V-3-4. 最適化と疑似逆行列

重回帰分析では、被説明変数を説明変数で説明するために、データを使って最適な係数を求めます。ここで、被説明変数を $\boldsymbol{y}$ として説明変数を$\boldsymbol{x}$とすると、データセットは以下のようになります。

$$\mathrm{y}_1, x_{11}, x_{21} \cdots x_{p1}$$
$$\mathrm{y}_2, x_{12}, x_{22} \cdots x_{p2}$$
$$\mathrm{y}_3, x_{11}, x_{23} \cdots x_{p3}$$
$$\vdots$$
$$\mathrm{y}_m, x_{1m}, x_{2m} \cdots x_{pm}$$

私たちは次のように被説明変数を説明変数で表したいと考えているのです。

$$\mathrm{y}_1 = a_1x_{11} + a_2x_{21} + \cdots + a_px_{p1}$$
$$\mathrm{y}_2 = a_1x_{12} + a_2x_{22} + \cdots + a_px_{p2}$$
$$\mathrm{y}_3 = a_1x_{13} + a_2x_{23} + \cdots + a_px_{p3}$$
$$\vdots$$
$$\mathrm{y}_m = a_1x_{1m} + a_2x_{2m} + \cdots + a_px_{pm}$$

しかし、被説明変数は未知の変動要因を含んでいますから、一般的には次のように誤差項を加えて、式を完成します

$$\mathrm{y}_1 = a_1x_{11} + a_2x_{21} + \cdots + a_px_{p1} + e_1$$
$$\mathrm{y}_2 = a_1x_{12} + a_2x_{22} + \cdots + a_px_{p2} + e_2$$
$$\mathrm{y}_3 = a_1x_{13} + a_2x_{23} + \cdots + a_px_{p3} + e_3$$
$$\vdots$$
$$\mathrm{y}_m = a_1x_{1m} + a_2x_{2m} + \cdots + a_px_{pm} + e_m$$

重回帰分析はこの誤差項の最適化・最小化です。最適化にはさまざまな方法が考えられます。最小二乗法も誤差項の最小化の一つの方法です。
これを行列で表すと、誤差項のベクトルは次のようになります。

$$\boldsymbol{E} = \boldsymbol{Y} - \boldsymbol{XA}$$

$$\boldsymbol{E} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mp} \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$

誤差項のベクトルの大きさを最小にすることを考えます。この場合、ベクトルの大きさは、誤差項の総和ではありません。ベクトルや行列の大きさとはノームです。 q次のノームL^qの定義底は以下の通りです。

$$\boldsymbol{X} = (x_1, x_2, \cdots x_n)、 1 \leq q \leq \infty$$
$$L^q = \sqrt[q]{|x_1|^q + |x_2|^q + \cdots + |x_n|^q}$$

L^qを$\|\boldsymbol{X}\|_q$と表すこともあります。これを使うと

$$\|\boldsymbol{X}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

$$\|\boldsymbol{X}\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}$$

となりますが、L^2はユークリッド距離です。最小二乗法はL^2を最小化する方法です。

$$\|\boldsymbol{E}\|_2 = \sqrt{|e_1|^2 + |e_2|^2 + \cdots + |e_m|^2}$$

$$= \sqrt{{e_1}^2 + {e_2}^2 + \cdots + {e_m}^2}$$

$$\|\boldsymbol{E}\|_2{}^2 = {e_1}^2 + {e_2}^2 + \cdots + {e_m}^2$$

$$= (e_1, e_2, \cdots . e_m)\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{pmatrix}$$

$$\boldsymbol{E}^T = (e_1, e_2, \cdots . e_m)$$

。

$$\boldsymbol{E} = \boldsymbol{Y} - \boldsymbol{XA}$$

$$\boldsymbol{E}^T = (\boldsymbol{Y} - \boldsymbol{XA})^T$$

$$\|\boldsymbol{E}\|_2{}^2 = (\boldsymbol{Y} - \boldsymbol{XA})^T(\boldsymbol{Y} - \boldsymbol{XA})$$

$$= \boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{A}^T\boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{Y}^T\boldsymbol{XA} + \boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{XA}$$

$$= \boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{2Y}^T\boldsymbol{XA} + \boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{XA}$$

