Multiple Linear Regression

Wednesday, September 1, 2021 6:24 PM

- Overview
  - $X \to \{x_1 \ldots x_p\}$, Y
  - Model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \cdots + \beta_p X_p + \varepsilon$$

* $\beta_j \to$ average effect on Y by X_j holding all other Xs constant!

- Estimation
  - $$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_p x_p$$
  - RSS

$$RSS = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \cdots \hat{\beta}_p x_p)^2$$

* OLS

$X \to n \times p$ ($p+1$ for constant)

$y \to n \times 1$

$\varepsilon \to n \times 1$

$\beta \to p \times 1$

1

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$y = X\beta + \varepsilon$$

$$\varepsilon = y - X\hat{\beta}$$

$$RSS = [\varepsilon_1 \ \varepsilon_2 \ \cdots \ \varepsilon_p] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

$$= \varepsilon^T \varepsilon$$

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

$$= y^T y - \hat{\beta}^T X^T y - y^T X\beta + \hat{\beta}^T X^T X \hat{\beta}$$

$$= y^T y - 2\hat{\beta}^T X^T y + \hat{\beta}^T X^T X \hat{\beta}$$

$$\frac{\partial \varepsilon^T \varepsilon}{\partial \hat{\beta}} = -2X^T y + 2X^T X \hat{\beta} = 0$$

$$\therefore (X^T X)\hat{\beta} = X^T y$$ Normal Equations

$$(n \times p)^T (n \times p)$$

$$p \times p$$

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

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

- Analysis
  ↳ Is there a relationship between Xs & Y

$H_0: \beta_1 = \beta_2 \cdots \beta_p = 0$
$H_a:$ at least 1 $\beta_j \neq 0$

- F-stat:

$$F = \frac{(TSS - RSS)/p}{RSS/(n-p-1)}$$

$$E\left[\frac{RSS}{n-p-1}\right] = \sigma^2 \rightarrow \text{if } H_0 \text{ is true}$$

$$E\left[\frac{TSS-RSS}{p}\right] = \sigma^2$$

* As F increases → reject H_0

* Marginal → subset q of features to test if they are 0

$$F = \frac{(RSS_0 - RSS)/q}{RSS/(n-q-1)}$$

RSS_0 is model without $\{q\}$ features

q is each individual feature
↳ $F = (t\text{-stat})^2$