Full Linear Regression 3-Derivations

Problem

Y $n \times 1$ response vector $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

X $n \times p$ full rank predictor matrix $\begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$

Objective

find β, a $n \times 1$ vector, minimizing $\|y - \beta X\|^2$

1st Derivation – Partial Derivatives

$$* = \|y - X\beta\|^2$$

$$= y^Ty - 2y^TX\beta + \beta^TX^TX\beta \quad \text{Expand}$$

$$\frac{\partial *}{\partial \beta} = -2X^Ty + 2X^TX\beta \quad \text{Take partial Derivative}$$

$$0 = -2X^Ty + 2X^TX\beta \quad \text{Set equal to zero to find minimum}$$

$$X^TX\beta = 2X^Ty \quad \text{normal equations}$$

$$\boxed{\hat\beta = (X^TX)^{-1}X^Ty} \quad \text{Solution}$$

Proof minimum

$$\frac{\partial *}{\partial \beta \partial \beta^T} = 2X^TX \quad \text{Second derivative}$$

X^TX is a positive definite matrix

$\therefore \hat\beta$ is the minimum

2nd Derivation – Hat Matrix

Hat Matrix

$$H_X = X(X^TX)^{-1}X^T$$

H_X Characteristics

$H_X^T = H_X$ symmetric

$H_XH_X = H_X$ idempotent

$(I - H_X)$ also symmetric and idempotent

$(I - H_X)x_a = 0$

Proof that $\hat\beta$ is the minimizer

$$\|y - X\beta\|^2 = \|y - X\hat\beta + X\hat\beta - X\beta\|^2 \quad \text{add/subtract } X\hat\beta$$

$$= \|y - X\hat\beta\|^2 + 2(y - X\hat\beta)^T(X\hat\beta - X\beta) + \|X\hat\beta - X\beta\|^2 \quad \text{expand}$$

$$\geq \|y - X\hat\beta\|^2 + 2(y - X\hat\beta)^T(X\hat\beta - X\beta) \quad \|X\hat\beta - X\beta\|^2 \geq 0$$

$$= \|y - X\hat\beta\|^2 + \underbrace{2(y - X(X^TX)^{-1}X^Ty)^T(X\hat\beta - X\beta)}_{\Rightarrow 0} \quad \text{plugging in } H_X$$

$$= \|y - X\hat\beta\|^2$$

$$\therefore \|y - X\beta\|^2 \geq \|y - X\hat\beta\|^2$$

3rd Derivation – Adjustment Mechanism

$$* \|y - X_1\beta_1 - \dots - X_p\beta_p\|$$

fix $\beta_2 \dots \beta_p$ considers $y - X_2\beta_2 \dots X_p\beta_p$ as single outcome

$$\therefore \beta_1(\beta_2 \dots \beta_p) = \frac{\langle y_1 - X_2\beta_2 - \dots X_p\beta_p, X_1 \rangle}{\langle X_1, X_1 \rangle}$$

How linear regression accounts for linear associations of other variables when estimating β for a variable

Quiz

1. True – based on Hat matrix characteristics
2. For design matrix, X, where $X^TX = I$
   least squares coefficients of Y
   X^TY and vector w/ i^{th} element $\langle X_i, Y \rangle$
   these are the same thing
   $\hat\beta = (\cancel{X^TX})^{-1}X^Ty$ I $\therefore \hat\beta = X^Ty$
3. X, X^T, X^TX, XX^T
   assuming X is full rank
   X rank p
   X^T rank p
   X^TX
   $(p \times n)(n \times p) \Rightarrow (p \times p) \therefore$ rank p
   $\underset{(n \times p)(p \times n)}{XX^T} \Rightarrow n \times n$ potentially $>$ rank p?
4. $1_a \otimes I_b = \begin{bmatrix} 1_{11} & 0 & 0 \cdots 0_{b1} & \cdots & 0_{ab1} \\ 0 & 1_{22} & & & \vdots \\ \vdots & & \ddots & & \vdots \\ 0_{1b} & & 1_{bb} & \cdots & 0_{abb} \\ 0_{1b+1} & & \vdots & \ddots & \vdots \\ \vdots & & & & \\ 0_{1ab} & \cdots & & \cdots & 1_{abab} \end{bmatrix} = I_{ab}$

   empirical means

5.

2nd Derivation – Hat Matrix

Hat Matrix

$$H_X = X(X^TX)^{-1}X^T$$

H_X Characteristics

$H_X^T = H_X$ symmetric

$H_XH_X = H_X$ idempotent

$(I - H_X)$ also symmetric and idempotent

$(I - H_X)x_a = 0$

Proof that $\hat\beta$ is the minimizer

$$\|y - X\beta\|^2 = \|y - X\hat\beta + X\hat\beta - X\beta\|^2 \quad \text{add/subtract } X\hat\beta$$

$$= \|y - X\hat\beta\|^2 + 2(y - X\hat\beta)^T(X\hat\beta - X\beta) + \|X\hat\beta - X\beta\|^2 \quad \text{expand}$$

$$\geq \|y - X\hat\beta\|^2 + 2(y - X\hat\beta)^T(X\hat\beta - X\beta) \quad \|X\hat\beta - X\beta\|^2 \geq 0$$

$$= \|y - X\hat\beta\|^2 + \underbrace{2(y - X(X^TX)^{-1}X^Ty)^T(X\hat\beta - X\beta)}_{\Rightarrow 0} \quad \text{plugging in } H_X$$

$$= \|y - X\hat\beta\|^2$$

$$\therefore \|y - X\beta\|^2 \geq \|y - X\hat\beta\|^2$$