OLS for Multiple Linear Regression

Linear regression model: $E[Y_i | X_i] = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}$

$$= \underset{\sim}{\beta}^T \underset{\sim}{X} \qquad \left(\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_1 \end{pmatrix}, \ \underset{\sim}{X} = \begin{pmatrix} X_{i1} \\ \vdots \\ X_{ip} \end{pmatrix} \right)$$

$$= \underset{\sim}{X}^T \underset{\sim}{\beta}$$

$$= \underset{\sim}{X} \cdot \underset{\sim}{\beta}$$

Note! $\underset{\sim}{X}, \underset{\sim}{\beta}$ are $(p+1)$ dimensional

Data: $\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \in \mathbb{R}^n$ $\qquad X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & & X_{np} \end{pmatrix} \in \mathbb{R}^{n \times p}$

$$= \begin{pmatrix} \underset{\sim}{X}_1^T \longrightarrow \\ \underset{\sim}{X}_2^T \longrightarrow \\ \vdots \\ \underset{\sim}{X}_n^T \longrightarrow \end{pmatrix}$$

Expectation of a vector:

$$X\beta = \begin{pmatrix} \underset{\sim}{X}_1^T \longrightarrow \\ \vdots \\ \underset{\sim}{X}_n^T \longrightarrow \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} \underset{\sim}{X}_1^T \beta \\ \vdots \\ \underset{\sim}{X}_n^T \beta \end{pmatrix}$$

Recall: $E[Y_i | X_i] = \underset{\sim}{X}_i^T \beta \iff Y_i = X_i^T \beta + \epsilon_i, \ E[\epsilon_i | \underset{\sim}{X}_i] = 0$

Linear Regression model

$$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_1^T \beta + \epsilon_1 \\ \vdots \\ X_n^T \beta + \epsilon_n \end{pmatrix} = \begin{pmatrix} X_1^T \beta \\ \vdots \\ X_n^T \beta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\underset{n \times 1}{\underset{\sim}{Y}} = \underset{n \times (p+1)}{X} \underset{(p+1) \times 1}{\beta} + \underset{n \times 1}{\underset{\sim}{\epsilon}}$$

Linear regression model (Matrix form)

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\epsilon} \qquad E[\underset{\sim}{\epsilon} | X] = \underset{\sim}{0}.$$

Residual Sum of Squares: for observed $\underset{\sim}{y}, \underset{\sim}{x}_i$,

$$RSS(\underset{\sim}{\beta}) = \sum (y_i - x_i^T\beta)^2$$

$$= (\underset{\sim}{y} - X\beta)\cdot(\underset{\sim}{y} - X\beta)$$

$$= (\underset{\sim}{y} - X\beta)^T(\underset{\sim}{y} - X\beta) = \|\underset{\sim}{y} - X\beta\|^2$$

OLS Estimates

The OLS estimates $\hat{\beta}$ of β are the values of β minimizing $RSS(\beta)$

$$\hat{\underset{\sim}{\beta}} = \arg\min_{\beta} RSS(\beta)$$

Derivation:

$$RSS(\beta) = (y - X\beta)^T(y - X\beta)$$
$$= y^Ty - \beta^TX^Ty - y^TX\beta + \beta^TX^TX\beta$$
$$= y^Ty - 2\beta^TX^Ty + \beta^TX^TX\beta$$

$$\frac{\partial RSS(\beta)}{\partial \beta} = -2X^Ty + 2X^TX\beta$$

$$\frac{\partial RSS(\beta)}{\partial \beta}\Big|_{\hat{\beta}} = -2X^Ty + 2X^TX\hat{\beta} = 0$$

$$X^TX\hat{\beta} = X^Ty$$

$$\hat{\beta} = \frac{X^Ty}{X^TX} = (X^TX)^{-1}X^Ty$$

Exercise: Confirm that if $p=1$, result from SLR holds.

