A seguir, vamos derivar da derivada parcial da função de custo $J(\theta)$ da regressão logística com relação aos parâmetros θ_j. Para isso considere que

$$\theta^T x^{(i)} := \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n^{(i)}.$$

Sendo assim,

$$\log\left(h_\theta(x^{(i)})\right) = \log \frac{1}{1+e^{-\theta^T x^{(i)}}} = -\log(1+e^{-\theta^T x^{(i)}}), \tag{1}$$

e

$$\begin{aligned} \log\left(1-h_\theta(x^{(i)})\right) &= \log\left[1-\frac{1}{1+e^{-\theta^T x^{(i)}}}\right] \\ &= \log(e^{-\theta^T x^{(i)}}) - \log(1+e^{-\theta^T x^{(i)}}) \\ &= -\theta^T x^{(i)} - \log(1+e^{-\theta^T x^{(i)}}). \end{aligned} \tag{2}$$

A passagem acima usou a identidade a seguir: $1 = \frac{1+e^{-\theta^T x^{(i)}}}{1+e^{-\theta^T x^{(i)}}}$. Os 1's no numerador se cancelam. A seguir, usamos a identidade $\log(x/y) = \log(x) - \log(y)$.

A função de custo é da forma:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] \tag{3}$$

Ao substituirmos 1 e 2 em 3, obtemos:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}(-y^{(i)}(\log(1+e^{-\theta^T x^{(i)}})) + (1-y^{(i)})(-\theta^T x^{(i)} - \log(1+e^{-\theta^T x^{(i)}})))$$

A expressão acima pode ser simplificada para:

$$\begin{aligned} J(\theta) &= -\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}\theta^T x^{(i)} - \theta^T x^{(i)} - \log(1+e^{-\theta^T x^{(i)}})) \\ &= -\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}\theta^T x^{(i)} - \log(1+e^{\theta^T x^{(i)}})] \end{aligned} \tag{4}$$

Em 4, a segunda igualdade segue de

$$-\theta^T x^{(i)} - \log(1 + e^{-\theta x^{(i)}}) = -[\log e^{\theta x^{(i)}} + \log(1 + e^{-\theta x^{(i)}})] = -\log(1 + e^{\theta x^{(i)}}).$$

(i.e., usamos a identidade $\log(x) + \log(y) = \log(xy)$).

Tudo que resta agora é computar as derivadas parciais de $J(\theta)$ (conforme apresentada em 4) com relação a θ_j:

$$\frac{\partial}{\partial\theta j} y^{(i)}\theta x^{(i)} = y^{(i)} x_j^{(i)} \tag{5}$$

$$\frac{\partial}{\partial\theta j} \log(1 + e^{\theta x^{(i)}}) = \frac{x_j^{(i)} e^{\theta x^{(i)}}}{1 + e^{\theta x^{(i)}}} = x_j^{(i)} h_\theta(x^{(i)}) \tag{6}$$

Sendo assim, de 5 e de 6, segue que:

$$\frac{\partial}{\partial\theta_j} J(\theta) = y^{(i)} x_j^{(i)} + x_j^{(i)} h_\theta(x^{(i)}) \tag{7}$$

Ao fatorarmos x_j em 7, obtemos finalmente:

$$\frac{\partial}{\partial\theta_j} J(\theta) = (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$