Combinar em uma única função de custo?

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$\mathrm{Cost}(h_\theta(x), y) = \begin{cases} \boxed{-\log(h_\theta(x))} & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

Note: $y = \boxed{0}$ or $\boxed{1}$ always

$$\mathrm{Cost}(h_\theta(x), y) = -\underbrace{y \log(h_\theta(x))}_{=0} - \underbrace{(1-y)}_{=0} \underbrace{\log(1 - h_\theta(x))}_{=1}$$

If $y=1$: $\mathrm{Cost}(h_\theta(x), y) = -\log h_\theta(x)$

If $y=0$: $\mathrm{Cost}(h_\theta(x), y) = -\log(1 - h_\theta(x))$

Gradiente Descendente

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

Want $\min_\theta J(\theta)$:

Repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

(simultaneously update all θ_j)

}

Gradiente Descendente

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

Want $\min_\theta J(\theta)$:

Repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

(simultaneously update all θ_j)

}

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Gradiente Descendente

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

Want $\min_\theta J(\theta)$:

$$\Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$

Repeat {

$$\boxed{\theta_j := \theta_j - \alpha \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}}$$

(simultaneously update all θ_j)

}

É a mesma regra!

Gradiente Descendente

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

$$\Theta = \begin{bmatrix} \Theta_0 \\ \Theta_1 \\ \Theta_2 \\ \vdots \\ \Theta_n \end{bmatrix}$$

Want $\min_\theta J(\theta)$:

Repeat {

$$\boxed{\theta_j := \theta_j - \alpha \sum_{i=1}^{m} (\boxed{h_\theta(x^{(i)})} - y^{(i)}) x_j^{(i)}}$$

(simultaneously update all θ_j)

}

$h_\Theta(x) = \Theta^T x$

$h_\Theta(x) = \frac{1}{1+e^{-\Theta^T x}}$

É a mesma regra!

Derivando a Função de Custo

$$\frac{\partial}{\partial\theta} J(\theta) = \frac{-1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial\theta} \Big[y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})) \Big],$$

$$h_\theta(x) = \frac{1}{1+e^{\theta^T x^{(i)}}}$$

Derivando a Função de Custo

$$\frac{\partial}{\partial\theta}J(\theta) = \frac{-1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial\theta}\Big[y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\Big],$$

$$h_\theta(x) = \frac{1}{1+e^{\theta^T x^{(i)}}}$$

Podemos escrever a função na forma:

$$kf(s(g(\theta))) + (1-k)f(t(s(g(\theta))))$$

tal que

$$f(x) = \log(x);\ f'(x) = \frac{1}{x}$$
$$s(x) = \frac{1}{1+e^{-x}};\ s'(x) = \frac{e^x}{(e^x+1)^2}$$
$$t(x) = 1-x;\ t'(x) = -1$$
$$g(\theta) = \theta^T x;\ g'(\theta) = x$$

Derivando a Função de Custo

Derivando

$$kf(s(g(\theta))) + (1-k)f(t(s(g(\theta))))$$

leva a

$$kf'(s(g(\boldsymbol{\theta}))s'(g(\boldsymbol{\theta}))g'(\boldsymbol{\theta}) + (1-k)f'(t(s(g(\boldsymbol{\theta}))))t'(s(g(\boldsymbol{\theta})))s'(g(\boldsymbol{\theta}))g'(\boldsymbol{\theta})$$

Derivando a Função de Custo

Fazendo $g_1 = e^{g(\theta)}$

teremos

$$k\frac{g_1+1}{g_1}\frac{g_1}{(g_1+1)^2}g'(\boldsymbol{\theta}) + (1-k)(g_1+1)(-1)\frac{g_1}{(g_1+1)^2}g'(\boldsymbol{\theta})$$

Derivando a Função de Custo

Simplificando teremos

$$k\frac{1}{g_1+1}g'(\boldsymbol{\theta}) + (k-1)\frac{g_1}{g_1+1}g'(\boldsymbol{\theta}) = \frac{k+kg_1-g_1}{g_1+1}g'(\boldsymbol{\theta}) = \left(k - \frac{g_1}{g_1+1}\right)g'(\boldsymbol{\theta})$$

Derivando a Função de Custo

Substituindo g_1 teremos

$$\left(k - \frac{1}{1 + \frac{1}{g_1}}\right) g'(\boldsymbol{\theta}) = (k - h_\theta(x))g'(\boldsymbol{\theta})$$

Derivando a Função de Custo

Por fim chegamos a

$$\frac{\partial}{\partial\theta_j}J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} - h_\theta(x)\right]x_j^{(i)} = \frac{1}{m}\sum_{i=1}^{m}\left[h_\theta(x) - y^{(i)}\right]x_j^{(i)}$$