Publication: lifemater.com, Author: peekAboo
Creation: 2021/12/30, Update: 2022/1/30


# 最小二乗法による直線単回帰

## 1. 最小二乗法による直線回帰

$n$ 個のデータ点$\{(x_i, y_i)\}$を最小二乗法を用いて直線回帰式$f(x)$する。ここでは、$x$は制御可能なインプト、$y$ をアウトプットして、$y$ に関する二乗誤差$I$を最小化するように直線回帰する。直線回帰なので$f(x)$は以下のように $x$ に関する一次関数として表現できる。

$$y = f(x) = ax + b, \tag{1-1}$$

$a$, $b$ は定数であり、この $a$, $b$ を$\{(x_i, y_i)\}$から決定する。$I$ は以下の式で表現することができ、$a, b$ に依存する。

$$I(a,b) = \sum_i \{y_i - f(x_i)\}^2, \tag{1-2}$$

Eq. (1-1)を代入して展開すると、

$$I(a,b) = \sum_i \{y_i - (ax_i + b)\}^2,$$

となる。$I(a,b)$ の最小化を考えるので、

$$\frac{\partial I}{\partial a} = \frac{\partial I}{\partial b} = 0, \tag{1-3}$$

を求める。それぞれの偏微分を計算すると、

$$\begin{aligned}\frac{\partial I}{\partial a} &= \sum_i 2\{y_i - (ax_i + b)\}(-x_i) = 2\sum_i (-x_i y_i + a x_i^2 + b x_i) \\ &= 2\left(-\sum_i x_i y_i + \sum_i a x_i^2 + \sum_i b x_i\right) = 2\left(-\sum_i x_i y_i + a\sum_i x_i^2 + b\sum_i x_i\right),\end{aligned}$$

となる。

ここで、ある $n$ 個のデータ $\{(x_i, y_i)\}$ について以下の式を定義する。

$$E[x] = \frac{1}{n}\sum_i x_i\,, \tag{1-4}$$

$$\mathrm{Cov}[x, y] = E\Big[\big(x - E[x]\big)\big(y - E[y]\big)\Big]. \tag{1-5}$$

$E[x]$は$\{x_i\}$の平均値、$\mathrm{Cov}[x,x]$は$\{x_i\}$の分散、$\mathrm{Cov}[x,y]$は$\{(x_i, y_i)\}$の共分散である。

Eq. (1-4)を用いると、

$$\frac{\partial I}{\partial a} = 2\left(-\sum_i x_i y_i + a\sum_i x_i^2 + b\sum_i x_i\right) = 2n\left(-E[xy] + aE\left[x^2\right] + bE[x]\right),$$

となる。同様にして、

$$\begin{aligned}\frac{\partial I}{\partial b} &= \sum_i 2\{y_i - (ax_i + b)\}(-1) = 2\sum_i(-y_i + ax_i + b)\\ &= 2\left(-\sum_i y_i + a\sum_i x_i + b\sum_i 1\right) = 2\left(-nE[y] + anE[x] + bnE[1]\right),\\ &= 2n\left(-E[y] + aE[x] + b\right)\end{aligned}$$

となるので、Eq. (1-3)より以下の式が成立する。

$$\frac{\partial I}{\partial a} = 2n\left(-E[xy] + aE\left[x^2\right] + bE[x]\right) = 0\,, \tag{1-6}$$

$$\frac{\partial I}{\partial b} = 2n\left(-E[y] + aE[x] + b\right) = 0\,. \tag{1-7}$$

Eqs. (1-6, 1-7)を行列表現すると、

$$\begin{bmatrix} E\left[x^2\right] & E[x] \\ E[x] & 1 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} E[xy] \\ E[y] \end{bmatrix}, \tag{1-8}$$

となるので、

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} E\left[x^2\right] & E[x] \\ E[x] & 1 \end{bmatrix}^{-1}\begin{bmatrix} E[xy] \\ E[y] \end{bmatrix}, \tag{1-9}$$

となる。ここで、

$$\begin{bmatrix} E\left[x^2\right] & E[x] \\ E[x] & 1 \end{bmatrix}^{-1} = \frac{1}{E\left[x^2\right]-\left(E[x]\right)^2}\begin{bmatrix} 1 & -E[x] \\ -E[x] & E\left[x^2\right] \end{bmatrix},$$

