# *NORMAL EQUATION DERIVATION*

Derivations of Normal Equation.

- Squared error cost function is:

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

Which can be written in vectorized form:

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

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

simplifying

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

Since $(X\theta)^T$ and $y$ are both vectors, so using their multiplicative property.

$$\because (X\theta)^T y = y^T (X\theta)$$

(2)

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

$$\theta^T X^T$$

$$(\theta) = \frac{1}{2m} \theta^T X^T X\theta - 2(X\theta)^T y + y^T y = 0$$

So, in order to find optimal values of $\theta$, we will have to derivate $J(\theta)$ w.r.t to vector $\theta$ and set that to 0 to find the minima.

Find derivatives of each term in the equation, note the $\frac{1}{2m}$ will be removed since we have set $J(\theta)$ to 0 on the right hand side.

Derivating $2(X\theta)^T y$.

For simplicity assume our input feature matrix $X$ is of dimension $2 \times 2$ as follow

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

$$= 2(X\theta)^T y$$

$$= 2\left(\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}\right)^T \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= 2\left(\begin{bmatrix} x_{11}\theta_1 + x_{12}\theta_2 \\ x_{21}\theta_1 + x_{22}\theta_2 \end{bmatrix}\right)^T \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

taking transform since $X\theta$ is a vector it will turn in $1 \times 2$ matrix.

$$= 2\left(\begin{bmatrix} x_{11}\theta_1 + x_{12}\theta_2 & x_{21}\theta_1 + x_{22}\theta_2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right)$$

lets call the below eq as $P_1(\theta)$

$$P_1(\theta) = 2 \quad y_1(x_{11}\theta_1 + x_{12}\theta_2) + y_2(x_{21}\theta_1 + x_{22}\theta_2)$$

$$\frac{\partial P_1}{\partial \theta_1} = 2(x_{11}y_1 + x_{21}y_2)$$

$$\frac{\partial P_1}{\partial \theta_2} = 2(y_1 x_{12} + x_{22} y_2)$$

The results of derivatives can be written as follows:

$$= 2 \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

→ the above matrix is $X^T$.

$$\text{So } \frac{\partial}{\partial \theta} (X\theta)^T y = 2X^T y$$

---

Taking the derivative of other term

$\theta^T X^T X \theta$, ~~it doesn't matter whether we write $\theta^T X^T$ or $X^T \theta^T$ as long as dimensions workout~~ -

let's multiply $X^T . X$ first

$$X^T = \begin{bmatrix} X_{11} & X_{21} \\ X_{12} & X_{22} \end{bmatrix} \qquad X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$$

$X^T \cdot X \cdot \theta$

$$= \begin{bmatrix} x_{11}^2 + x_{21}^2 & x_{11}x_{12} + x_{21}x_{22} \\ x_{12}x_{11} + x_{22}x_{21} & x_{12}^2 + x_{22}^2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

$X^T \cdot X\theta$

$$= \begin{bmatrix} \theta_1 (x_{11}^2 + x_{21}^2) + \theta_2 (x_{11}x_{12} + x_{21}x_{22}) \\ \theta_1 (x_{12}x_{11} + x_{22}x_{21}) + \theta_2 (x_{12}^2 + x_{22}^2) \end{bmatrix}$$

$\theta^T \cdot X^T \cdot X\theta$

$$= [\theta_1 \quad \theta_2] \begin{bmatrix} \theta_1 (x_{11}^2 + x_{21}^2) + \theta_2 (x_{11}x_{12} + x_{21}x_{22}) \\ \theta_1 (x_{12}x_{11} + x_{22}x_{21}) + \theta_2 (x_{12}^2 + x_{22}^2) \end{bmatrix}$$

$$= [\theta_1^2 (x_{11}^2 + x_{21}^2) + \theta_1\theta_2 (x_{11}x_{12} + x_{21}x_{22})$$
$$+ \theta_2\theta_1 (x_{12}x_{11} + x_{22}x_{21}) + \theta_2^2 (x_{12}^2 + x_{22}^2)]$$

lets call the above function $Z(\theta)$

$$Z(\theta) = [\theta_1^2 (x_{11}^2 + x_{21}^2) + 2\theta_1\theta_2 (x_{11}x_{12} + x_{21}x$$

$$+ \theta_2^2 (x_{12}^2 + x_{22}^2)]$$

$$\frac{\partial Z}{\partial \theta_1} = [2\theta_1 (x_{11}^2 + x_{21}^2) + 2\theta_2 (x_{11}x_{12} + x_{21}x$$

$$\frac{\partial Z}{\partial \theta_2} = [2\theta_1 (x_{11}x_{12} + x_{21}x_{22}) + 2\theta_2 (x_{12}^2 + x$$

$$= \begin{bmatrix} 2(x_{11}^2 + x_{21}^2) & 2(x_{11}x_{12} + x_{21}x_{22}) \\ 2(x_{11}x_{12} + x_{21}x_{22}) & 2(x_{12}^2 + x_{22}^2) \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

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

$$\frac{\partial}{\partial \theta} \theta^T X^T X\theta = 2X^T X\theta$$

$$\frac{\partial}{\partial \theta} y^T y = 0$$

Combining all:

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

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

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

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

Normal Equation.