ERRORE DI COSTO MEDIO $\forall x$ in $\vec{X}$

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i \ln\left(h_\theta(x_i)\right) + (1-y_i)\ln\left(1-h_\theta(x_i)\right)\right]$$

DISCESA DEL GRADIENTE

$$\boxed{\theta_j = \theta_j - \eta \cdot \frac{\partial J(\theta)}{\partial \theta_j}}$$

CALCOLO DERIVATE

Calcoliamo prima $\frac{\partial h(\theta)}{\partial \theta}$ ← servirà poi per calcolare $\partial J(\theta)/\partial\theta_j$

LOGISTIC FUNCTION

$$\sigma(x)' = \left[\frac{1}{1+e^{-x}}\right]' = \frac{-[1+e^{-x}]'}{(1+e^{-x})^2} = -\frac{[1]' + [e^{-x}]'}{(1+e^{-x})^2} =$$

$$= \frac{-[e^{-x}]'}{(1+e^{-x})^2} = \frac{-[g(f(x))]'}{(1+e^{-x})^2} = \frac{-[f(x)]'\, g'(f(x))}{(1+e^{-x})^2} =$$

$t = -x$

$$= \frac{-[-x]' \cdot [e^t]'}{(1+e^{-x})^2} = \frac{-(-1)\cdot e^t}{(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})^2} =$$

$$\frac{e^{-x}}{(1+e^{-x})(1+e^{-x})} = \underbrace{\boxed{\frac{1}{(1+e^{-x})}}}_{\sigma(x)} \cdot \frac{e^{-x}}{(1+e^{-x})} =$$

$$= \sigma(x)\frac{e^{-x}}{(1+e^{-x})} = \sigma(x)\left[\frac{1+e^{-x}-1}{(1+e^{-x})}\right] =$$

$$= \sigma(x)\left[\underbrace{\boxed{\frac{1+e^{-x}}{1+e^{-x}}}}_{1} - \underbrace{\boxed{\frac{1}{1+e^{-x}}}}_{\sigma(x)}\right] = \sigma(x)\left(1-\sigma(x)\right)$$

DA CUI

$$\boxed{\frac{\partial \sigma(x)}{\partial x} = \sigma(x)\left(1-\sigma(x)\right)}$$

DA CUI

$$\frac{\partial \sigma(x)}{\partial x} = \sigma(x)\left(1 - \sigma(x)\right)$$

DERIVATA PRIMA SIGMOIDE

DA QUI POSSIAMO CALCOLARE LA DERIVATA PRIMA DELLA FUNZIONE D'ERRORE

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j}\left(-\frac{1}{m}\sum_{i=1}^{m}\left[y_i \ln\left(h_\theta(x_i)\right) + (1-y_i)\ln\left(1-h_\theta(x_i)\right)\right]\right) =$$

$$= -\frac{1}{m}\sum \frac{\partial}{\partial \theta_j}\left(y_i \ln\left(h_\theta(x_i)\right) + (1-y_i)\ln\left(1-h_\theta(x_i)\right)\right)$$

$$= -\frac{1}{m}\sum\left[\frac{\partial}{\partial \theta_j} y_i \ln\left(h_\theta(x_i)\right) + \frac{\partial}{\partial \theta_j}(1-y_i)\ln\left(1-h_\theta(x_i)\right)\right]$$

$$= -\frac{1}{m}\sum\left[y_i \frac{\partial}{\partial \theta_j}\ln\left(h_\theta(x_i)\right) + (1-y_i)\frac{\partial}{\partial \theta_j}\ln\left(1-h_\theta(x_i)\right)\right]$$

