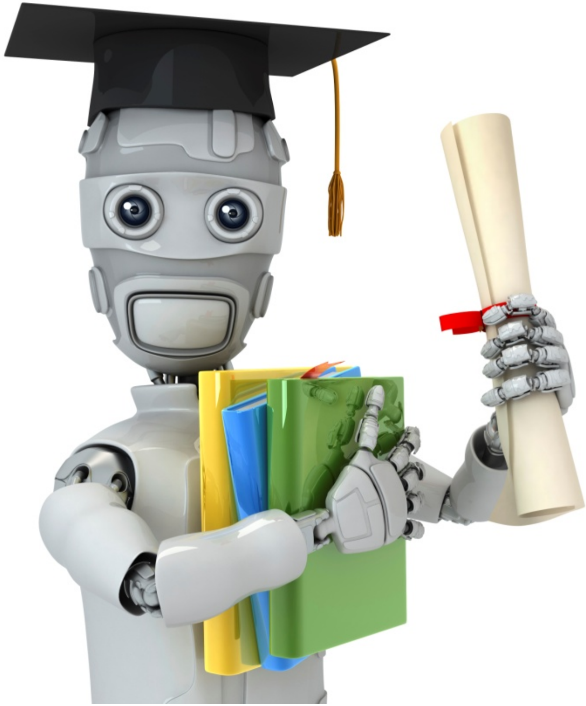
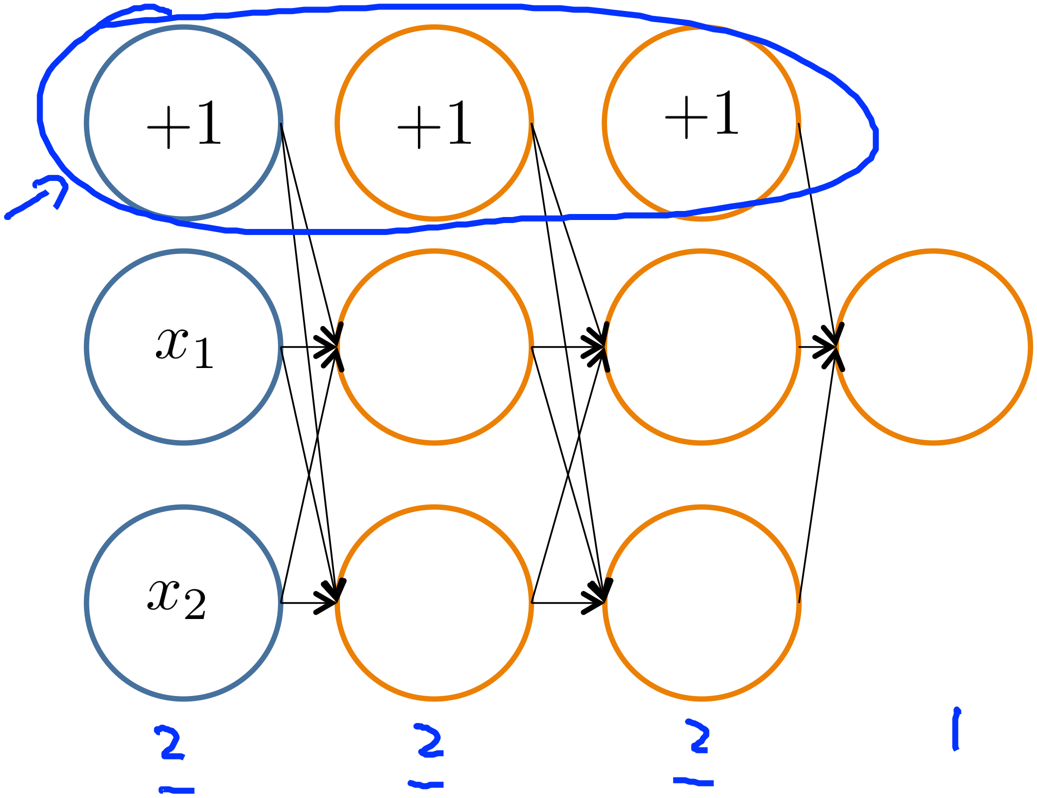