Properties of OLS estimators

(A1): $\underset{\sim}{y} = X\beta + \underset{\sim}{\epsilon}$, $E[\underset{\sim}{\epsilon} | X] = 0$, for some $\beta \in \mathbb{R}^{p+1}$

(R1): $E[\hat{\beta} | X] = E[(X^TX)^{-1}X^T y | x]$

$= (X^TX)^{-1} X^T E[y | x]$

$= (X^TX)^{-1} X^T X \beta = \beta \Rightarrow \hat{\beta}$ is unbiased est. of β.

(A2): $\epsilon_1 \dots \epsilon_n \sim$ IID $N(0, \sigma^2)$ or equivalently,

$\underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \sigma^2 I)$

(R2) $\text{Var}(\hat{\beta} | x) \equiv \begin{bmatrix} V(\hat{\beta}_0) & C(\hat{\beta}_0, \hat{\beta}_1) & & \\ C(\hat{\beta}_0, \hat{\beta}_1) & V(\hat{\beta}_1) & & \\ \vdots & & \ddots & \\ & & & V(\hat{\beta}_p) \end{bmatrix} = \sigma^2 (X^TX)^{-1}$

(compare to $V(\hat{\beta}_1) = \sigma^2 / S_{xx} = \sigma^2 (S_{xx})^{-1}$ for simple linear regression)

Proof of (R2)

(Lemma 1) (a) If $E[\underset{\sim}{z}] = 0$, then $\text{Cov}(\underset{\sim}{z}) = E[\underset{\sim}{z}\underset{\sim}{z}^T]$

In general, $\text{Cov}(\underset{\sim}{w}) = E[(\underset{\sim}{w} - \underset{\sim}{\mu})(\underset{\sim}{w} - \underset{\sim}{\mu})^T]$, $E(\underset{\sim}{w}) = \underset{\sim}{\mu}$

(b) Therefore, if $\underset{\sim}{w} = \underset{\sim}{\mu} + \underset{\sim}{z}$, $E[\underset{\sim}{z}] = 0$

$\Rightarrow \text{Cov}(\underset{\sim}{w}) = \text{Cov}(\underset{\sim}{z}) = E(\underset{\sim}{z}\underset{\sim}{z}^T)$

$\Rightarrow \text{Cov}(\underset{\sim}{y}) = \text{Cov}(\underset{\sim}{\epsilon}) = E(\underset{\sim}{\epsilon}\underset{\sim}{\epsilon}^T) = \sigma^2 I$

Lemma 2: Let $\underset{\sim}{V} = \underset{\sim}{A}(\underset{\sim}{\mu} + \underset{\sim}{Z})$, where $E(\underset{\sim}{Z}) = 0$

then $\underset{\sim}{V} = \underset{\sim}{A}\underset{\sim}{\mu} + \underset{\sim}{A}\underset{\sim}{Z}$

$= \underset{\sim}{\mu^*} + \underset{\sim}{Z^*}$ where $Z^* = AZ$

$E(Z^*) = E(AZ) = A\,E(Z) = 0$

then $\mathrm{Var}(\underset{\sim}{V}) = E[\underset{\sim}{Z^*}\underset{\sim}{Z^{*T}}]$

$= A\,E(ZZ^T)A^T = A\,\mathrm{Cov}(Z)A^T$

Back to regression: $\hat{\beta} = (X^TX)^{-1}X^Ty$

$= (X^TX)^{-1}X^T(X\beta + \epsilon)$

$= \beta + (X^TX)^{-1}X^T\epsilon$

$\mathrm{Cov}(\hat{\beta}) = \mathrm{Cov}((X^TX)^{-1}X^T\epsilon)$

$= E[(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1}]$

$= (X^TX)^{-1}X^T E[\epsilon\epsilon^T]\,X\,(X^TX)^{-1}$ ($E[\epsilon\epsilon^T] = \sigma^2 I$)

$= (X^TX)^{-1}X^TX(X^TX)^{-1}\sigma^2 = \underline{\underline{\sigma^2(X^TX)^{-1}}}$