_2 \bar{x}_2 + \dots + \hat{a}_p \bar{x}_p$$

- 予測値 Y_i は説明変数で表される
- 目的変数 y と予測値 Y の単相関係数 r_{yY}
→ 目的変数 y と説明変数 x_1, x_2, \dots, x_p の重相関係数

$$r_{y.12\dots p} = \frac{s_{yY}}{\sqrt{s_{yy} s_{YY}}} = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

– 回帰平面 $(p+1)$ 次元に近いかどうかを表す

2012/06/15

18

重相関係数

- 重相関係数のとる範囲は？

$$r_{y.12\cdots p} = \frac{s_{yY}}{\sqrt{s_{yy}s_{YY}}} = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

– 分子について考える

$$s_{yY} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(Y_i - \bar{Y})$$

重相関係数

– 目的変数と予測値の平均

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n (y_i - e_i) = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}$$

– 分子

$$s_{yY} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(Y_i - \bar{Y}) = \frac{1}{n} \sum_{i=1}^n (Y_i + e_i - \bar{Y})(Y_i - \bar{Y})$$

$$= \frac{1}{n} \sum_{i=1}^n \left\{ (Y_i - \bar{Y})^2 + e_i (Y_i - \bar{Y}) \right\}$$

$$= \frac{1}{n} \sum_{i=1}^n \left\{ (Y_i - \bar{Y})^2 + e_i (a_0 + a_1 x_{1i} + \cdots + a_p x_{pi} - \bar{Y}) \right\}$$

$$= \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 = s_{YY} \geq 0 \quad \Rightarrow \quad r_{y.12\cdots p} = \frac{s_{yY}}{\sqrt{s_{yy}s_{YY}}} \geq 0$$

重相関係数

– 回帰式の残差平方和を最小にする係数の条件

$$\begin{aligned} \frac{\partial F}{\partial a_0} &= \frac{\partial}{\partial a_0} \sum_{i=1}^n \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\}^2 \\ &= -2 \sum_{i=1}^n \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\} = \sum_{i=1}^n e_i = 0 \\ \frac{\partial F}{\partial a_j} &= \frac{\partial}{\partial a_j} \sum_{i=1}^n \{y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_p x_{pi})\}^2 \\ &= -2 \sum_{i=1}^n x_{ji} \{y_i - (a_0 + a_1 x_{1i} + \dots + a_p x_{pi})\} = -2 \sum_{i=1}^n x_{ji} e_i = 0 \end{aligned}$$

2012/06/15

21

重相関係数

- 両辺二乗 $r_{y \cdot 12 \dots p}^2 = \frac{s_{yY}^2}{s_{yy} s_{YY}} = \frac{s_{YY}^2}{s_{yy} s_{YY}} = \frac{s_{YY}}{s_{yy}}$
- 目的変数・予測値の分散の関係を求める

$$\begin{aligned} s_{yy} &= \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n (Y_i + e_i - \bar{Y})^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ (Y_i - \bar{Y})^2 + 2e_i (Y_i - \bar{Y}) + e_i^2 \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ (Y_i - \bar{Y})^2 + 2e_i (a_0 + a_1 x_{1i} + \dots + a_p x_{pi} - \bar{Y}) + e_i^2 \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ (Y_i - \bar{Y})^2 + e_i^2 \right\} = s_{YY} + \frac{1}{n} \sum_{i=1}^n e_i^2 \end{aligned}$$

2012/06/15

22

重相関係数

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n e_i^2 &= s_{yy} - s_{YY} = s_{yy} \left(1 - \frac{s_{YY}}{s_{yy}} \right) \\ &= s_{yy} (1 - r_{y \cdot 12 \dots p}^2) \geq 0\end{aligned}$$

$$1 - r_{y \cdot 12 \dots p}^2 \geq 0$$

$$-1 \leq r_{y \cdot 12 \dots p} \leq 1$$

$$r_{y \cdot 12 \dots p} = \frac{s_{yY}}{\sqrt{s_{yy} s_{YY}}} \geq 0$$

よリ

$$0 \leq r_{y \cdot 12 \dots p} \leq 1$$