Machine Learning

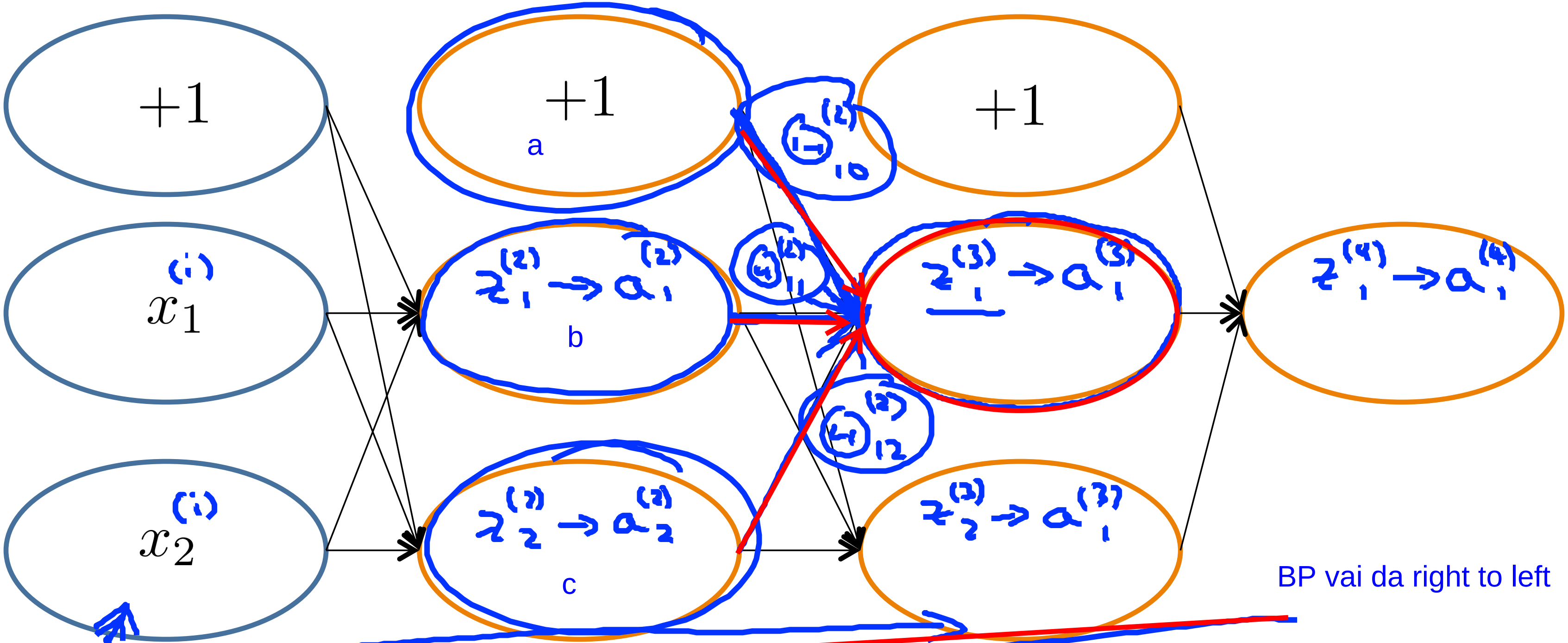
Neural Networks: Learning

Backpropagation intuition

Forward Propagation



Forward Propagation



$(x^{(i)}, y^{(i)})$

$$z_1^{(3)} = \underbrace{\Theta_{10}^{(2)} \times 1}_{a} + \underbrace{\Theta_{11}^{(2)} a_1^{(2)}}_{b} + \underbrace{\Theta_{12}^{(2)} \cdot a_1^{(2)}}_{c}$$

formula par apenas 1 output unit

What is backpropagation doing?

$$J(\Theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} \underbrace{y^{(i)}\log(h_\Theta(x^{(i)})) + (1-y^{(i)})\log(1-(h_\Theta(x^{(i)})))}_{(x^{(i)},\, y^{(i)})}\right]$$

$$\cancel{+\frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_{ji}^{(l)})^2}$$

Focusing on a single example $x^{(i)}$, $y^{(i)}$, the case of 1 output unit, and ignoring regularization ($\lambda = 0$),

$$\boxed{\text{cost(i)} = y^{(i)}\log h_\Theta(x^{(i)}) + (1-y^{(i)})\log h_\Theta(x^{(i)})}$$

(Think of $\text{cost(i)} \approx (\boxed{h_\Theta(x^{(i)})} - \boxed{y^{(i)}})^2$)

I.e. how well is the network doing on example i?

