end for

$$\begin{cases} D_{ij}^{(l)} := \frac{1}{m}\Delta_{ij}^{(l)} + \lambda\Theta_{ij}^{(l)}, \text{ if } j \neq 0 \\ D_{ij}^{(l)} := \frac{1}{m}\Delta_{ij}^{(l)} \text{ if } j = 0 \end{cases}$$

sem regularização na bias unit

Then, finally:

$$\boxed{\frac{\partial}{\partial\Theta_{ij}^{(l)}} J(\Theta) = D_{ij}^{(l)}}$$

Cajamar, 11 de novembro de 2018

Análise dimensional do algoritmo de backpropagation

$\delta_j^{(4)} = a_j^{(4)} - y_j$ → node j layer l

$\delta^{(3)} = (\Theta^{(3)})^T \cdot \delta^{(4)} .* g'(z^{(3)})$ deriving w.r.t. $z^{(3)}$

$z^{(3)} = \Theta^{(2)} \cdot a^{(2)}$

$\delta^{(2)} = (\Theta^{(2)})^T \cdot \delta^{(3)} .* g'(z^{(2)})$

↳ element-wise multiplication

Temos: $a^{(4)}_{4\times5000}$ $\quad y_{5000\times4}$ ← # classes

Então: $\boxed{\delta^{(4)}_{4\times5000} = a^{(4)} - y^T}$

Também temos: $\cancel{\Theta^{(3)}_{4\times7}}$

OBS: não calculamos δ com a bias unit, então $\Theta^{(3)}_{4\times6}$

Por isso: $(\Theta^{(3)})^T_{6\times4} \cdot \delta^{(4)}_{4\times5000} \rightarrow M_{6\times5000}$

na implementação a 1ª coluna de $\Theta^{(3)}_{4\times7}$ não deverá ser percorrida, pois se refere à bias unit

E como $g(z^{(3)}) = \frac{1}{1+e^{-\Theta^{(2)}a^{(2)}}}$

Segue que:

$$g'(z^{(3)}) = \frac{d}{dz^{(3)}} g(z^{(3)}) = \frac{d}{d(\Theta^{(2)}a^{(2)})} g(\Theta^{(2)}a^{(2)}) = \frac{-(-1)\cdot e^{-\Theta^{(2)}a^{(2)}}}{(1+e^{-\Theta^{(2)}a^{(2)}})^2}$$

$$\Rightarrow g'(z^{(3)}) = \frac{1 \cdot e^{-\Theta^{(2)}a^{(2)}}}{(1+e^{-\Theta^{(2)}a^{(2)}})^2} = \varphi$$

Se desenvolvermos: $a^{(3)} \cdot (1-a^{(3)})$, lembrando que $a^{(3)} = \frac{1}{1+e^{-\Theta^{(2)}a^{(2)}}}$, teremos:

$$\frac{1}{1+e^{-\Theta^{(2)}a^{(2)}}} \cdot \left(1 - \frac{1}{1+e^{-\Theta^{(2)}a^{(2)}}}\right) = \frac{1}{1+e^{-\Theta^{(2)}a^{(2)}}} \cdot \left(\frac{1+e^{-\Theta^{(2)}a^{(2)}} - 1}{1+e^{-\Theta^{(2)}a^{(2)}}}\right)$$