$t = h_\theta(x_i)$ $\quad w = 1 - h_\theta(x_i)$

$$= -\frac{1}{m}\sum\left[y_i \frac{\partial}{\partial \theta_j}\ln(t) + (1-y_i)\frac{\partial}{\partial \theta_j}\ln(w)\right] =$$

$$= -\frac{1}{m}\sum\left[y_i \frac{\partial \ln(t)}{\partial t}\cdot\frac{\partial t}{\partial \theta_j} + (1-y_i)\frac{\partial \ln(w)}{\partial w}\cdot\frac{\partial(t)}{\partial \theta_j}\right] =$$

$$= -\frac{1}{m}\sum\left[y_i \frac{1}{h_\theta(x_i)}\cdot\frac{\partial h_\theta(x_i)}{\partial \theta_j} + (1-y_i)\frac{1}{(1-h_\theta(x_i))}\cdot\frac{\partial(1-h_\theta(x_i))}{\partial \theta_j}\right]$$

$$\boxed{\frac{\partial(1)}{\partial \theta_j} - \frac{\partial(h_\theta(x_i))}{\partial \theta_j}}$$

$$= -\frac{1}{m}\sum\left[y_i \frac{1}{h_\theta(x_i)}\cdot\frac{\partial h_\theta(x_i)}{\partial \theta_j} + (1-y_i)\frac{1}{(1-h_\theta(x_i))}\left(-\frac{\partial h_\theta(x_i)}{\partial \theta_j}\right)\right]$$

$$= -\frac{1}{m}\sum\left[(y_i)\frac{\frac{\partial h_\theta(x_i)}{\partial \theta_j}}{h_\theta(x_i)} - (1-y_i)\frac{\frac{\partial h_\theta(x_i)}{\partial \theta_j}}{(1-h_\theta(x_i))}\right] =$$

IN FORMA VETTORIALE $h_\theta(x_i) = \sigma(\theta^T x_i)$

$$= -\frac{1}{m}\sum\left[y_i \frac{\frac{\partial \sigma(\theta^T x_i)}{\partial \theta_j}}{h_\theta(x_i)} - (1-y_i)\frac{\frac{\partial \sigma(\theta^T x_i)}{\partial \theta_j}}{(1-h_\theta(x_i))}\right] =$$

$$\frac{\partial g(f(t))}{\partial t} = \frac{\partial g(f(t))}{\partial f(t)}\cdot\frac{\partial f(t)}{\partial t}$$

$$= -\frac{1}{m}\sum\left[y_i \frac{\frac{\partial \sigma(\theta^T x_i)}{\partial \theta^T x_i}\frac{\partial \theta^T x_i}{\partial \theta_j}}{h_\theta(x_i)} - (1-y_i)\frac{\frac{\partial \sigma(\theta^T x_i)}{\partial \theta^T x_i}\cdot\frac{\partial \theta^T x_i}{\partial \theta_j}}{1-h_\theta(x_i)}\right] =$$

SAPPIAMO CHE $\frac{\partial \sigma(t)}{\partial t} = \sigma(t)\left(1-\sigma(t)\right)$

DA CUI

$$= -\frac{1}{m}\sum\left[y_i\,\sigma(\theta^T x_i)\left(1-\sigma(\theta^T x_i)\right)\frac{\partial \theta^T x_i}{\partial \theta_j} - (1-y_i)\,\sigma(\theta^T x_i)\left(1-\sigma(\theta^T x_i)\right)\frac{\partial \theta^T x_i}{\partial \theta_j}\right]$$

Inoltre possiamo dire che

$$\frac{\partial(\vec{\theta}^T \vec{x}_i)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j}\left[\theta_0 x_0 + \dots \theta_j x_j + \dots + \theta_m x_m\right] =$$

$$= \cancel{\frac{\partial}{\partial \theta_j}\theta_0 x_0^{(i)}}^{\,0} + \dots + \cancel{\frac{\partial}{\partial \theta_j}\theta_j x_j^{(i)}}^{\,x_j^{(i)}} + \dots + \cancel{\frac{\partial}{\partial \theta_j}\theta_m x_m^{(i)}}^{\,0} =$$