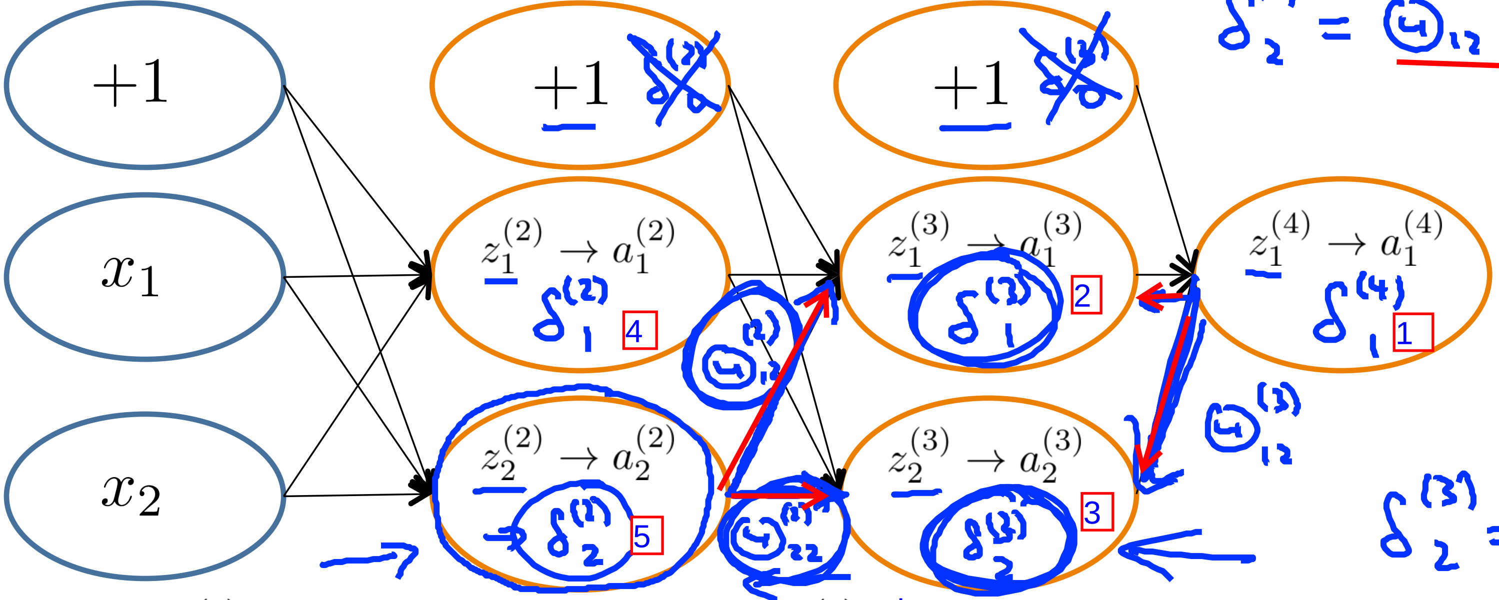
$y^{(i)}$

Andrew Ng

Forward Propagation

$$\delta_1^{(4)} = y^{(i)} - a_1^{(4)}$$

$$\delta_2^{(2)} = \Theta_{12}^{(2)} \delta_1^{(3)} + \Theta_{22}^{(2)} \delta_2^{(3)}$$



$$\delta_2^{(3)} = \Theta_{12}^{(3)} \cdot \delta_1^{(4)}.$$

$\delta_j^{(l)}$ = "error" of cost for $a_j^{(l)}$ (ativaçao) (unit j in layer l).

Formally, $\delta_j^{(l)} = \frac{\partial}{\partial z_j^{(l)}} \text{cost}(i)$ (for $j \geq 0$), where

$$\text{cost}(i) = y^{(i)} \log h_\Theta(x^{(i)}) + (1 - y^{(i)}) \log h_\Theta(x^{(i)})$$