Matrix form of Multiple Linear Regression

– Dr J. Dhalia Sweetlin
Associate Professor, IT Dept.

- To Develop a multi linear regression model with one dependent variable and many independents variables (features).
- Finding best regression line/plane to fit these points to

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots \beta_k x_k + \varepsilon.$$

where $\beta_0 \Rightarrow$ intercept term.
$\beta_1 \ldots \beta_k \rightarrow$ regression coefficients.
$x_1, x_2 \ldots x_k \rightarrow$ features.
$\varepsilon \longrightarrow$ Error term (epsilon)

Errors can be negative/positive. To rule out this sum of squared errors is considered. Also called as sum of residuals.

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \ldots \beta_k x_{1k} + \varepsilon_1$$
$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \ldots \beta_k x_{2k} + \varepsilon_2$$
$$\vdots$$
$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \ldots \beta_k x_{nk} + \varepsilon_n.$$

Representing the above equations in matrix form we get

$$Y = X\beta + \varepsilon$$

$$\underset{n\times 1}{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}} = \underset{n\times k}{\begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}} \underset{k\times 1}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}} + \underset{n\times 1}{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}$$

from (1) we know that $\boxed{\varepsilon = Y - X\beta}$ — (2).

Applying sum of least squared errors principle. squaring error term, differentiating partially and equating to zero will minimize the error term

Sum of residuals $S_r = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon_i^T \cdot \varepsilon$ — (3)

$$\boxed{\begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \cdots & \varepsilon_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \varepsilon_i^2}$$

sub (2) in (3)

$$S_r = \varepsilon_i^2 = (Y - X\beta)^T \cdot (Y - X\beta)$$

$$S_r = y^T y - y^T X\beta - (X\beta)^T y + (X\beta)^T X\beta \longrightarrow (4)$$

Matrix Property 1 (MP1):

Transpose rule for multiplication:

$$(A\beta)^T = \beta^T A^T.$$

Applying MP1 in (4)

$$S_r = y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta \longrightarrow (5).$$

Consider the term $\beta^T X^T y$ and finding its transpose, we can conclude that

$$(\beta^T X^T y)^T \Rightarrow y^T \beta X \quad \left[\text{According to } \boxed{\text{MP1}}\right].$$

Hence replacing $y^T \beta X$ in (5) gives

$$S_r = y^T y - (\beta^T X^T y)^T - \beta^T X^T y + \beta^T X^T X\beta \rightarrow (6)$$

Recall:

$y = n \times 1$ matrix.

$y^T \Rightarrow 1 \times n$ matrix.

$X \Rightarrow n \times k$ matrix $\Rightarrow X^T = k \times n$ matrix.

$\beta \Rightarrow k \times 1$ matrix

$\beta^T \Rightarrow 1 \times k$ matrix.

Taking the term $\beta^T X^T Y$ from (6).

$\beta^T X^T \Rightarrow$ $1\times k$ matrix is multiplied with $k\times n$ resulting in an o/p matrix of size $1\times n$

$(\beta^T X^T)Y \Rightarrow$ $1\times n$ matrix is multiplied with $n\times 1$ resulting in an o/p of size 1×1 (ie) it results in a single scalar element as output.

eg [24]

* Transpose of single scalar value does not have any effect on the position of the value (ie)

Transpose of $[24] = [24]^T = [24]$ only.

Matrix Property MP2:
If A is a scalar, then $A^T = A$

Hence, it can be understood that in (6), the term $\beta^T X^T Y$ results in a scalar of size 1×1 and also $(\beta^T X^T Y)^T$ is the same as $\beta^T X^T Y$ as per MP2 property

Applying this reduction to (6)

$$S_r = y^T y - 2\beta^T x^T y + \beta^T x^T x \beta \rightarrow (7)$$

Computing the partial derivatives of (7) w.r.t β

$$\frac{\partial S_r}{\partial \beta} = -2x^T y + 2x^T x\beta \rightarrow (8) \quad [x^2 \Rightarrow 2x] \text{ similar to}$$

Setting and equating (8) to zero

$$\frac{\partial S_r}{\partial \beta} = 0 \Rightarrow 2x^T y = 2x^T x\beta$$

$$x^T x \beta = x^T y \rightarrow (9)$$

Pre multiplying $(x^T x)^{-1}$ on both sides to (9)

gives $\beta = (x^T x)^{-1} x^T y$.

The regression coefficients can be computed using $\beta = (x^T x^{-1}) x^T y$.

This can be applied to simple linear regression also where only one dependent and one independent variables are present.