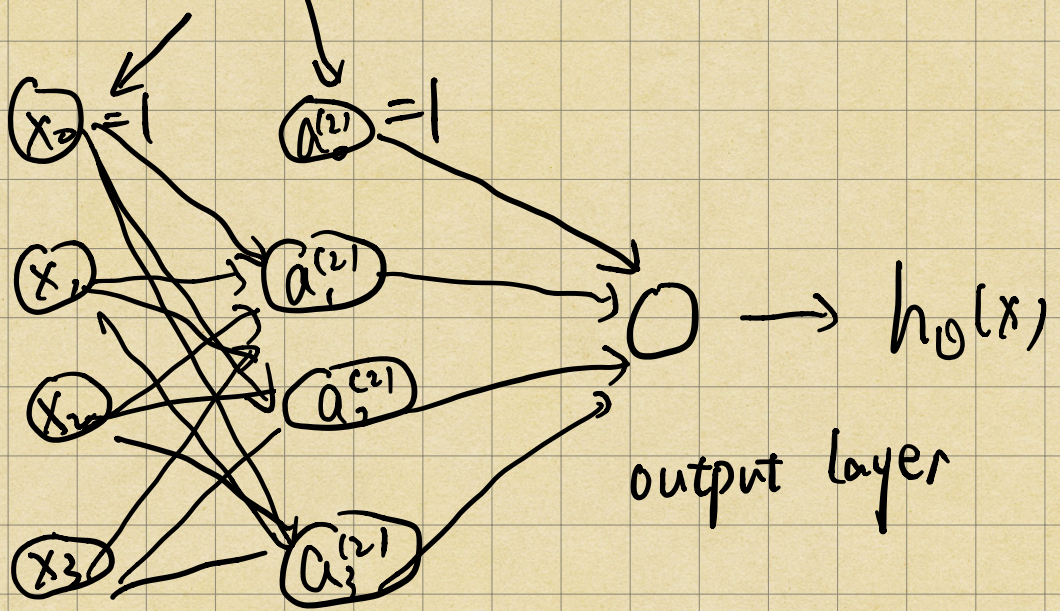
activaction function.

$g(z) = \frac{1}{1+e^{-z}}$ (sigmoid)

weight (parameters)

$\begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$ bias unit



$a_i^{(j)} \Rightarrow$ unit i in layer j

$\theta^j \Rightarrow$ weights controlling function mapping from layer j to $j+1$

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

$z_1^{(2)}$ ← related to 2 layer

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

$z_2^{(2)}$

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

$z_3^{(2)}$

$$h_\theta(x) = g(\theta_{10}^{(2)} a_0^{(2)} + \theta_{11}^{(2)} a_1^{(2)} + \theta_{12}^{(2)} a_2^{(2)} + \theta_{13}^{(2)} a_3^{(2)})$$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad z^{(2)} = \begin{bmatrix} z_1^{(2)} \\ z_2^{(2)} \\ z_3^{(2)} \end{bmatrix} = \theta^{(1)} x = \theta^{(1)} a^{(1)}$$

$$a^{(2)} = g(z^{(2)}) \rightarrow h_\theta(x) = a^{(3)} = g(z^{(3)})$$

forward (input → hidden → output)