MULTIPLE LINEAR REGRESSION

(OLS Method)

We know,

$$Y = mx + b$$

$$\hat{Y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots\ldots \beta_n x_n$$

where,

$\hat{Y}$ = target value

β_0 = Intercept

β_1 = slope

We can represent our data as :-

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} \beta_0 & \beta_1 x_{11} & \cdots & \beta_m x_{1m} \\ \beta_0 & \beta_1 x_{21} & \cdots & \beta_m x_{2m} \\ \beta_0 & \beta_1 x_{31} & \cdots & \beta_m x_{3m} \\ \vdots & \vdots & \cdots & \vdots \\ \beta_0 & \beta_1 x_{n1} & \cdots & \beta_m x_{nm} \end{bmatrix}$$

Writing the matrix in different way $\Rightarrow$

$$\overbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ 1 & x_{21} & \cdots & x_{2m} \\ 1 & x_{31} & \cdots & x_{3m} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}}^{X} \overbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_m \end{bmatrix}}^{\beta}$$

$$\Rightarrow \underline{\hat{Y} = X\beta} \rightarrow (1)$$

where, $\hat{Y}$ is the matrix that includes predicted values of the data

β is the matrix of all the coefficients
X is the matrix of all the input variables

Now,

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

$$e = Y - \hat{Y}$$

$$e = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ y_3 - \hat{y}_3 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}$$

We know that for Simple Linear Regression

$$E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$E = e^T e$$

Proof:

$$\begin{bmatrix} (y_1 - \hat{y}_1) & (y_2 - \hat{y}_2) & \ldots\ldots & y_n - \hat{y}_n \end{bmatrix} \begin{bmatrix} (y_1 - \hat{y}_1) \\ (y_2 - \hat{y}_2) \\ \vdots \\ (y_n - \hat{y}_n) \end{bmatrix}$$

$$\Rightarrow (y_1-\hat{y}_1)^2+(y_2-\hat{y}_2)^2+(y_3-\hat{y}_3)^2 \dots\dots (y_n-\hat{y}_n)^2$$

$$\Rightarrow \sum_{i=1}^{n} (y_i-\hat{y}_i)^2$$

Hence, $E = e^T e$

$$\Rightarrow (y-\hat{y})^T (y-\hat{y})$$

We know, $(A-B)^T = A^T - B^T$

$$\Rightarrow (y^T-\hat{y}^T)(y-\hat{y})$$

From equation 1, $\hat{Y} = X\beta$

$$\Rightarrow [y^T-(X\beta)^T](y-X\beta)$$

$$\Rightarrow y^T y - y^T X\beta - X\beta^T y + X\beta^T X\beta$$

$$\therefore \{ y^T X\beta = X\beta^T y \}$$

$$\Rightarrow y^T y - 2y^T X\beta + X\beta^T X\beta$$

Now, we have to find the minimum value of the error.

So we will do, $\frac{dE}{d\beta} = 0$

$$\Rightarrow \frac{d}{d\beta}\left[y^T y - 2y^T X\beta + \beta^T X^T X\beta \right] = 0$$

$$\Rightarrow 0 - 2y^T X + \frac{d}{d\beta}[\beta^T X^T X\beta] = 0$$

$$\Rightarrow -2y^T X + 2X^T X\beta^T = 0$$

$$\Rightarrow 2X^TX\beta^T = 2Y^TX$$

$$\Rightarrow \beta^T = Y^TX\frac{1}{[X^TX]}$$

$$\Rightarrow \beta^T = Y^TX(X^TX)^{-1}$$

$$\Rightarrow (\beta^T)^T = y^TX(X^TX)^{-1}]^T$$

$$\Rightarrow [(X^TX)^{-1}]^T(y^TX)^T$$

$$\beta = [(X^TX)^{-1}]^T X^Ty$$

$$\beta = [x^Tx]^{-1}[x^Ty)$$

$\therefore$ Transpose of a square matrix is same

$$\boxed{\beta = (x^TX)^{-1}X^Ty}$$

where,

X = X_train data matrix

y = y_train data matrix