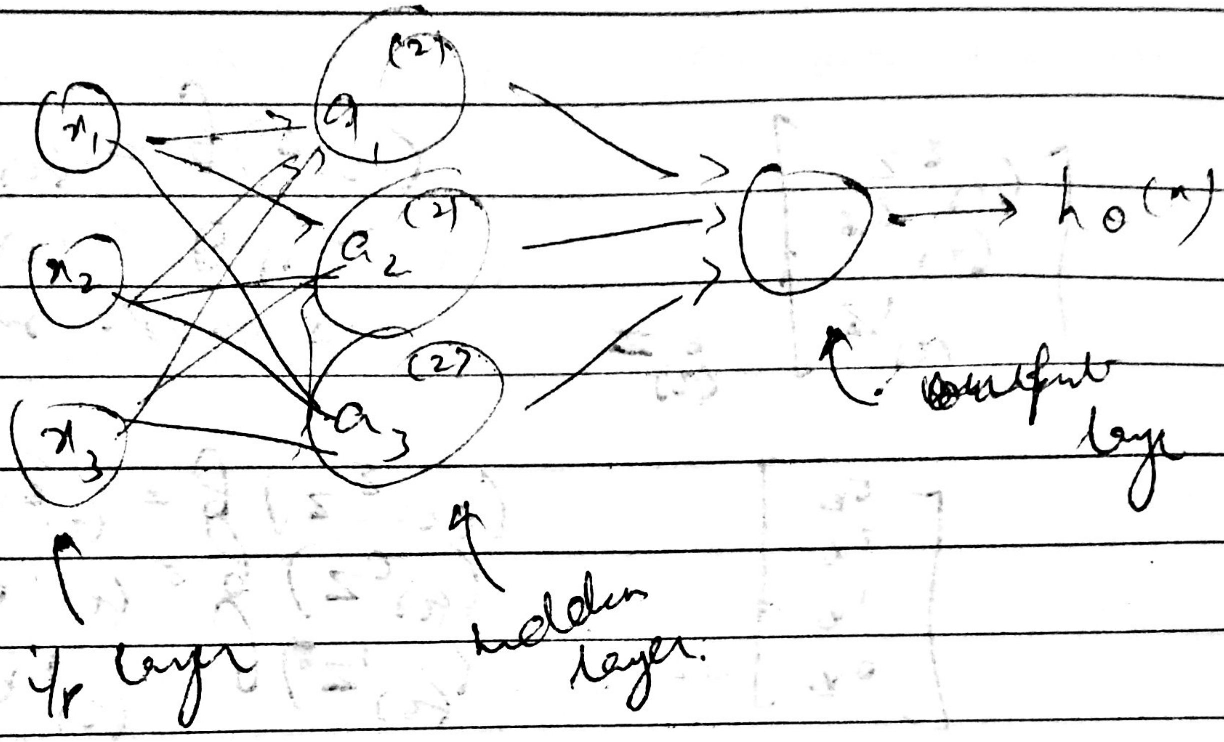
Week 4 Alberta.

Neural Network

theta = parameters or weights



$a_i^{(j)}$ = activation of unit 'i' in layer 'j'

θ_j = matrix of weights controlling function mapping from layer j to layer j+1

$$a_1^{(2)} = g(\theta_{10}^{(1)} x_0 + \theta_{11}^{(1)} x_1 + \theta_{12}^{(1)} x_2 + \theta_{13}^{(1)} x_3)$$
$$a_2^{(2)} = g(\theta_{20}^{(1)} x_0 + \theta_{21}^{(1)} x_1 + \theta_{22}^{(1)} x_2 + \theta_{23}^{(1)} x_3)$$
$$a_3^{(2)} = g(\theta_{30}^{1} x_0 + \theta_{31}^{1} x_1 + \theta_{32}^{1} x_2 + \theta_{33}^{1} x_3)$$
$$h_\theta(x) = g(\theta_{10}^{(2)} a_0^{(2)} + \theta_{11}^{(2)} a_1^{(2)} + \theta_{12}^{(2)} a_2^{(2)} + \theta_{13}^{(2)} a_3^{(2)})$$

