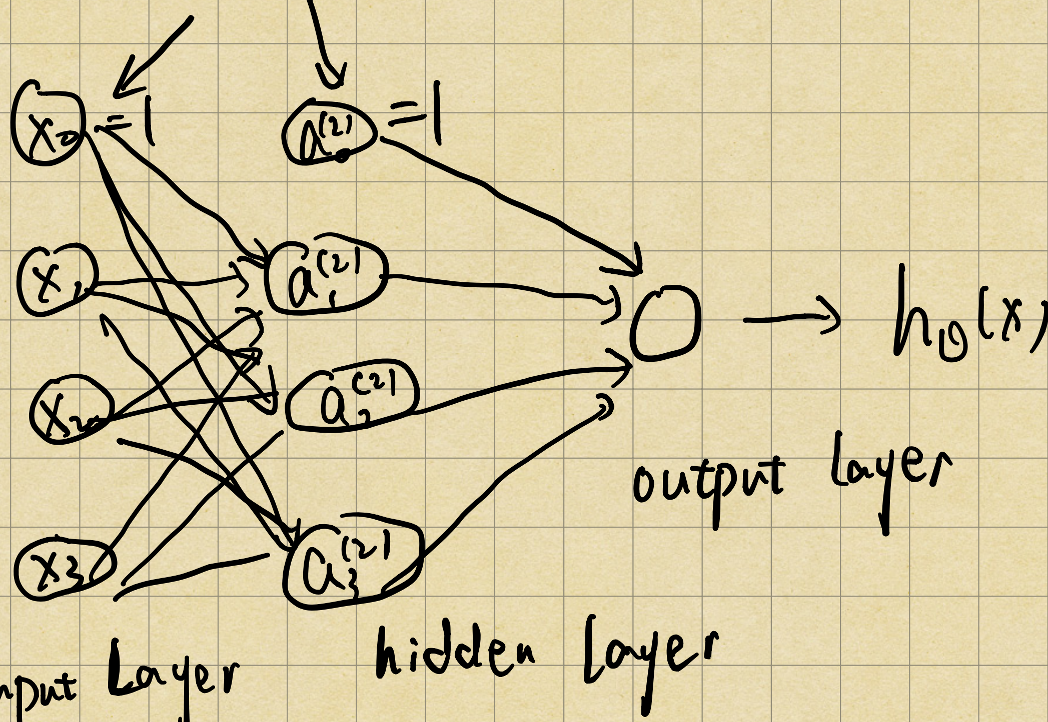
activaction function.

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

$$a_1^{(2)} = g(\underbrace{\Theta_{10}^{(1)}x_0 + \Theta_{11}^{(1)}x_1 + \Theta_{12}^{(1)}x_2 + \Theta_{13}^{(1)}x_3}_{z_1^{(2)}})$$

$z_1^{(2)} \leftarrow$ related to 2 layer

$$a_2^{(2)} = g(\underbrace{\Theta_{(20)}^{(1)}x_0 + \Theta_{21}^{(1)}x_1 + \Theta_{(22)}^{(1)}x_2 + \Theta_{(23)}^{(1)}x_3}_{z_2^{(2)}})$$

$$a_3^{(2)} = g(\underbrace{\Theta_{(30)}^{(1)}x_0 + \Theta_{(31)}^{(1)}x_1 + \Theta_{(32)}^{(1)}x_2 + \Theta_{(33)}^{(1)}x_3}_{z_3^{(2)}})$$

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

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

$$\underbrace{a^{(2)} = g(z^{(2)}) \longrightarrow h_\Theta(x) = a^{(3)} = g(z^{(3)})}_{\text{forward (input} \to \text{hidden} \to \text{output)}}$$

Cost function.

$$J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} \log h(x^{(i)})_k + (1 - y_k^{(i)})\log(1 - h_\Theta(x^{(i)}))_k\right]$$

$$+ \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}(\Theta_j^{(l)})^2.$$

squares of each θ_s

logistic regression cost of each cell

Backpropagation algorithm

$\delta_j^{(l)}$:= erro of node j in layer l

$$\delta_j^{(4)} = a_j^{(4)} - y_j$$

$$\delta^{(3)} = (\theta^{(3)})^T \delta^{(4)} \cdot g'(z^{(3)})$$

$$\frac{\partial}{\partial \theta_{ij}^{(l)}} J(\theta) = a_j^{(l)} \delta_i^{(l+1)}$$

Gradient Checking

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{J(\theta_j+\varepsilon) - J(\theta_j-\varepsilon)}{2\varepsilon}$$