linear_regression

Masayuki Sakai

yyyy/mm/dd

Contents

1　最小二乗法

ここでは，n 個の対になる観測値，$\boldsymbol{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ と $\boldsymbol{y} = (y_1, y_2, \dots, y_n)^T \in \mathbb{R}^n$ について最小二乗法により線形モデルである

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \varepsilon_i \sim N(0, \sigma^2)$$

のパラメーター β_0, β_1 を求める．

これは，直線から上下の距離 $|y_i - \beta_0 - \beta_1 x_i|$ の二乗和 L を最小にする $\boldsymbol{\beta} = (\beta_0, \beta_1)^T \in \mathbb{R}^2$ を求める時に，L の偏微分を用いる．まず

$$L := \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

として，これを β_0, β_i の関数と見てそれぞれの偏微分を求めると

$$\frac{\partial L}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \frac{\partial L}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

となる．この連立方程式を解くと

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

を得る．ただし $\sum_{i=1}^{n}(x_i-\bar{x})^2 \neq 0$ である．これは，共変量 $\boldsymbol{x}$ の値が全て同じことを意味しており，そのような変数はなんら情報を持たないためこのケースは実務上無視して良い．また，$\bar{x}=\sum x_i/n, \bar{y}=\sum y_i/n$ である．

ここで，$\boldsymbol{x}'=\boldsymbol{x}-\bar{x}, \boldsymbol{y}'=\boldsymbol{y}-\bar{y}$ とすると，$\bar{y'}=\bar{x'}=0$ で $\bar{\beta}_0=0$，$\hat{\beta}_1=\sum x_i y_i/\sum x_i^2$ となり計算が簡単になる．加減による傾きの影響は無いので，$\hat{\beta}_1$ の推定はこのままで問題はないが，切片項 β_0 はもとのスケールで計算する必要がある．

1.1 実装

ここでは，シンプルに $\boldsymbol{x}, \boldsymbol{y}$ という 2 つのベクトルを受け取った場合に，$\hat{\beta}_0, \hat{\beta}_1$ を返す関数を書く．

```
lsm <- function(x, y){
  if(length(unique(x)) == 1){
    stop("x has only 1 unique values")
  }
  # estimation
  bar_x <- mean(x); bar_y <- mean(y)
  hat_b1 <- sum((x-bar_x)*(y-bar_y)) / sum((x-bar_x)^2)
  hat_b0 <- bar_y - hat_b1 * bar_x

  # return
  return(list(b0 = hat_b0, b1 = hat_b1))
}
```

