Least Square

Kangwon Lee, Mechanical Engineering, Korea Polytechnic University 2016

Declaration

Independent variable

$$\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

Label

$$\mathbf{y}^T = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}$$

Model

$$y_i = ax_i + b$$

Estimation by Model

$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} \hat{a}x_1 + \hat{b} \\ \hat{a}x_2 + \hat{b} \\ \vdots \\ \hat{a}x_n + \hat{b} \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \mathbf{X}\mathbf{w}$$

Error

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \hat{a}x_1 + \hat{b} \\ \hat{a}x_2 + \hat{b} \\ \vdots \\ \hat{a}x_n + \hat{b} \end{bmatrix} = \mathbf{y} - \mathbf{X}\mathbf{w}$$

Scalar

Cost to minimize by changing $\hat{a}$ and $\hat{b}$

$$J = \frac{1}{2}\sum_{i=1}^{n} e_i^{\,2} = \frac{1}{2}\sum_{i=1}^{n}\left(y_i - \hat{a}x_i - \hat{b}\right)^2$$

$$= \frac{1}{2}\sum_{i=1}^{n}\left(y_i^{\,2} - y_i\hat{a}x_i - y_i\hat{b}_i - \hat{a}x_i y_i + \hat{a}^2 x_i^{\,2} + \hat{a}x_1\hat{b} - \hat{b}y_i + \hat{b}\hat{a}x_i + \hat{b}^2\right)$$

$$= \frac{1}{2}\sum_{i=1}^{n}\left(y_i^{\,2} + \hat{a}^2 x_i^{\,2} + \hat{b}^2 - 2\hat{a}x_i y_i - 2\hat{b}y_i + 2\hat{a}x_1\hat{b}\right)$$

First Parameter $\hat{a}$

$$\frac{\partial J}{\partial \hat{a}} = \frac{\partial}{\partial \hat{a}}\frac{1}{2}\sum_{i=1}^{n}\left(y_i - \hat{a}x_i - \hat{b}\right)^2 = 0$$

$$= \frac{1}{2}\sum_{i=1}^{n}\frac{\partial}{\partial \hat{a}}\left(y_i - \hat{a}x_i - \hat{b}\right)^2$$

$$= \sum_{i=1}^{n}\left(y_i - \hat{a}x_i - \hat{b}\right)\frac{\partial}{\partial \hat{a}}\left(y_i - \hat{a}x_i - \hat{b}\right)$$

$$= \sum_{i=1}^{n}\left(y_i - \hat{a}x_i - \hat{b}\right)\left(-x_i\right) = 0$$

$$= \sum_{i=1}^{n}\left(-x_i y_i + \hat{a}x_i^{\,2} + x_i\hat{b}\right) = 0$$

$$\hat{a}\sum_{i=1}^{n} x_i^{\,2} + \hat{b}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$

Second Parameter $\hat{b}$

$$\frac{\partial J}{\partial \hat{b}} = \frac{\partial}{\partial \hat{b}}\frac{1}{2}\sum_{i=1}^{n}\left(y_i - \hat{a}x_i - \hat{b}\right)^2 = 0$$

$$= \frac{1}{2}\sum_{i=1}^{n}\frac{\partial}{\partial \hat{b}}\left(y_i - \hat{a}x_i - \hat{b}\right)^2$$

$$= \sum_{i=1}^{n}\left(y_i - \hat{a}x_i - \hat{b}\right)\frac{\partial}{\partial \hat{b}}\left(y_i - \hat{a}x_i - \hat{b}\right)$$

$$= \sum_{i=1}^{n}\left(y_i - \hat{a}x_i - \hat{b}\right)\left(-1\right) = 0$$

$$= \sum_{i=1}^{n}\left(-y_i + \hat{a}x_i + \hat{b}\right) = 0$$

$$\hat{a}\sum_{i=1}^{n} x_i + \hat{b}\sum_{i=1}^{n} 1 = \sum_{i=1}^{n} y_i$$

Simultaneous Equations about $\hat{a}$ and $\hat{b}$

$$\begin{cases} \hat{a}\sum_{i=1}^{n} x_i^2 + \hat{b}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ \hat{a}\sum_{i=1}^{n} x_i + \hat{b}\sum_{i=1}^{n} 1 = \sum_{i=1}^{n} y_i \end{cases}$$

$$\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} 1 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix}$$

Left side

$$\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} 1 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} = \mathbf{X}^T \mathbf{X}$$

