input $X = x_1, x_2, x_3, \dots x_N$

(x_i is a D-dimensional vector)

output $Y = y_1, y_2, \dots, y_N$

$$X = \underbrace{\begin{bmatrix} \vdots & \cdots \\ \vdots & \end{bmatrix}}_{D} \Big\} N \quad \rightarrow N \times D$$

x_{ij} = sample i, dimension j

$i = 1 \dots N$, $j = 1 \dots D$

$\hat{y} = w_1 x_1 + \dots w_D x_D$ need to find

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_D \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_D \end{bmatrix}$$

$$\rightarrow \hat{y} = w \cdot x \qquad (1 \times D)\,(D \times 1) = 1 \times 1$$
$$= w^T \cdot x$$

$$\hat{y} = w_0 + w_1 x_1 + \dots w_D x_D = w_0 x_0 + \dots w_D x_D \qquad (x_0 = 1)$$
$$= w^T x$$

Objective:

$$\boxed{E = \sum (y_i - \hat{y}_i)^2}$$
$$= \sum (y_i - w^T x_i)^2$$

Again D-derivative & set to 0 each, but we can use matrix derivation

$$\frac{\partial E}{\partial w_j} = \sum 2 \, (y_i - w^T x_i) \, \frac{\partial (w^T x_i)}{\partial w_j}$$
$$= \sum 2 \, (y_i - w^T x_i) \, x_{ij}$$

$$\Leftrightarrow \quad \sum 2 \, (y_i - w^T x_i) \, x_{ij} = 0$$

$$\Leftrightarrow \quad \underbrace{\sum (y_i x_{ij})} = \sum (w^T x_i) x_{ij}$$

$$\Leftrightarrow \quad \underbrace{\underset{1 \times N}{x_{:,j}^T} \cdot \underset{N \times 1}{y}}_{1 \times 1} = \underbrace{\underset{1 \times D}{w^T} \, (\underset{D \times N}{X^T} \cdot \underset{N \times 1}{x_{:,j}})}_{1 \times 1}$$

$$\boxed{\begin{aligned} X^T \cdot y &= [w^T \cdot (X^T \cdot X)]^T \\ &= (X^T \cdot X)\, w \end{aligned}} \rightarrow \boxed{w = (X^T \cdot X)^{-1} \cdot X^T \cdot y}$$