The hypothesis is defined as:

$$h_\theta(X) = X \cdot \theta$$

The cost function is defined as:

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

$$\Rightarrow J(\theta) = \frac{1}{2m} \sum_{i=1}^{n} \left(X \cdot \theta - Y \right)^2$$

We know that $X \cdot \theta - Y = e$, with $e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$ is the error vector.

now let try to find this term $\sum_{i=1}^{n} (X \cdot \theta - Y)^2$

$$\sum_{i=1}^{n} (X \cdot \theta - Y)^2 = e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2$$

and we've $\quad e^T \cdot e = (e_1, e_2, \dots, e_n) \times \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$

$$\Rightarrow e^T \cdot e = e_1^2 + e_2^2 + \dots + e_n^2$$

then $\quad \sum_{i=1}^{n} (X \cdot \theta - Y)^2 = e^T \cdot e = (X \cdot \theta - Y)^T \times (X \cdot \theta - Y)$

Now our cost function will look like this:

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

Ignore the term $\frac{1}{2m}$ since we're going to compare the derivative to zero.

$$J(\theta) = (X \cdot \theta - Y)^T \times (X \cdot \theta - Y)$$

Made by: Anas BRITAL

$$\Rightarrow J(\theta) = ((X\theta)^T - y^T)(X\theta - y)$$

$$\Rightarrow J(\theta) = (X\theta)^T X\theta - (X\theta)^T y - y^T(X\theta) + y^T y$$

$$\Rightarrow J(\theta) = \theta^T X^T X\theta - 2(X\theta)^T y + y^T y$$

the derivative of the cost function is :

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

The minimum of $J(\theta)$ correspond to where the derivative $\frac{\partial J(\theta)}{\partial \theta} = 0$.

$$\Rightarrow \frac{\partial J(\theta)}{\partial \theta} = 0 \Rightarrow 2X^T X\theta - 2X^T y = 0$$

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

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

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

Made by : Anas BRITAL