If network has s_j units in layer j, s_{j+1} units in layer $j+1$, then $\theta^{(j)}$ will be of dimension

$$(s_{j+1} \times s_j + 1)$$

$$z_1^{(2)}$$

$$a_1^{(2)} = g(\theta_{10}^{(1)} x_0^{(1)} + \theta_{11}^{(1)} x_1^{(1)} + \theta_{12}^{(1)} x_2^{(1)} + \theta_{13}^{(1)} x_3^{(1)})$$

$$a_2^{(2)} = g(\theta_{20}^{(1)} x_0^{(1)} + \theta_{21}^{(1)} x_1^{(1)} + \theta_{22}^{(1)} x_2^{(1)} + \theta_{23}^{(1)} x_3^{(1)})$$

$$a_3^{(2)} = g(\theta_{30}^{(1)} x_0^{(1)} + \theta_{31}^{(1)} x_1^{(1)} + \theta_{32}^{(1)} x_2^{(1)} + \theta_{33}^{(1)} x_3^{(1)})$$

$$\therefore \; a_1^{(2)} = g(z_1^{(2)}) \qquad \text{if } x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$a_2^{(2)} = g(z_2^{(2)})$$

$$a_3^{(2)} = g(z_3^{(2)})$$

Now, we vectorize.

$$z^{(2)} = \begin{bmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \end{bmatrix}$$

$$z^{(2)} = \theta^{(1)} x$$

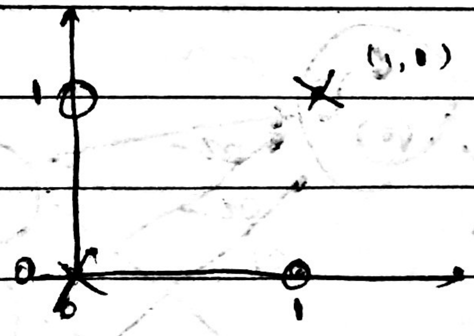
$$a^{(2)} = g(z^{(2)})$$

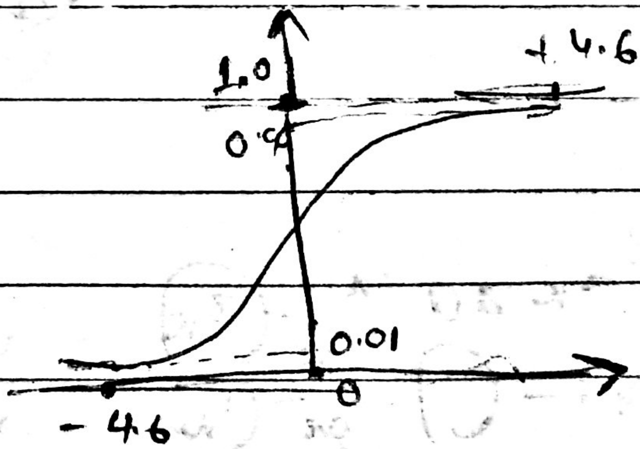
That was <u>forward</u> propagation because, we got from i/p to o/p.

Neural Network learns its own features.

x_1, x_2 are binary

$y = x_1$ ~~XOR~~ x_2 XNOR

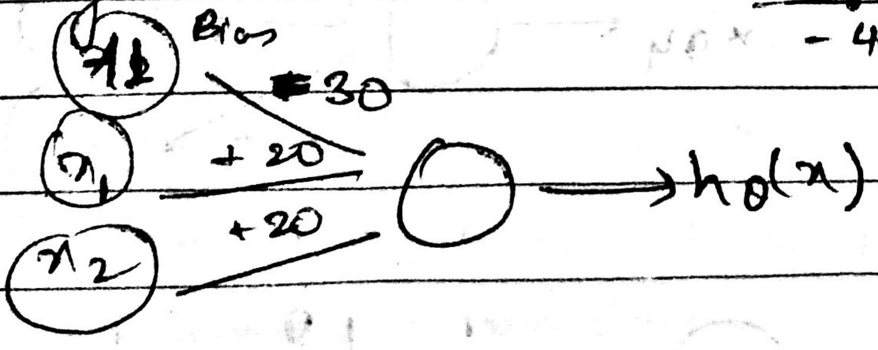




Simple example: AND

$x_1, x_2 \in \{0, 1\}$

$y = x_1$ AND x_2



$h_\theta(x) = g(-30 + 20x_1 + 20x_2)$

x_1	x_2	$h_\theta(x)$
0	0	$g(-30) \approx 0$
0	1	$g(-10) \approx 0$
1	0	$g(-10) \approx 0$
1	1	$g(10) \approx 1$