로버스트 선형회귀

No.
year month day (9-1

- 단순선형모형 : $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i=1,2,\cdots,n$

  $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

(행렬형태) $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\underset{\beta_0,\beta_1}{\text{Min}} \sum_{i=1}^{n} \varepsilon_i^2 \Leftrightarrow \underset{\underline{\beta}}{\text{Min}}\ \underline{\varepsilon}'\underline{\varepsilon} \quad (\text{LSM})$$

$$Q = \underline{\varepsilon}'\underline{\varepsilon} = (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta})$$

$$\frac{\partial Q}{\partial \underline{\beta}} = 0 \Rightarrow X'X\hat{\underline{\beta}} = X'\underline{y} \quad \therefore \hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$$

(정규방정식) (BLUE)

$$E(\hat{\underline{\beta}}) = \underline{\beta}, \quad Var(\hat{\underline{\beta}}) = (X'X)^{-1}\sigma^2$$

$$\underset{\beta_0,\beta_1}{\text{Min}} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2$$

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2$$

$$\begin{cases} \dfrac{\partial Q}{\partial \beta_0} = 0 \Rightarrow -2\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^1 = 0 \quad \therefore n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i = \sum Y_i \\ \dfrac{\partial Q}{\partial \beta_1} = 0 \Rightarrow -2\sum_{i=1}^{n} x_i (Y_i - \beta_0 - \beta_1 x_i)^1 = 0 \quad \therefore \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 = \sum x_i Y_i \end{cases}$$

: 정규방정식

$$\therefore \begin{cases} \hat{\beta}_0 = \bar{y} - \beta_1 \bar{x} \\ \hat{\beta}_1 = \dfrac{S(xy)}{S(xx)} \end{cases}$$

$$\left(\bar{x} = \frac{\sum x_i}{n},\ \bar{y} = \frac{\sum y_i}{n},\ S(xx) = \sum_{i=1}^{n}(x_i - \bar{x})^2, \right.$$

$$S(xy) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$r_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$; ith residual (잔차)

$\leftarrow r_i = \hat{\varepsilon}_i$

• M 추정 (후버(Huber)의 M 추정)

LS(Least Squares) 추정 ; $\rho_0(z) = z^2$

$r_i = y_i - \beta_0 - \beta_1 x_i$

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \rho_0\left(\frac{r_i}{\sigma}\right)$$; 이상값에 민감함 (특이값)

$$\rho_1(z) = \begin{cases} z^2, & |z| \le c \\ c(2|z| - c), & \text{o.w.} \end{cases}$$

$(c = 1.345)$

Huber's M estimation :

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \rho_1\left(\frac{r_i}{\sigma}\right)$$

σ에 $\hat{\sigma}$ 대입함.

$(\hat{\sigma} = \text{MAD}($Median of Absolute Deviation$)$

$= \text{median}(|r_1|, |r_2|, \cdots, |r_n|)/0.675$

$$h(\beta_0,\beta_1) = \sum_{i=1}^{n} \rho\left(\frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma}\right)$$

M-estimation : $\displaystyle\min_{\beta_0,\beta_1} h(\beta_0,\beta_1)$

rho 함수 : convex ftn

$$\frac{\partial h}{\partial \beta_0} = 0, \quad \frac{\partial h}{\partial \beta_1} = 0$$

$$\frac{\partial h}{\partial \beta_0} = \frac{1}{\sigma}\sum_{i=1}^{n} (-1)\,\psi\left(\frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma}\right) = 0$$

$$\frac{\partial h}{\partial \beta_1} = \frac{1}{\sigma}\sum_{i=1}^{n} (-X_i)\,\psi\left(\frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma}\right) = 0$$

$$(\psi(z) = \rho'(z))$$

$$w(z) = \frac{\psi(z)}{z}$$

$$\frac{\partial h}{\partial \beta_0} = \sum_{i=1}^{n} (-1)\, w_i (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\frac{\partial h}{\partial \beta_1} = \sum_{i=1}^{n} (-X_i)\, w_i (Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\left(w_i = \frac{\psi\left(\frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma}\right)}{\frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma}} \right)$$

$$\begin{cases} \hat\beta_0 \sum w_i + \hat\beta_1 \sum w_i X_i = \sum w_i Y_i \\ \hat\beta_0 \sum w_i X_i + \hat\beta_1 \sum w_i X_i^2 = \sum w_i X_i Y_i \end{cases}$$

No.
year month day (9-4)

$$\psi_1(z) = \psi(z)/2$$
$$= \begin{cases} -c, & z \le -c \\ z, & -c < z < c \\ c, & z > c \end{cases}$$

$$w_1(z) = w(z)/2$$
$$= \begin{cases} -c/z, & z \le -c \\ 1, & -c < z \le c \\ c/z, & z > c \end{cases}$$

• LMS / LTS (Least Median of Squares / Least Trimmed Squares)

LS : $\min_{\beta_0, \beta_1} \sum_{i=1}^{n} r_i^2 \iff \min_{\beta_0, \beta_1} \text{mean}(r_1^2, r_2^2, \cdots, r_n^2)$

Least Squares = Least Mean of Squares

LMS : $\min_{\beta_0, \beta_1} \text{median}(r_1^2, r_2^2, \cdots, r_n^2)$

($r_{(1)}^2 \le \cdots \le r_{(n)}^2$ 에서 $[\frac{n+1}{2}]$ 번째 값을 최소화함.)

LTS : $\min_{\beta_0, \beta_1} \sum_{i=1}^{q} r_{(i)}^2$

(q : $r_{(1)}^2 \le \cdots \le r_{(n)}^2$ 에서의 순서
default $= [n/2] + [(p+1)/2]$
$= [\frac{n+p+1}{2}]$

※ LQS : $\min_{\beta_0, \beta_1} r_{(q)}^2$

(q : $r_{(1)}^2 \le \cdots \le r_{(n)}^2$ 에서 $[\frac{n+p+1}{2}]$ 번째 값을 최소화함.)

CHUNIL