Logo: $\boxed{\delta^{(2)}_{4\times 5000}}$

E, como não existirá $\delta^{(1)}$, terminamos os cálculos de $\delta^{(2)}$.

OBS4: O Prof. Andrew Ng disse que ele calcula os $\delta_0^{(l)}$ terms (das bias units) mas que ele descarta esses valores, uma vez que não serão usados para os cálculos das derivadas.

OU SEJA, resumindo, uma forma mais eficiente de implementar a backpropagation seria:

$$\underbrace{\delta^{(4)}}_{4\times5000} = \underbrace{a^{(4)}}_{4\times5000} - \underbrace{y^{T}}_{4\times5000} ;$$

a3 c/ linha de bias

$$\underbrace{\delta^{(3)}}_{7\times5000} = \underbrace{(\Theta^{(3)})^{T}}_{7\times4} . \underbrace{\delta^{(4)}}_{4\times5000} .* \underbrace{a^{(3)}}_{7\times5000} .* \underbrace{(1-a^{(3)})}_{7\times5000}$$

// descarta a linha de $\delta_0^{(3)}$

$\delta^{(3)}(1,:) = [\,]$;

// agora $\delta^{(3)}_{6\times5000}$

a2 c/ linha de bias

$$\underbrace{\delta^{(2)}}_{5\times5000} = \underbrace{(\Theta^{(2)})^{T}}_{5\times6} . \underbrace{\delta^{(3)}}_{6\times5000} .* \underbrace{a^{(2)}}_{5\times5000} .* \underbrace{(1-a^{(2)})}_{5\times5000}$$

// descarta linha $\delta_0^{(2)}$

$\delta^{(2)}(1,:) = [\,]$;

// agora $\delta^{(2)}_{4\times5000}$

Perceba também que seria conveniente criar uma matriz com 3 dimensões para armazenar $\delta^{(2)}, \delta^{(3)}, \delta^{(4)}$.
Dessa forma, podemos criar uma matriz deltinha(:,:,3), assim:

~~(deltinha) = max(s(l))~~