Right side

$$\begin{bmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \mathbf{X}^T \mathbf{y}$$

Equation in Matrix Form

$$\mathbf{X}^T \mathbf{X} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \mathbf{X}^T \mathbf{y}$$

$$\left(\mathbf{X}^T \mathbf{X}\right)^{-1} \left(\mathbf{X}^T \mathbf{X}\right) \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{y}$$

Vector

Error in vector form

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \hat{a}x_1 + \hat{b} \\ \hat{a}x_2 + \hat{b} \\ \vdots \\ \hat{a}x_n + \hat{b} \end{bmatrix} = \mathbf{y} - \mathbf{X}\mathbf{w}$$

Cost to minimize

$$\begin{aligned} J &= \frac{1}{2}\sum_{i=1}^{n} e_i^{\,2} = \frac{1}{2}\mathbf{e}^T\mathbf{e} = \frac{1}{2}(\mathbf{y} - \mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w}) \\ &= \frac{1}{2}\left(\mathbf{y}^T - \mathbf{w}^T\mathbf{X}^T\right)(\mathbf{y} - \mathbf{X}\mathbf{w}) \\ &= \frac{1}{2}\left(\mathbf{y}^T\mathbf{y} - \mathbf{w}^T\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\mathbf{w} + \mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w}\right) \end{aligned}$$

Partial derivatives of cost about parameter **w** (i)

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{w}} &= \frac{1}{2}\frac{\partial}{\partial \mathbf{w}}\left[(\mathbf{y} - \mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w})\right] = 0 \\ &= \frac{1}{2}\left[\left[\frac{\partial}{\partial \mathbf{w}}(\mathbf{y} - \mathbf{X}\mathbf{w})^T\right](\mathbf{y} - \mathbf{X}\mathbf{w}) + (\mathbf{y} - \mathbf{X}\mathbf{w})^T\left[\frac{\partial}{\partial \mathbf{w}}(\mathbf{y} - \mathbf{X}\mathbf{w})\right]\right] \\ &= \frac{1}{2}\left[-\mathbf{X}^T(\mathbf{y} - \mathbf{X}\mathbf{w}) + (\mathbf{y} - \mathbf{X}\mathbf{w})^T[-\mathbf{X}]\right] \\ &= \frac{1}{2}\left[\left(-\mathbf{X}^T\mathbf{y} + \mathbf{X}^T\mathbf{X}\mathbf{w}\right) + \left(\mathbf{y}^T - \mathbf{w}^T\mathbf{X}^T\right)[-\mathbf{X}]\right] \\ &= \frac{1}{2}\left[\left(-\mathbf{X}^T\mathbf{y} + \mathbf{X}^T\mathbf{X}\mathbf{w}\right) + \left(-\mathbf{y}^T\mathbf{X} + \mathbf{w}^T\mathbf{X}^T\mathbf{X}\right)\right] \end{aligned}$$

Partial derivatives of cost about parameter **w** (ii)

$$\frac{\partial J}{\partial \mathbf{w}} = \begin{bmatrix} \dfrac{\partial J}{\partial \hat{a}} \\ \dfrac{\partial J}{\partial \hat{b}} \end{bmatrix} = \frac{1}{2}\frac{\partial}{\partial \mathbf{w}}\left[\left(\mathbf{y}_{n\times 1} - \mathbf{X}_{n\times 2}\mathbf{w}_{n\times 2}\right)^{T}{}_{1\times n}\left(\mathbf{y}_{n\times 1} - \mathbf{X}_{n\times 2}\mathbf{w}_{n\times 2}\right)_{n\times 1}\right] = \mathbf{0}$$

$$= \frac{1}{2}\left[\left[\frac{\partial}{\partial \mathbf{w}}\left(\mathbf{y}_{n\times 1} - \mathbf{X}_{n\times 2}\mathbf{w}_{n\times 2}\right)^{T}\right]\left(\mathbf{y}_{n\times 1} - \mathbf{X}_{n\times 2}\mathbf{w}_{n\times 2}\right)_{n\times 1} + \left(\mathbf{y}_{n\times 1} - \mathbf{X}_{n\times 2}\mathbf{w}_{n\times 2}\right)^{T}{}_{1\times n}\left[\frac{\partial}{\partial \mathbf{w}}(\mathbf{y} - \mathbf{X}\mathbf{w})\right]\right]$$

