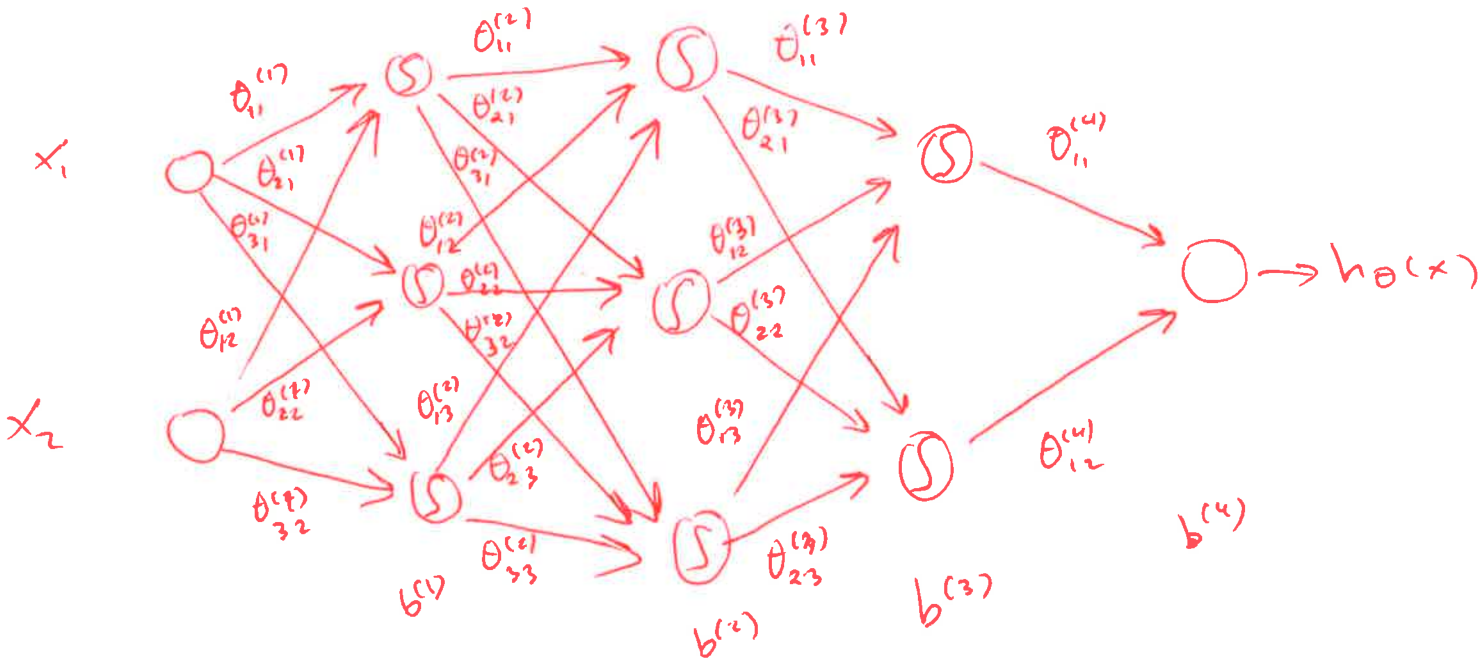
Final!



$$\theta^{(1)} = \begin{bmatrix} \theta_{11}^{(1)} & \theta_{12}^{(1)} \\ \theta_{21}^{(1)} & \theta_{22}^{(1)} \\ \theta_{31}^{(1)} & \theta_{32}^{(1)} \end{bmatrix} \quad b^{(1)} = \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ b_3^{(1)} \end{bmatrix}$$

3×2 $\qquad$ 3×1

$$\theta^{(2)} \begin{bmatrix} \theta_{11}^{(2)} & \theta_{12}^{(2)} & \theta_{13}^{(2)} \\ \theta_{21}^{(2)} & \theta_{22}^{(2)} & \theta_{23}^{(2)} \\ \theta_{31}^{(2)} & \theta_{32}^{(2)} & \theta_{33}^{(2)} \end{bmatrix} \quad b^{(2)} = \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \end{bmatrix}$$

3×3 $\qquad$ 3×1

$$\theta^{(3)} = \begin{bmatrix} \theta_{11}^{(3)} & \theta_{12}^{(3)} & \theta_{13}^{(3)} \\ \theta_{21}^{(3)} & \theta_{22}^{(3)} & \theta_{23}^{(3)} \end{bmatrix} \quad b^{(3)} = \begin{bmatrix} b_1^{(3)} \\ b_2^{(3)} \end{bmatrix}$$