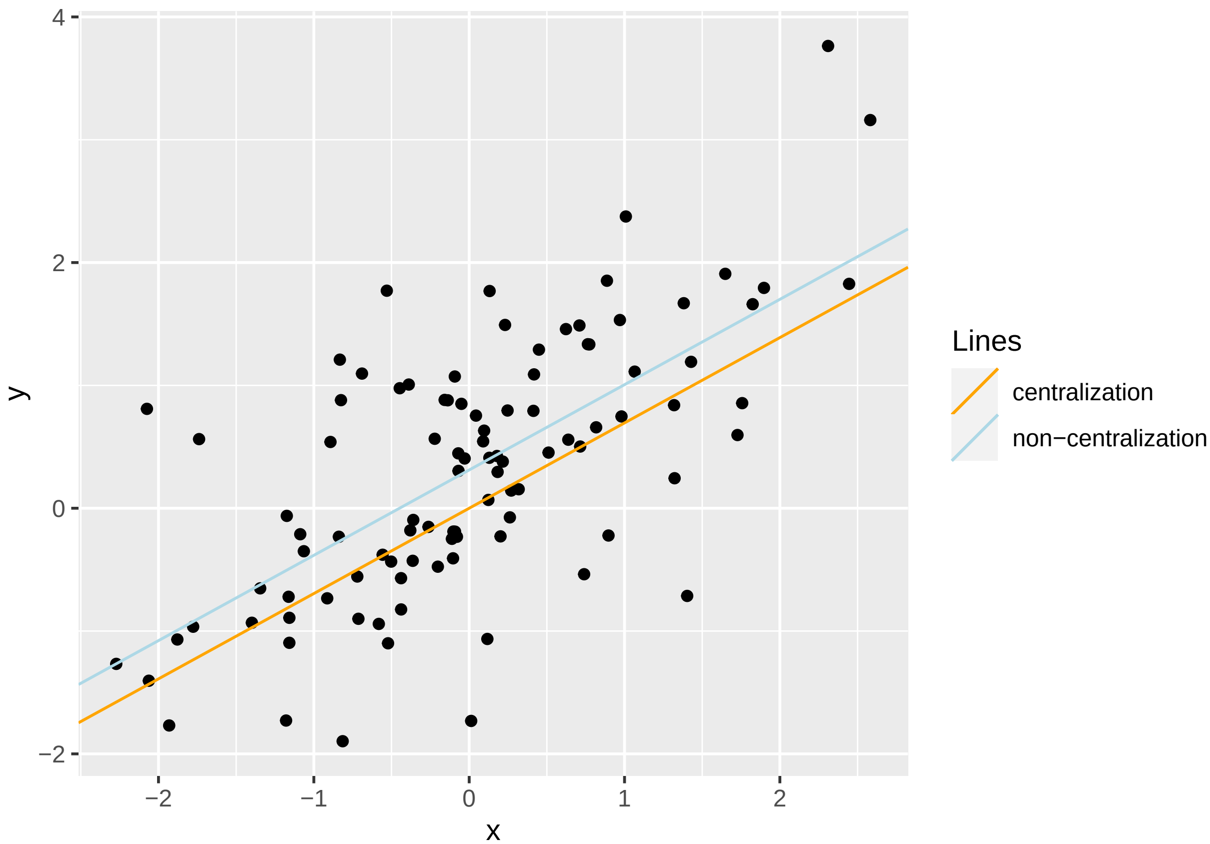
```
set.seed(100)
beta <- c(0.3, 0.8)
n <- 100
x <- c(rep(1,n), rnorm(n)) %>% matrix(nrow=100, byrow=FALSE)
y <- x %*% beta + rnorm(n)

fit <- lsm(x=x[,2], y=y[,1])
fit_centered <- lsm(x=x[,2]-mean(x[,2]), y=y-mean(y))

print(fit)
```

```
## $b0
## [1] 0.3114477
##
## $b1
## [1] 0.6946326
```

```
print(fit_centered)
```

```
## $b0
## [1] 2.193677e-17
##
## $b1
## [1] 0.6946326
```

```r
data.frame(x=x[,2], y=y) %>%
  ggplot(aes(x=x, y=y)) +
  geom_point() +
  geom_abline(aes(slope = fit$b1, intercept = fit$b0, color = 'non-centralization'), show.legend = TRUE)
  geom_abline(aes(slope = fit_centered$b1, intercept = fit_centered$b0, color = 'centralization'), show
  labs(color="Lines") +
  scale_color_manual(values = c('orange', 'lightblue'))
```



2 重回帰モデル

$$y := \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X := \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} := \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

とすると，

$$L = \|y - X\boldsymbol{\beta}\|^2$$

とかけて，

$$\nabla L := \begin{bmatrix} \frac{\partial L}{\partial \beta_0} \\ \frac{\partial L}{\partial \beta_1} \end{bmatrix} = -2X^T(y - X\boldsymbol{\beta})$$

と表すことができる．ここから共変量の数を増やして

$$X := \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \quad \boldsymbol{b}eta := \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

とする．このように拡張してもやることは変わらず，

$$-2X^T(y - X\boldsymbol{\beta}) = \mathbf{0}$$

を解く．結果として，

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T y$$

を得る．ただし，$(X^T X)^{-1}$ は存在するとする．

2.1　実装

```
lsm <- function(X, y){
  res <- solve(t(X)%*%X) %*% t(X) %*% y
  return(res)
}
```

```
n <- 100
p <- 3
set.seed(888); beta <- rnorm(4)
x <- matrix(1, nrow=n, ncol=p+1)
x[,-1] <- rnorm(100*p)
set.seed(666);y <- x %*% beta + rnorm(n)

fit <- lsm(X = x, y = y)
data.frame(beta = beta, fit = fit, diff = beta - fit)
```

```
##          beta        fit         diff
## 1 -1.9513433 -2.0154629  0.064119522
## 2 -1.5443662 -1.5412662 -0.003099945
## 3  0.7298327  0.7150430  0.014789653
## 4 -0.2775818 -0.3555181  0.077936312
```

2.2　考察

2.2.1　$(X^T X)^{-1}$ の存在

次の場合は $(X^T X)^{-1}$ が存在しない．
1. $N < p + 1$ 2. X の異なる 2 列が等しい

2.2.2 関連する補題

Lemma 1. 正方行列 $A \in \mathbb{R}^{n\times n}$ について以下は同値である．*1.* A が正則 *2.* $\mathrm{rank}(A) = n$ *3.* $\det(A) \neq 0$

Lemma 2. 正方行列 A, B について，次が成り立つ．*1.* $\det(AB) = \det(A)\det(B)$ *2.* $\det(A^T) = \det(A)$