$\|E\|_2{}^2$は正定値なので、以下の偏微分によって最小値を求めることが出来ます。

$$\frac{\partial\|\boldsymbol{E}\|_2{}^2}{\partial\boldsymbol{A}} = 0$$

$$\frac{\partial\|\boldsymbol{E}\|_2{}^2}{\partial\boldsymbol{A}} = \frac{\partial(\boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{2Y}^T\boldsymbol{XA} + \boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{XA})}{\partial\boldsymbol{A}}$$

$$= \frac{\partial(\boldsymbol{Y}^T\boldsymbol{Y})}{\partial\boldsymbol{A}} - \boldsymbol{2}\frac{\partial(\boldsymbol{Y}^T\boldsymbol{XA})}{\partial\boldsymbol{A}} + \frac{\partial(\boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{XA})}{\partial\boldsymbol{A}} = \boldsymbol{0}$$

$$\frac{\partial(\boldsymbol{Y}^T\boldsymbol{Y})}{\partial\boldsymbol{A}} = 0$$

$$\frac{\partial(\boldsymbol{Y}^T\boldsymbol{XA})}{\partial\boldsymbol{A}} = \boldsymbol{Y}^T\boldsymbol{X}$$

$$\frac{\partial(\boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{XA})}{\partial\boldsymbol{A}} = \boldsymbol{2X}^T\boldsymbol{XA}$$

$$\frac{\partial\|\boldsymbol{E}\|_2{}^2}{\partial\boldsymbol{A}} = -\boldsymbol{2Y}^T\boldsymbol{X} + \boldsymbol{2X}^T\boldsymbol{XA} = \boldsymbol{0}$$

$$\boldsymbol{Y}^T\boldsymbol{X} = \boldsymbol{X}^T\boldsymbol{XA}$$

$$\boldsymbol{A} = (\boldsymbol{X}^T\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{Y}^T\boldsymbol{X}$$

$$\boldsymbol{A} = (\boldsymbol{X}^T\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}^T\boldsymbol{Y}$$

ここで、 $(\boldsymbol{X}^T\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}^T$ を次のように一つの行列だと考えます。

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \boldsymbol{X}^\#$$

$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A}$と表せるときに（これを行列が正則だと言います。)、次のように、m$\boldsymbol{Y}$に逆行列$\boldsymbol{X}^{-1}$を左から掛けて、$\boldsymbol{A}$を求めます。

$$\boldsymbol{X}^{-1}\boldsymbol{Y} = \boldsymbol{X}^{-1}\boldsymbol{X}\boldsymbol{A} = \boldsymbol{A}$$

同じように、正則でない場合にも、最適解$\boldsymbol{A}$を$\boldsymbol{X}^\#$を左から掛けて求めることが出来ます。

$$\boldsymbol{X}^\#\boldsymbol{Y} = \boldsymbol{A}$$

$$\boldsymbol{X}^\# = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

式 71

$$\boldsymbol{X}^\#\boldsymbol{X} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{X}) = \boldsymbol{I}$$

唯一解ではなくて最適解が求まるのですが、$\boldsymbol{X}^\#$には逆行列と似たよう機能があります。そこで、この行列を疑似逆行列と呼びます。逆に言えば、$\boldsymbol{X}$ が正則であれば、$\boldsymbol{X}^\# = \boldsymbol{X}^{-1}$であり、疑似逆行列$\boldsymbol{X}^\#$は連立方程式の解を与えます。

練習のために、この方法で単回帰分析の式を作ってみます。

$$y = ax + b$$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A}$$

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$\boldsymbol{X}^T = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{pmatrix}\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \boldsymbol{x_i}^2 & \sum_{i=1}^{n} \boldsymbol{x_i} \\ \sum_{i=1}^{n} \boldsymbol{x_i} & \boldsymbol{n} \end{pmatrix}$$

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{pmatrix} \sum_{i=1}^{n} \boldsymbol{x_i}^2 & \sum_{i=1}^{n} \boldsymbol{x_i} \\ \sum_{i=1}^{n} \boldsymbol{x_i} & \boldsymbol{n} \end{pmatrix}^{-1} = \frac{1}{\boldsymbol{n}\sum_{i=1}^{n} \boldsymbol{x_i}^2 - (\sum_{i=1}^{n} \boldsymbol{x_i})^2}\begin{pmatrix} \boldsymbol{n} & -\sum_{i=1}^{n} \boldsymbol{x_i} \\ -\sum_{i=1}^{n} \boldsymbol{x_i} & \sum_{i=1}^{n} \boldsymbol{x_i}^2 \end{pmatrix}$$