DA CUI

$$\frac{\partial \theta^T x_i}{\partial \theta_j} = x_j^{(i)}$$

NOTA
$\vec{\theta}^T$ è il vettore dei parametri
$\vec{x}_i$ è il vettore che rappresenta le I features del dato i-esimo

Da cui abbiamo

$$= -\frac{1}{m}\sum\left[y_i \frac{(\theta^T x_i)(1-\theta^T x_i)}{h_\theta(x_i)}\,x_j^{(i)} - (1-y_i)\frac{(\theta^T x_i)(1-\theta^T(x_i))}{1-h_\theta(x_i)}\,x_j^{(i)}\right] =$$

$$= -\frac{1}{m}\sum\left[\frac{y_i\left(h_\theta(x_i)\left(1-h_\theta(x_i)\right)\right)}{h_\theta(x_i)}x_j^{(i)} - (1-y_i)\frac{h_\theta(x_i)\left(1-h_\theta(x_i)\right)}{1-h_\theta(x_i)}x_j^{(i)}\right] =$$

$$= -\frac{1}{m}\sum\left[y_i x_j^{(i)}\left(1-h_\theta(x_i)\right) - (1-y_i)x_j^{(i)} h_\theta(x_i)\right] =$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left\{\left[y_i\left(1-h_\theta(x_i)\right) - (1-y_i)\cdot h_\theta(x_i)\right]x_j^{(i)}\right\}$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left\{\left[y_i - \cancel{y_i h_\theta(x_i)} - h_\theta(x_i) + \cancel{y_i h_\theta(x_i)}\right]x_j^{(i)}\right\}$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left[\left(y_i - h_\theta(x_i)\right)x_j^{(i)}\right]$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left[h_\theta(x_i) - y_i\right]x_j^{(i)}$$

:)

IN FORMA VETTORIALE

$$\nabla J(\theta) = \frac{1}{m}\left[X^T \cdot g(\theta X) - Y\right]$$

OGNI RIGA HA $x_j^{(i)}$ con $1 \geq i \geq$

DOVE

$$X^T = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_f^{(1)} \\ & & \vdots & \\ x_1^{(m)} & x_2^{(m)} & \dots & x_f^{(m)} \end{bmatrix}^T = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_m^{(f)} \\ & & \vdots & \\ x_m^{(1)} & x_m^{(2)} & \dots & x_m^{(f)} \end{bmatrix}$$

m dati in $\|X\|$.
ogni dato ha f features $\quad \|X\| = m \times f$, $\|x^{(i)}\| = f$

DA CUI ABBIAMO CHE

$$\Theta_j = \Theta_j - \eta \frac{\partial}{\partial \Theta_j} J(\Theta)$$

sigmoide

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

DA CUI

$$\Theta_j = \Theta_j - \frac{\eta}{m} \sum_{i=1}^{m} \left[\left(h_\Theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)} \right] \quad \forall j \in [0, n]$$

numero features

con $h_\Theta(x^{(i)}) = \dfrac{1}{1 + e^{-(\Theta_0^{(i)} + \Theta_1^{(i)} x_1^{(i)} + \dots + \Theta_j^{(i)} x_j^{(i)} + \dots \Theta_f^{(i)} x_f^{(i)})}}$

- dove in $x_j^{(i)}$
  - $x^{(i)}$ rappresenta il vettore $\vec{x}^{(i)}$ composto delle f features di x con $1 \leq i \leq m$
    $\vec{x}^{(i)} = [x_0^{(i)}, x_1^{(i)}, \dots, x_j^{(i)}, \dots, x_f^{(i)}]$
  - $x_j^{(i)}$ rappresenta il singolo elemento j delle f features in $x^{(i)}$ $1 \leq j \leq f$

- $y^{(i)}$ rappresenta il valore dell'etichetta di output associata al dato i-esimo con $0 \leq i \leq m$

- Θ_j rappresenta il coefficiente di $x_j^{(i)}$ con $1 \leq j \leq f$ e $1 \leq i \leq m$