Lemma 3. $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times l}$ として，以下が成立．

$$\begin{aligned} \mathrm{rank}(AB) &\le \min\{\mathrm{rank}(A), \mathrm{rank}(B)\} \\ \mathrm{rank}(A^T) = \mathrm{rank}(A) &\le \min\{m, n\} \end{aligned}$$

Lemma 4. V, W をそれぞれ $\mathbb{R}^n, \mathbb{R}^m$ の部分空間として，行列 $A \in \mathbb{R}^{m\times n}$ による線形写像 $V \to W$ の像と核は，それぞれ V, W の部分空間であって，それらの次元の和は n になる．また，その次元は，A の階数に一致する．

2.2.3 実務上の仮定

以降では，$X \in \mathbb{R}^{N\times(p+1)}$ の階数は $p+1$ であることを仮定する．$p = 1$ の場合は $\mathrm{rank(X) = 2 = p + 1}$ と同値．

3 β の分布

被説明変数 $\boldsymbol{y} \in \mathbb{R}^N$ が共変量 $X \in \mathbb{R}^{N\times(p+1)}$ とその係数 $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$ と誤差項 $\boldsymbol{\varepsilon} \sim N(\boldsymbol{\mu}, \sigma^2 I)$ によって下記のような関係にあると仮定する．

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

ここで，前項のように $\boldsymbol{\beta}$ と推定された $\hat{\boldsymbol{\beta}}$ は異なるものであることに注意しよう．あくまで $\boldsymbol{\beta}$ の値は未知である．いま，あらためて $\hat{\boldsymbol{\beta}}$ は次のように表せることを確認する．

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (X^T X)^{-1} X^T y \\ &= (X^T X)^{-1} X^T (X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + (X^T X)^{-1} \boldsymbol{\varepsilon} \end{aligned}$$

となり，$\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\varepsilon}$ によって決まることがわかる．$\boldsymbol{\varepsilon}$ が従う分布はその平均を $E[\boldsymbol{\varepsilon}] = \mathbf{0}$ と仮定していることから，$\hat{\boldsymbol{\beta}}$ の分布の平均は $\boldsymbol{\beta}$ となる．このように，ある推定量の平均が真値に一致する時，その推定量を不偏推定量 (unbiased estimator) と呼ぶ．また，$\hat{\boldsymbol{\beta}}$ の分散を考えよう．それぞれの要素の分散 $V(\hat{\beta}_i)$ は定義通り $E(\hat{\beta}_i - \beta_i)^2$ である．また，それぞれの要素間の共分散 $\sigma_{i,j} = \mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j), (i \neq j)$ と定めれば，$\hat{\boldsymbol{\beta}}$ の共分散行列は，次のようになる．

$$\begin{aligned} V(\hat{\boldsymbol{\beta}}) &= E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T] \\ &= E[(\boldsymbol{\beta} + (X^T X)^{-1} X^T \boldsymbol{\varepsilon} - \boldsymbol{b}eta)(\boldsymbol{\beta} + (X^T X)^{-1} X^T \boldsymbol{\varepsilon} - \boldsymbol{\beta})^T] \\ &= E[(X^T X)^{-1} X^T \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T (X^T X)^{-1} X^T] \\ &= (X^T X)^{-1} E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} \end{aligned}$$

となる．これより $\hat{\boldsymbol{\beta}}$ の分布は下記のようになることがわかった．

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 (X^T X)^{-1})$$

4　二乗誤差の分布：Residual sum of squares

ここでは，$L = \|\boldsymbol{y} - X\hat{\boldsymbol{\beta}}\|^2$ の分布を導出する．まず $H := X(X^TX)^{-1}X^T$ として，その性質を確認していく．

以下を確認しなさい．

$$\begin{aligned} H^2 &= H \\ (I-H)^2 &= I - H \\ HX &= X \\ \hat{\boldsymbol{y}} &= H\boldsymbol{y} \\ \boldsymbol{y} - \hat{\boldsymbol{y}} &= (I-H)\varepsilon \end{aligned}$$

また，$H, I-H$ については次の補題が成り立つ．

Lemma 5. $H, I-H$ の固有値は，$0, 1$ のみであり，H の固有値 *1* と $I-H$ の固有値 *0* の固有空間の次元は $p+1$，H の固有値 *0* と $I-H$ の固有値 *1* の固有空間の次元は $N-p-1$ となる．

証明は後述する．

ここで RSS を次のように定義する．

$$\begin{aligned} RSS &:= \|\boldsymbol{y} - \hat{\boldsymbol{y}}\| \\ &= [(I-H)\varepsilon]^T[(I-H)\varepsilon] \\ &= \varepsilon^T(I-H)^2\varepsilon \\ &= \varepsilon^T(I-H)\varepsilon \end{aligned}$$