となるので、

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} E\left[x^2\right] & E[x] \\ E[x] & 1 \end{bmatrix}^{-1}\begin{bmatrix} E[xy] \\ E[y] \end{bmatrix} = \frac{1}{E\left[x^2\right]-\left(E[x]\right)^2}\begin{bmatrix} 1 & -E[x] \\ -E[x] & E\left[x^2\right] \end{bmatrix}\begin{bmatrix} E[xy] \\ E[y] \end{bmatrix}$$
$$= \frac{1}{E\left[x^2\right]-\left(E[x]\right)^2}\begin{bmatrix} E[xy]-E[x]E[y] \\ E\left[x^2\right]E[y]-E[xy]E[x] \end{bmatrix},$$

となる。平均、分散、共分散の関係式、

$$\mathrm{Cov}[x,x] = E\left[x^2\right]-\left(E[x]\right)^2. \tag{1-10}$$

$$\mathrm{Cov}[x,y] = E[xy]-E[x]E[y]. \tag{1-11}$$

を用いると、

$$E\left[x^2\right]E[y]-E[xy]E[x] = E\left[x^2\right]E[y]-E[y]\left(E[x]\right)^2+E[y]\left(E[x]\right)^2-E[xy]E[x]$$
$$= \left\{E\left[x^2\right]-\left(E[x]\right)^2\right\}E[y]+\left\{E[y]E[x]-E[xy]\right\}E[x]$$
$$= \mathrm{Cov}[x,x]E[y]-\mathrm{Cov}[x,y]E[x]$$

とできるので、

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{\mathrm{Cov}[x,x]}\begin{bmatrix} \mathrm{Cov}[x,y] \\ \mathrm{Cov}[x,x]E[y]-\mathrm{Cov}[x,y]E[x] \end{bmatrix}$$
$$= \begin{bmatrix} \dfrac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]} \\ E[y]-\dfrac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]}E[x] \end{bmatrix} = \begin{bmatrix} \dfrac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]} \\ E[y]-aE[x] \end{bmatrix},$$

が成立する。

この式から $a, b$ は以下の式で与えられる。

$$a=\frac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]}, \tag{1-12}$$

$$b=E[y]-\frac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]}E[x]=E[y]-aE[x], \tag{1-13}$$

これを、Eq. (1-1)に代入すると、

$$y=f(x)=\frac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]}x+E[y]-\frac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]}E[x]=\frac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]}\left(x-E[x]\right)+E[y],$$

となる。つまり、直線回帰式は以下の式になる。

$$y=f(x)=\frac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]}\left(x-E[x]\right)+E[y]. \tag{1-14}$$

また、$\{x_i\},\{y_i\}$ の標準偏差を $\sigma_{\mathrm{x}},\sigma_{\mathrm{y}}$、$\{(x_i,y_i)\}$ の相関係数を $\rho_{\mathrm{xy}}$ とすると、

$$\rho_{\mathrm{xy}}=\frac{\mathrm{Cov}[x,y]}{\sqrt{\mathrm{Cov}[x,x]}\sqrt{\mathrm{Cov}[y,y]}}, \tag{1-15}$$

$$\sigma_{\mathrm{x}}=\sqrt{\mathrm{Cov}[x,x]}, \tag{1-16}$$

となるので、

$$\begin{aligned} a&=\frac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]}=\frac{\rho_{\mathrm{xy}}\sqrt{\mathrm{Cov}[x,x]}\sqrt{\mathrm{Cov}[y,y]}}{\mathrm{Cov}[x,x]}\\ &=\rho_{\mathrm{xy}}\frac{\sqrt{\mathrm{Cov}[y,y]}}{\sqrt{\mathrm{Cov}[x,x]}}=\rho_{\mathrm{xy}}\frac{\sigma_{\mathrm{y}}}{\sigma_{\mathrm{x}}} \end{aligned},$$

つまり、

$$a=\rho_{\mathrm{xy}}\frac{\sigma_{\mathrm{y}}}{\sigma_{\mathrm{x}}}, \tag{1-17}$$

