1 Back Propagation algorithm

Algorithm 1 Back Propagation

set $\Delta_{ij}^{(l)} = 0$ (for all i, j, l)
for i = 1 to m **do**
  set $a^{(1)}$ to $x^{(i)}$
  perform forward propagation to compute $a^{(l)}$ for $l = 2, 3...L$
  using $y^{(i)}$ compute $\delta^{(L)} = a^{(L)} - y^{(i)}$
  compute $\delta^{(i)}$ for $i = L-1, L-2, .., 2$
  $\Delta_{ij}^{(l)} := \Delta_{ij}^{(l)} + a_j^{(l)} \cdot \delta_j^{(l+1)}$
  or equivalently $\Delta^{(l)} := \Delta^{(l)} + \delta^{(l+1)} \cdot (a^{(l)})^T$
end for
$D_{ij}^{(l)} := \frac{1}{m}\Delta_{ij}^{(l)} + \lambda\Theta_{ij}^{(l)}$ if $j \neq 0$
$D_{ij}^{(l)} := \frac{1}{m}\Delta_{ij}^{(l)}$ otherwise

It is possible to prove that

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