$$\boldsymbol{X}^\# = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \frac{1}{\boldsymbol{n}\sum_{i=1}^{n} \boldsymbol{x_i}^2 - (\sum_{i=1}^{n} \boldsymbol{x_i})^2}\begin{pmatrix} \boldsymbol{n} & -\sum_{i=1}^{n} \boldsymbol{x_i} \\ -\sum_{i=1}^{n} \boldsymbol{x_i} & \sum_{i=1}^{n} \boldsymbol{x_i}^2 \end{pmatrix}\begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

$$= \frac{1}{\boldsymbol{n}\sum_{i=1}^{n}\boldsymbol{x_i}^2 - (\sum_{i=1}^{n}\boldsymbol{x_i})^2}\begin{pmatrix} nx_1 - \sum_{i=1}^{n}\boldsymbol{x_i} & \boldsymbol{n}x_2 - \sum_{i=1}^{n}\boldsymbol{x_i} & \cdots & \boldsymbol{n}x_n - \sum_{i=1}^{n}\boldsymbol{x_i} \\ -x_1\sum_{i=1}^{n}\boldsymbol{x_i} + \sum_{i=1}^{n}\boldsymbol{x_i}^2 & -x_2\sum_{i=1}^{n}\boldsymbol{x_i} + \sum_{i=1}^{n}\boldsymbol{x_i}^2 & \cdots & x_n\sum_{i=1}^{n}\boldsymbol{x_i} + \sum_{i=1}^{n}\boldsymbol{x_i}^2 \end{pmatrix}$$

$$\boldsymbol{X}^{\#}\boldsymbol{Y} = \frac{1}{\boldsymbol{n}\sum_{i=1}^{n}\boldsymbol{x_i}^2 - (\sum_{i=1}^{n}\boldsymbol{x_i})^2}\begin{pmatrix} nx_1 - \sum_{i=1}^{n}\boldsymbol{x_i} & \boldsymbol{n}x_2 - \sum_{i=1}^{n}\boldsymbol{x_i} & \cdots & \boldsymbol{n}x_n - \sum_{i=1}^{n}\boldsymbol{x_i} \\ -x_1\sum_{i=1}^{n}\boldsymbol{x_i} + \sum_{i=1}^{n}\boldsymbol{x_i}^2 & -x_2\sum_{i=1}^{n}\boldsymbol{x_i} + \sum_{i=1}^{n}\boldsymbol{x_i}^2 & \cdots & x_n\sum_{i=1}^{n}\boldsymbol{x_i} + \sum_{i=1}^{n}\boldsymbol{x_i}^2 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$= \frac{1}{\boldsymbol{n}\sum_{i=1}^{n}\boldsymbol{x_i}^2 - (\sum_{i=1}^{n}\boldsymbol{x_i})^2}\begin{pmatrix} \boldsymbol{n}\sum_{i=1}^{n}\boldsymbol{x_i}\boldsymbol{y_i} + \sum_{i=1}^{n}\boldsymbol{x_i}\sum_{i=1}^{n}\boldsymbol{y_i} \\ \sum_{i=1}^{n}\boldsymbol{x_i}\boldsymbol{y_i}\sum_{i=1}^{n}\boldsymbol{x_i} + \sum_{i=1}^{n}\boldsymbol{y_i}\sum_{i=1}^{n}\boldsymbol{x_i}^2 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

$$a = \frac{\boldsymbol{n}\sum_{i=1}^{n}\boldsymbol{x_i}\boldsymbol{y_i} + \sum_{i=1}^{n}\boldsymbol{x_i}\sum_{i=1}^{n}\boldsymbol{y_i}}{\boldsymbol{n}\sum_{i=1}^{n}\boldsymbol{x_i}^2 - (\sum_{i=1}^{n}\boldsymbol{x_i})^2}$$

$$b = \frac{\sum_{i=1}^{n}\boldsymbol{x_i}\boldsymbol{y_i}\sum_{i=1}^{n}\boldsymbol{x_i} + \sum_{i=1}^{n}\boldsymbol{y_i}\sum_{i=1}^{n}\boldsymbol{x_i}^2}{\boldsymbol{n}\sum_{i=1}^{n}\boldsymbol{x_i}^2 - (\sum_{i=1}^{n}\boldsymbol{x_i})^2}$$

a が回帰係数になっていることを確かめてください。

ここで言いたかったことは、重回帰の解法を線形代数的に説明すると、疑似逆行列を掛けることになるということです。