となる。

## 2. 正規化の影響

$k_{\mathrm{x}}, k_{\mathrm{y}}, d_{\mathrm{x}}, d_{\mathrm{y}},$ を定数として、以下の式で $X_i, Y_i$ を定義する。

$$X_i = k_{\mathrm{x}}\left(x_i - d_{\mathrm{x}}\right), \tag{2-1}$$

$$Y_i = k_{\mathrm{y}}\left(y_i - d_{\mathrm{y}}\right). \tag{2-2}$$

この量についても同様に $Y_i$ 最小二乗法で以下のように直線回帰する。

$$Y = a_{\mathrm{X}} X + b_{\mathrm{Y}}. \tag{2-3}$$

Eqs. (1-12, 1-13)より、

$$a_{\mathrm{X}} = \frac{\mathrm{Cov}[X,Y]}{\mathrm{Cov}[X,X]} = \frac{\mathrm{Cov}\left[k_{\mathrm{x}}\left(x-d_{\mathrm{x}}\right), k_{\mathrm{y}}\left(y-d_{\mathrm{y}}\right)\right]}{\mathrm{Cov}\left[k_{\mathrm{x}}\left(x-d_{\mathrm{x}}\right), k_{\mathrm{x}}\left(x-d_{\mathrm{x}}\right)\right]},$$

$$b_{\mathrm{X}} = E[Y] - a_{\mathrm{X}} E[X] = E\left[k_{\mathrm{y}}\left(y-d_{\mathrm{y}}\right)\right] - a_{\mathrm{X}} E\left[k_{\mathrm{x}}\left(x-d_{\mathrm{x}}\right)\right],$$

となる。ここで、任意の定数 $a, b, c, d$ について、

$$\mathrm{Cov}\left[a(x-b), c(y-d)\right] = ac\,\mathrm{Cov}[x,y], \tag{2-4}$$

$$E\left[a(x-b)\right] = a\left(E[x]-b\right), \tag{2-5}$$

が成立するので、

$$a_{\mathrm{X}} = \frac{\mathrm{Cov}\left[k_{\mathrm{x}}\left(x-d_{\mathrm{x}}\right), k_{\mathrm{y}}\left(y-d_{\mathrm{y}}\right)\right]}{\mathrm{Cov}\left[k_{\mathrm{x}}\left(x-d_{\mathrm{x}}\right), k_{\mathrm{x}}\left(x-d_{\mathrm{x}}\right)\right]} = \frac{k_{\mathrm{x}} k_{\mathrm{y}} \mathrm{Cov}[x,y]}{k_{\mathrm{x}} k_{\mathrm{x}} \mathrm{Cov}[x,x]} = \frac{k_{\mathrm{y}}}{k_{\mathrm{x}}} \frac{\mathrm{Cov}[x,y]}{\mathrm{Cov}[x,x]} = \frac{k_{\mathrm{y}}}{k_{\mathrm{x}}} a,$$

$$\begin{aligned} b_{\mathrm{X}} &= E\left[k_{\mathrm{y}}\left(y-d_{\mathrm{y}}\right)\right] - a_{\mathrm{X}} E\left[k_{\mathrm{x}}\left(x-d_{\mathrm{x}}\right)\right] = k_{\mathrm{y}}\left(E[y]-d_{\mathrm{y}}\right) - \frac{k_{\mathrm{y}}}{k_{\mathrm{x}}} a k_{\mathrm{x}}\left(E[x]-d_{\mathrm{x}}\right) \\ &= k_{\mathrm{y}}\left\{\left(E[y]-d_{\mathrm{y}}\right) - a\left(E[x]-d_{\mathrm{x}}\right)\right\} = k_{\mathrm{y}}\left\{\left(E[y]-aE[x]\right) - \left(d_{\mathrm{y}} - ad_{\mathrm{x}}\right)\right\} \\ &= k_{\mathrm{y}}\left\{b - \left(d_{\mathrm{y}} - ad_{\mathrm{x}}\right)\right\} \end{aligned},$$

となる。よって、

$$a_{\mathrm{X}} = \frac{k_{\mathrm{y}}}{k_{\mathrm{x}}} a , \tag{2-6}$$

$$b_{\mathrm{X}} = k_{\mathrm{y}} \left\{ b - \left( d_{\mathrm{y}} - a d_{\mathrm{x}} \right) \right\} , \tag{2-7}$$

となる。