$$= \frac{1}{2}\left[-\mathbf{X}^{T}(\mathbf{y} - \mathbf{X}\mathbf{w}) + (\mathbf{y} - \mathbf{X}\mathbf{w})^{T}[-\mathbf{X}]\right]$$

$$= \frac{1}{2}\left[\left(-\mathbf{X}^{T}\mathbf{y} + \mathbf{X}^{T}\mathbf{X}\mathbf{w}\right) + \left(\mathbf{y}^{T} - \mathbf{w}^{T}\mathbf{X}^{T}\right)[-\mathbf{X}]\right]$$

$$= \frac{1}{2}\left[\left(-\mathbf{X}^{T}\mathbf{y} + \mathbf{X}^{T}\mathbf{X}\mathbf{w}\right) + \left(-\mathbf{y}^{T}\mathbf{X} + \mathbf{w}^{T}\mathbf{X}^{T}\mathbf{X}\right)\right]$$

Size

$$J = \frac{1}{2}\left(\mathbf{y}^{T}{}_{1\times n}\mathbf{y}_{n\times 1} - \mathbf{w}^{T}{}_{1\times 2}\mathbf{X}^{T}{}_{2\times n}\mathbf{y}_{n\times 1} - \mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} + \mathbf{w}^{T}{}_{1\times 2}\mathbf{X}^{T}{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1}\right)$$

$$\frac{\partial J}{\partial \mathbf{w}} = \begin{bmatrix} \dfrac{\partial J}{\partial w_1} \\ \dfrac{\partial J}{\partial w_2} \end{bmatrix} = \frac{\partial}{\partial \mathbf{w}}\frac{1}{2}\left(\mathbf{y}^{T}{}_{1\times n}\mathbf{y}_{n\times 1} - \mathbf{w}^{T}{}_{1\times 2}\mathbf{X}^{T}{}_{2\times n}\mathbf{y}_{n\times 1} - \mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} + \mathbf{w}^{T}{}_{1\times 2}\mathbf{X}^{T}{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1}\right)$$

$$= \frac{1}{2}\left(\frac{\partial}{\partial \mathbf{w}}\mathbf{y}^{T}{}_{1\times n}\mathbf{y}_{n\times 1} - \frac{\partial}{\partial \mathbf{w}}\mathbf{w}^{T}{}_{1\times 2}\mathbf{X}^{T}{}_{2\times n}\mathbf{y}_{n\times 1} - \frac{\partial}{\partial \mathbf{w}}\mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} + \frac{\partial}{\partial \mathbf{w}}\mathbf{w}^{T}{}_{1\times 2}\mathbf{X}^{T}{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1}\right)$$

$$= \frac{1}{2}\left(-\mathbf{X}^{T}{}_{2\times n}\mathbf{y}_{n\times 1} - \frac{\partial}{\partial \mathbf{w}}\mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} + \frac{\partial}{\partial \mathbf{w}}\mathbf{w}^{T}{}_{1\times 2}\mathbf{X}^{T}{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1}\right)$$

Simplifying partial derivatives of cost function

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} = \begin{bmatrix} \dfrac{\partial}{\partial w_1}\left(\mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1}\right) \\ \dfrac{\partial}{\partial w_2}\left(\mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1}\right) \end{bmatrix}$$

$$\frac{\partial}{\partial w_1}\left(\mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1}\right) = \mathbf{y}^{T}{}_{1\times n}\mathbf{x}_{n\times 1}$$

$$\frac{\partial}{\partial w_2}\left(\mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1}\right) = \mathbf{y}^{T}{}_{1\times n}\mathbf{1}_{n\times 1}$$

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} = \begin{bmatrix} \dfrac{\partial}{\partial w_1}\left(\mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1}\right) \\ \dfrac{\partial}{\partial w_2}\left(\mathbf{y}^{T}{}_{1\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1}\right) \end{bmatrix} = \begin{bmatrix} \mathbf{y}^{T}{}_{1\times n}\mathbf{x}_{n\times 1} \\ \mathbf{y}^{T}{}_{1\times n}\mathbf{1}_{n\times 1} \end{bmatrix} = \mathbf{X}^{T}{}_{2\times n}\mathbf{y}_{n\times 1}$$

$$\frac{\partial J}{\partial \mathbf{w}} = \begin{bmatrix} \dfrac{\partial J}{\partial w_1} \\ \dfrac{\partial J}{\partial w_2} \end{bmatrix} = \frac{1}{2}\left(-\mathbf{X}^T{}_{2\times n}\mathbf{y}_{n\times 1} - \mathbf{X}^T{}_{2\times n}\mathbf{y}_{n\times 1} + \frac{\partial}{\partial \mathbf{w}}\mathbf{w}^T{}_{1\times 2}\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} \right)$$

$$= \frac{1}{2}\left(\frac{\partial}{\partial \mathbf{w}}\mathbf{w}^T{}_{1\times 2}\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} \right) - \mathbf{X}^T{}_{2\times n}\mathbf{y}_{n\times 1}$$

