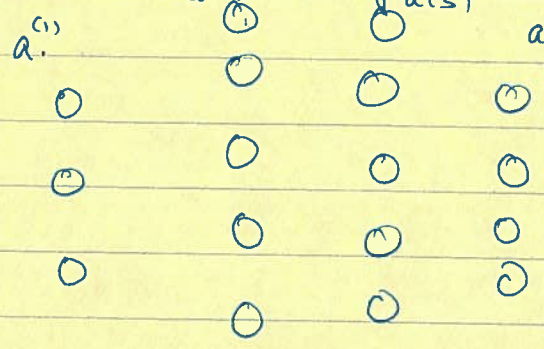
26 × 1

$a1 = 1 \times 401$

NN – Non linear hypothesis $z2 = 25 \times 401$

Forward Prop



$X = 5000 \times 401$

Theta1 $= 25 \times 401$

Theta2 $= 10 \times 26$

$a1 = 5000 \times 401$

Theta1' $= 401 \times 25$

$a2 = 25 \times 5000$

$a^{(1)} = x$

$z^{(2)} = \Theta^{(1)} a^{(1)}$ → bias Unit

$a^{(2)} = g(z^{(2)})$ add $a_0^{(2)}$ sigmoid $z^{(2)}$

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

$a^{(3)} = g(z^3)$ add $a_0^{(3)}$ → biag Unit

$z^{(4)} = \Theta^{(3)} a^{(3)}$

$a^{(4)} = h_\Theta(x) = g(z^{(4)})$

Back Propagation

$\delta_j^{(l)}$ = "error" of node j in layer l

For each output unit (layer $L=4$)

$\delta_j^{(4)} = a_j^{(4)} - y_j$ $\quad (h_\Theta(x))_j$ $\quad \delta^{(4)} = a^{(4)} - y$

$\delta^{(3)} = (\Theta^{(2)})^T \delta^{(4)} .* g'(z^{(3)})$ $\quad a^{(3)} .* (1 - a^{(3)})$

$\delta^{(2)} = (\Theta^{(2)})^T \delta^{(3)} .* g'(z^{(2)})$ $\quad a^{(2)} .* (1 - a^{(2)})$

$\frac{\partial}{\partial \Theta_{ij}^{(l)}} J(\Theta) = a_j^{(l)} \delta_i^{(l+1)}$ Ignore λ if $\lambda = 0$

Backpropagation Algorithm

Set $\Delta_{ij}^{(l)} = 0$ (for all l, i, j)

For $i = 1:m$

Set $a^{(1)} = x^{(i)}$

Perform forward for $l = 2, 3, \ldots, L$

Using $y^{(i)}$ compute $\delta^{(L)} = a^{(L)} - y^{(i)}$

Compute $\delta^{(L-1)}, \delta^{(L-2)}, \ldots, \delta^{(2)}$

$$\Delta_{ij}^{(l)} := \Delta_{ij}^{(l)} + a_j^{(l)} \delta_i^{(l+1)}$$

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

$$D_{ij}^{(l)} := \frac{1}{m} \Delta_{ij}^{(l)} \quad \text{if } j = 0$$

$$\frac{d}{d\theta} J(\theta) \approx \frac{J(\theta + \varepsilon) - J(\theta - \varepsilon)}{2\varepsilon}$$

$\varepsilon = 10^{-4}$