Derivation

We Know that Normal equation for logistic regression is

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

where θ is the vector of parameters we want to ~~exis~~ estimate

X: Input feature value of each instance

Y: Output value of each instance.

We know hypothesis equation:-

$$h(\theta) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

$$\Rightarrow h(\theta) = \theta^T \cdot x = h_\theta(x) \text{ (Representation)}$$

$$\theta^T = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 & \cdots & \theta_n \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Cost function:-

$$J(\theta) = \frac{1}{2m} (h_\theta(x) - Y)^T (h_\theta(x) - Y)$$

also

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

x^i: the Input variable i^{th} training Exampl

m: no of training samples

y^i: expected result of i^{th} instance.

cost function in vector form-

$$\begin{bmatrix} h_\theta(x^0) \\ h_\theta(x^1) \\ \cdots \\ h_\theta(x^m) \end{bmatrix} - \begin{bmatrix} y^0 \\ y^1 \\ \cdots \\ y^m \end{bmatrix}$$

$$= \begin{bmatrix} \theta^T(x^0) \\ \theta^T(x^1) \\ \cdots \\ \theta^T(x^m) \end{bmatrix} - y$$

we know

$$\theta^T(x^{(i)}) = \theta_0(x_0^i) + \theta_1(x_1^i) + \cdots \theta_n(x_n^i)$$

$$= \begin{bmatrix} \theta_0(x_0^0) + \theta_1(x_1^0) + \cdots \theta_n(x_n^0) \\ \theta_0(x_0^1) + \theta_1(x_1^1) + \cdots \theta_n(x_n^1) \\ \theta_0(x_0^m) + \theta_1(x_1^m) + \cdots \theta_n(x_n^m) \end{bmatrix} - y$$

$x_j^i \rightarrow j^{th}$ feature in i^{th} training sample

we can write computed matrix as
$(X\theta - y)$

Hence, cost $= (X\theta - y)^T (X\theta - y) = J$

motive is to minimize cost

$$\frac{\partial J}{\partial \theta} = \frac{\partial}{\partial \theta}\left[(X\theta - y)^T (X\theta - y)\right]$$

$\Rightarrow$ θ is the variable

Applying product variable ~~with~~
we get

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

$$\Rightarrow \frac{\partial J}{\partial \theta} = X^T X\theta - X^T y + XX^T\theta - Xy^T$$

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

as X is square matrix)

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

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

Pre multiply $(X^T X)^{-1}$ on both side
we get
$$(X^T X)^{-1} (X^T X)\theta = (X^T X)^{-1} (X^T y)$$

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