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{w}^T{}_{1\times 2}\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} = \begin{bmatrix} \dfrac{\partial}{\partial w_1}\mathbf{w}^T{}_{1\times 2}\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} \\ \dfrac{\partial}{\partial w_2}\mathbf{w}^T{}_{1\times 2}\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} \end{bmatrix}$$

$$\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times 2} = \begin{bmatrix} \sum_{i=1}^{n} x_i^{\,2} & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i \end{bmatrix}$$

$$\mathbf{w}^T{}_{1\times 2}\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} x_i^{\,2} & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$= \begin{bmatrix} w_1\sum_{i=1}^{n} x_i^{\,2} + w_2\sum_{i=1}^{n} x_i & w_1\sum_{i=1}^{n} x_i + w_2\sum_{i=1}^{n} 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$= w_1^{\,2}\sum_{i=1}^{n} x_i^{\,2} + 2w_1w_2\sum_{i=1}^{n} x_i + w_2\sum_{i=1}^{n} 1$$

$$\frac{\partial}{\partial w_1}\mathbf{w}^T{}_{1\times 2}\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times 2}\mathbf{w}_{2\times 1} = \frac{\partial}{\partial w_1}\left(w_1^{\,2}\sum_{i=1}^{n} x_i^{\,2} + 2w_1w_2\sum_{i=1}^{n} x_i + w_2\sum_{i=1}^{n} 1 \right)$$

$$= 2w_1\sum_{i=1}^{n} x_i^{\,2} + 2w_2\sum_{i=1}^{n} x_i$$

$$= 2\begin{bmatrix} \sum_{i=1}^{n} x_i^{\,2} & \sum_{i=1}^{n} x_i \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$= 2\begin{bmatrix} \sum_{i=1}^{n} x_i^{\,2} & \sum_{i=1}^{n} x_i \end{bmatrix} \mathbf{w}$$

$$\frac{\partial}{\partial w_2}\mathbf{w}^T{}_{1\times2}\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times2}\mathbf{w}_{2\times1} = \frac{\partial}{\partial w_2}\left({w_1}^2\sum_{i=1}^{n}{x_i}^2 + 2w_1w_2\sum_{i=1}^{n}x_i + {w_2}^2\sum_{i=1}^{n}1 \right)$$

$$= 2w_1\sum_{i=1}^{n}x_i + 2w_2\sum_{i=1}^{n}1_i$$

$$= 2\begin{bmatrix} \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}1 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$= 2\begin{bmatrix} \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}1 \end{bmatrix}\mathbf{w}$$

$$\frac{\partial}{\partial \mathbf{w}}\mathbf{w}^T{}_{1\times2}\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times2}\mathbf{w}_{2\times1} = 2\mathbf{X}^T\mathbf{X}\mathbf{w}$$

Finally

$$\frac{\partial J}{\partial \mathbf{w}} = \frac{1}{2}\left(\frac{\partial}{\partial \mathbf{w}}\mathbf{w}^T{}_{1\times2}\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times2}\mathbf{w}_{2\times1}\right) - \mathbf{X}^T{}_{2\times n}\mathbf{y}_{n\times1}$$

$$= \frac{1}{2}\left(2\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times2}\mathbf{w}_{2\times1}\right) - \mathbf{X}^T{}_{2\times n}\mathbf{y}_{n\times1}$$

$$= \mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times2}\mathbf{w}_{2\times1} - \mathbf{X}^T{}_{2\times n}\mathbf{y}_{n\times1} = 0$$

$$\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times2}\mathbf{w}_{2\times1} = \mathbf{X}^T{}_{2\times n}\mathbf{y}_{n\times1}$$

$$\mathbf{w}_{2\times1} = \left(\mathbf{X}^T{}_{2\times n}\mathbf{X}_{n\times2}\right)^{-1}\mathbf{X}^T{}_{2\times n}\mathbf{y}_{n\times1}$$