(2) $h_\theta(x) = \theta^T x = x^T \theta$

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^i) - y^i \right)^2$$

Write $h_\theta(x)$ as ~~the~~ product of design matrix X and parameter vector $\rightarrow$ $h_\theta(x) = X\beta \; (= X\theta)$

e.g $X = \begin{bmatrix} 1 & x_1^T \\ 1 & \vdots \\ \vdots & \vdots \\ 1 & x_n^T \end{bmatrix}$ $\quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \end{bmatrix}$

$$\rightarrow \quad J(\theta) = \frac{1}{2m} (X\theta - y)^T (X\theta - y)$$

$$= \frac{1}{2m} \left[((X\theta)^T - y^T)(X\theta - y) \right]$$

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

$1 + 1 = 2$

$1 + 4 = 5$

$1 +$

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

$$\frac{dJ}{d\theta} = 0$$

$$\Rightarrow \frac{d}{d\theta} \frac{1}{2m} \left[\theta^T X^T X \theta - 2(X\theta)^T y + y^T y \right] = 0$$

$$\rightarrow \frac{d}{d\theta} \theta^T X^T X \theta - 2(X\theta)^T y + y^T y = 0$$

(*) $\theta^T X^T X \theta$: $\begin{bmatrix} \theta_1 & \cdots & \theta_n \end{bmatrix} \underbrace{\begin{bmatrix} x_{11} & \cdots & x_{m1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \cdots & x_{mn} \end{bmatrix}} \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$

$= \theta_1(x_{11} + \cdots x_{m1}) + \theta_2(x_{12} + \cdots x_{m2}) + \cdots + \theta_n(x_{1n} + \cdots x_{mn})$

(~~row~~)

$$= \begin{bmatrix} \theta_1 x_{11}^2 + \cdots + \theta_n x_{1n}^2 \\ \theta_1 x_{21}^2 + \cdots \\ \vdots \\ \theta_1 x_{m1}^2 + \cdots + \theta_n x_{mn}^2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

Note: $X^T X \rightarrow$ square, symmetric $\rightarrow m = n$

$$= \begin{bmatrix} \theta_1^2 x_{11}^2 + \cdots + \theta_n^2 x_{1n}^2 \\ \vdots \\ \theta_1^2 x_{n1}^2 \cdots \theta_n^2 x_{nn}^2 \end{bmatrix}$$

$$\rightarrow \frac{dJ}{d\theta_1} = \frac{d}{d\theta_1} \theta_1^2$$

Compute $X^T X$ first $\qquad X^T X$: square, symmetric $\rightarrow m = n$

$$X^T X = \begin{bmatrix} x_{11}^2 + x_{21}^2 + \cdots & x_{1n}^2 \\ \vdots & \vdots \\ x_{n1}^2 + \cdots & x_{nn}^2 \end{bmatrix}$$

$$X^T X \theta = \begin{bmatrix} \theta_1 x_{11}^2 + \cdots + \theta_n x_{1n}^2 \\ \vdots \\ \theta_1 x_{n1}^2 + \cdots \theta_n x_{nn}^2 \end{bmatrix}$$

$$\theta^T X^T X \theta = \begin{bmatrix} \theta_1 \cdots \theta_n \end{bmatrix} \begin{bmatrix} \theta_1 x_{11}^2 + \cdots + \theta_n x_{1n}^2 \\ \vdots \\ \theta_1 x_{n1}^2 + \cdots + \theta_n x_{nn}^2 \end{bmatrix}$$

$$= \left(\theta_1^2 x_{11}^2 + \theta_2 \theta_1 x_{21}^2 + \cdots + \theta_n \theta_1 x_{n1}^2\right) + \left(\theta_1 \theta_2 x_{12}^2 + \theta_2^2 + x_{22}^2 + \cdots \theta_n \theta_2 x_{n2}^2\right)$$

$$+ \cdots \theta_n \left(\theta_1 x_{n1}^2 + \cdots \theta_n x_{nn}^2\right)$$

$$\frac{dJ'}{d\theta_1} = 2\theta_1 x_{11}^2 + \theta_2 x_{21}^2 + \cdots \theta_n x_{n1}^2 + \cdots + \theta_n x_{1n}^2$$

X^TX Symmetric $\Rightarrow x_{12} = x_{21} \Rightarrow x_{12}^2 = x_{21}^2 \dots$

$$\Rightarrow \frac{dJ'}{d\theta_1} = 2\theta_1 x_{11}^2 + 2\theta_2 x_{12}^2 + \dots \quad 2\theta_n x_{1n}^2$$

$$\frac{dJ'}{d\theta_2} = 2\theta_1 x_{11}^2 + 2\theta_2 x_{12}^2 + \dots \quad 2\theta_n x_{1n}^2$$

$$\Rightarrow \frac{dJ}{d\theta} = 2X^2\theta = 2(X^TX)\theta$$

(2) $2(X\theta)^Ty$

$$= 2\left[\begin{array}{c} m\times n \\ \begin{bmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} \end{bmatrix} \end{array} \begin{array}{c} n\times 1 \\ \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \end{array}\right]^T \begin{array}{c} m\times 1 \\ \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \end{array}$$

$$= \begin{matrix} x_{11}\theta_1 + x_{12}\theta_2 \dots \\ x_{21}\theta_1 + \dots \end{matrix}$$

$\rightarrow 1\times m \times m\times 1 = 1\times 1$

$$= 2 \begin{array}{cc} & m\times 1 \\ \begin{matrix} m \\ \times 1 \end{matrix} & \begin{bmatrix} \theta_1 x_{11} + \dots + \theta_n x_{1n} \\ \vdots \\ \theta_1 x_{m1} + \dots + \theta_n x_{mn} \end{bmatrix}^T \end{array} \begin{array}{c} m\times 1 \\ \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \end{array}$$

$$= \cancel{2\theta_1 x_{11} y_1 + 2\theta_1 x_{21} y_2 + \dots 2\theta_1 x_{m1} y_m}$$

$$= (2\theta_1 x_{11} y_1 + 2\theta_2 x_{12} y_1 + \dots \quad 2\theta_n x_{1n} y_1) + (2\theta_1 x_{21} y_2 + \dots \quad 2\theta_n x_{2n} y_2)$$

$$+ \dots \quad 2\theta_n x_{mn} y_m$$

$$= 2y_1(\theta_1 x_{11} + \theta_2 x_{12} + \dots \quad \theta_n x_{1n}) + \dots \quad 2y_m(\theta_1 x_{m1} + \dots \quad x_{mn}\theta_n)$$

$$\frac{dJ'}{d\theta_1} = 2(y_1 x_{11} + \dots + y_m x_{m1})$$

$$\frac{dJ'}{d\theta_2} = 2(y_1 x_{12} + \dots + y_m x_{m2})$$

$$\frac{dJ'}{d\theta} = \cancel{2X^Ty} \quad 2y \quad \underbrace{2X^T}_{(m\times 1)}y$$

$m \times m$

$m \times n$

$$\Rightarrow \frac{dJ}{d\theta} = 2X^T X\theta - 2X^T y = 0$$

$$\Leftrightarrow \cancel{2X^T(X\theta - y) = 0}$$

$$\Leftrightarrow 2X^T X\theta = 2X^T y$$

$$\Leftrightarrow X^T X\theta = X^T y$$

$$\Leftrightarrow \theta = \frac{X^T y}{X^T X} = \cancel{X^T y (X^T X)^{-1}}$$

$$= \underbrace{(X^T X)^{-1}}_{m \times m} \underbrace{X^T y}_{m \times 1}$$