ith feature of 1st sample

$$\begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{m} \theta_i x_1^{(i)} - y_1 \\ \\ \sum_{i=0}^{m} \theta_i x_n^{(i)} - y_n \end{bmatrix}$$

$\therefore$

$$\begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} \sum \theta_i x_1^{(i)} - y_1 \\ \vdots \\ \sum \theta_i x_n^{(i)} - y_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum \theta_i x_1^{(i)} \\ \vdots \\ \sum \theta_i x_n^{(i)} \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$= \begin{bmatrix} x_1^{(0)} & \cdots & x_1^{(m)} \\ \\ x_n^{(0)} & \cdots & x_n^{(m)} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_m \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$= \bar{X} \bar{\Theta} - \bar{y}$$

Sum of squares of errors

$$\sum_{i=1}^{n} e_i^2 = e_1^2 \cdots e_n^2$$

$$= [e_1 \cdots e_n] \circ [e_1 \cdots e_n]$$

$$= \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}^T \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

$$= (\bar{X}\bar{\Theta} - \bar{y})^T (\bar{X}\bar{\Theta} - \bar{y})$$

$$J(\Theta) = (\bar{X}\bar{\Theta})^T (\bar{X}\bar{\Theta}) - y^T X\bar{\Theta} - (\bar{X}\bar{\Theta})^T \bar{y} - y^T y$$

~~In norm~~ grad desct we reduce $J(\Theta)$ till it reaches min i.e. derivative is 0

$\rightarrow = 0$ $\qquad \rightarrow (X^T y \Theta)^T$

$$\frac{dJ(\Theta)}{dJ(\Theta)} = \frac{\partial(\Theta^T X^T X \Theta)}{\partial\Theta} + \frac{\partial(y^T X\Theta)}{\partial\Theta} - \frac{\partial(\Theta^T X^T y)}{\partial\Theta} - \frac{\partial(y^T y)}{\partial\Theta}$$

$$= 2X^T X\Theta - 2X^T y = 0$$

$$= \cancel{2}X^T X\Theta = \cancel{2}X^T y$$

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

Scanned with CamScanner

We have

$X\theta = y$

Exp We can't directly write $\theta = X^{-1}y$ as it may not be invertible

we multiply w/ $X^T X$

Lim: $X^T X$ may not be invertible

$(X^T X)^{-1} \rightarrow$ expensive computations