OLS (Ordinary List Square) Method For Multiple Linear Regression:-

- Backend Process...

1) If you have 2D dataset,
   - 1 input column
   - 1 output column
   - $\beta_0 + \beta_1 x_1 = y$

2) If you have 3D dataset,
   - 2 input column
   - 1 output column
   - $\beta_0 + \beta_1 x_1 + \beta_2 x_2 = y$

3) If you have nD dataset
   - (n-1) input column
   - 1 output column
   - $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_{n-1} x_{n-1} = y$

where, $\beta_0, \beta_1, \beta_2, \dots, \beta_{n-1}$ are weights.

y is prediction.

$x_1, x_2, x_3, \dots, x_{n-1}$ are independent variables.

Start...

$Y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_4 X_m$ ($\because X_m$ = last column)

I assumed I have m column.

Y_1 is the prediction for person 1

let I have n person / n rows.

$y_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_4 x_m$

$\vdots$

$y_n \dots$

$x_1, x_2, x_3, \dots, x_m$ varry for each person but $\beta_0, \beta_1, \beta_2, \dots \beta_m$ are fixed.

"I can write above equation in the form of matrix."

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ 1 & x_{31} & x_{32} & \dots & x_{3m} \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}_{n \times m} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}_{m \times 1} \quad (9)$$

Notation: $X_{11} \rightarrow 1^{st}$ person of 1^{st} column value.

$X_{3m} \rightarrow 3^{rd}$ person's m^{th} column value

$X_{nm} \rightarrow n^{th}$ person's m^{th} column value.

Justification of my sentence,

"I can write above equation in the form of matrix."

let's solve matrix,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_m x_{1m} \\ \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_m x_{2m} \\ \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} + \cdots + \beta_m x_{3m} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_m x_{nm} \end{bmatrix}$$

Hence,

Prediction

for person 1 $(y_1) = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_m x_{1m}$.

Prediction

for person n $(y_n) = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_m x_{nm}$.

Do more simplification of our original matrix, (a)

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = X_{n \times m} \beta_{m \times 1} \implies \boxed{\hat{y}_{n \times 1} = X_{n \times m} \beta_{m \times 1}} \quad \text{— (b)}$$

$\hat{y}$ is prediction matrix.

Now, In simple Linear Regression we calculate distance / error.

$$E = (y_i - \hat{y}_i)^2$$ i from 1 to n where n is the number of students/persons in our case.

This same equation work for Multiple Linear Regression also, just written it as $E = e.e^T$

$$\underset{n\times 1}{e} = Y_{n\times 1} - \hat{Y}_{n\times 1}$$

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

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

For $E = e.e^T$

$$= \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix}_{n\times 1} \begin{bmatrix} y_1 - \hat{y}_1 & y_2 - \hat{y}_2 & \cdots & y_n - \hat{y}_n \end{bmatrix}_{1\times n}$$

$$= (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \cdots + (y_n - \hat{y}_n)^2$$

$$= (y_i - \hat{y}_i)^2$$ i from 1 to n.

which is our equation, of L.R. also.

$E = ee^T$

Now for finding constant / bius we differentiate E w.r.t. b in Linear regression, same thing we do here,

→ But E is not containing $B_0 | B_1 | B_2 \approx \beta$ so, write it as,

$$E = ee^T$$
$$= (y-\hat{y})(y-\hat{y})^T$$
$$= (y-\hat{y})(y^T-\hat{y}^T) \quad (\because (a \pm b)^T = a^T \pm b^T) \text{ property.}$$
$$= (y - X\beta)(y^T - (X\beta)^T)$$
$$= \underbrace{yy^T - y(X\beta)^T - (X\beta)y^T} + (X\beta)(X\beta)^T \quad (\because \text{from } (b))$$
$$(\hat{y} = X\beta)$$

~~let~~ $y = n \times 1$

$(X\beta) = n \times 1$ is known,

let $n = 2$

$$\underset{2\times1}{y}\,\underset{1\times2}{(X\beta)^T} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

let $y = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$(X\beta)^T = \begin{bmatrix} 1 & 2 \end{bmatrix}$

For, $(X\beta)y^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix}$

$$= \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$\boxed{y(X\beta)^T = (X\beta)y^T}$ put in our equation,

$$E = yy^T - 2y^T(X\beta) + (X\beta)(X\beta)^T$$

$$E = yy^T - 2y^T(X\beta) + XX^T\beta\beta^T \quad \because ((AB)^T = A^TB^T)$$

Now I have β so, I can differentiate,

$$\frac{dE}{d\beta} = 0 - 2y^TX + 2\beta^TXX^T \quad \because \left(\frac{d(\beta\beta^T)}{d(\beta)} = 2\beta^T\right)$$

For maxima & minima. $\frac{dE}{d\beta} = 0$

$$\cancel{2}\beta^TXX^T = \cancel{2}y^TX$$

$$\beta^T = \frac{y^TX}{\cancel{X}^TX}$$

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

take both the side (transpose).

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

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

$$= (y^T X^T)\left((X^T X)^{-1}\right)^T$$

$\because$ Note:- $(X^T X)^{-1}$ & $\left((X^T X)^{-1}\right)^T$ are symmetric matrix

& for symmetric matrix $A^T = A$.

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

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

formula used by sci-kit learn

linear regression ~~f~~ class for Simple / Multiple Linear regression.