2×3 $\qquad$ 2×1

$$\theta^{(4)} = \begin{bmatrix} \theta_{11}^{(4)} & \theta_{12}^{(4)} \end{bmatrix} \quad b^{(4)} = \begin{bmatrix} b_1^{(4)} \end{bmatrix}$$

$$h^{(1)} = x \qquad 2\times 1$$

$$h^{(2)} = g\Big(\underbrace{\underset{3\times 2}{\theta^{(1)}}\,\underset{2\times 1}{h^{(1)}}}_{3\times 1} + \underset{3\times 1}{b^{(1)}}\Big)$$

$$h^{(3)} = g\Big(\underset{3\times 1}{\underset{3\times 3}{\theta^{(2)}}\,\underset{3\times 1}{h^{(2)}}} + \underset{3\times 1}{b^{(2)}}\Big)$$

$$h^{(4)} = g\Big(\underbrace{\underset{2\times 3}{\theta^{(3)}}\,\underset{3\times 1}{h^{(3)}}}_{2\times 1} + \underset{2\times 1}{b^{(3)}}\Big)$$

$$h^{(5)} = g\Big(\underbrace{\underset{1\times 2}{\theta^{(4)}}\,\underset{2\times 1}{h^{(4)}}}_{1\times 1} + \underset{1\times 1}{b^{(4)}}\Big) \quad \checkmark$$

$$\delta^{(5)} = 2\underset{1\times 1}{(h^{(5)} - y)} \odot g'\Big(\underset{1\times 1}{\underset{1\times 2}{\theta^{(4)}} \cdot \underset{2\times 1}{h^{(4)}}} + \underset{1\times 1}{b^{(4)}}\Big) \quad \text{output layer.}$$

$$\delta^{(4)} = \underset{2\times 1}{\Big(\underset{2\times 1}{\theta^{(4)T}}\,\underset{1\times 1}{\delta^{(5)}}\Big)} \odot g'\underset{2\times 1}{\Big(\theta^{(3)} h^{(3)} + b^{(3)}\Big)} \qquad 2\times 1$$

$$\delta^{(3)} = \underbrace{\underset{3\times 2}{\theta^{(3)T}}\,\underset{2\times 1}{\delta^{(4)}}}_{3\times 1} \odot g'\underset{3\times 1}{\Big(\theta^{(2)} h^{(2)} + b^{(2)}\Big)}$$

$$\delta^{(2)} = \underbrace{\underset{3\times 3}{\theta^{(2)T}}\,\underset{3\times 1}{\delta^{(3)}}}_{3\times 1} \odot g'\underset{3\times 1}{\Big(\theta^{(1)} h^{(1)} + b^{(1)}\Big)}$$

$\Delta\theta^{(1)} = \delta^{(2)} h^{(1)T}$ 3×2 $\theta^{(1)}$ 3×2

3×1 1×2

$\Delta\theta^{(2)} = \delta^{(3)} h^{(2)T}$ 3×3 $\theta^{(2)}$ 3×3

3×1 1×3

3×3

$\Delta\theta^{(3)} = \delta^{(4)} h^{(3)T}$ 2×3 $\theta^{(3)}$ 2×3

2×1 1×3

2×3

$\Delta\theta^{(4)} = \delta^{(5)} h^{(4)T}$ 1×2 $\theta^{(4)}$ 1×2

1×1 1×2

$\Delta b^{(1)} = \delta^{(2)}$ 3×1

$\Delta b^{(2)} = \delta^{(3)}$ 3×1

$\Delta b^{(3)} = \delta^{(4)}$ 2×1

$\Delta b^{(4)} = \delta^{(5)}$ 1×1

$$h^{(1)} = x$$

$$h^{(2)} = g(\theta^{(1)} h^{(1)} + b^{(1)})$$

$$h^{(L-1)} = g(\theta^{(L-2)} h^{(L-2)} + b^{(L-2)})$$

$$h(x) = h^{(L)} = g(\theta^{(L-1)} h^{(L-1)} + b^{(L-1)}$$

Given $h^{(1)}, h^{(2)}, \ldots, h^{(L)}$

output layer

$$\delta_1^{(L)} = 2(h^{(L)} - y) \odot g'(\theta^{(L-1)} h^{(L-1)} + b^{(L-1)})$$

$$l = L-1, L-2, \ldots, 2$$

$$\delta^{(l)} = \theta^{(l)T} \delta^{(l+1)} \odot g'(\theta^{(l-1)} h^{(l-1)} + b^{(l-1)})$$

$$\Delta \theta^{(l)} = \delta^{(l+1)} h^{(l)T}$$

$$\Delta b^{(l)} = \delta^{(l+1)}$$