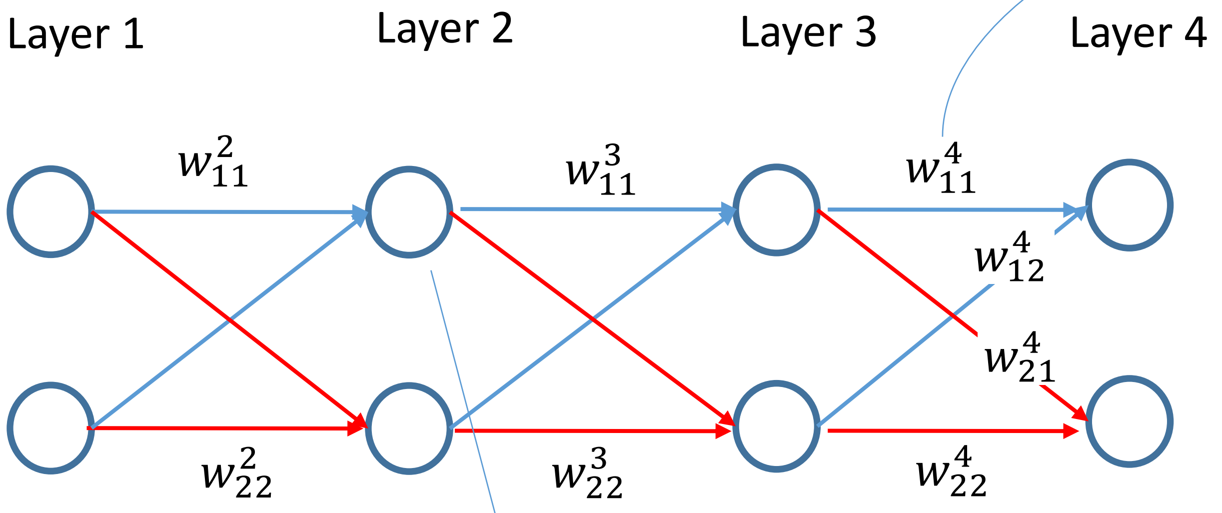
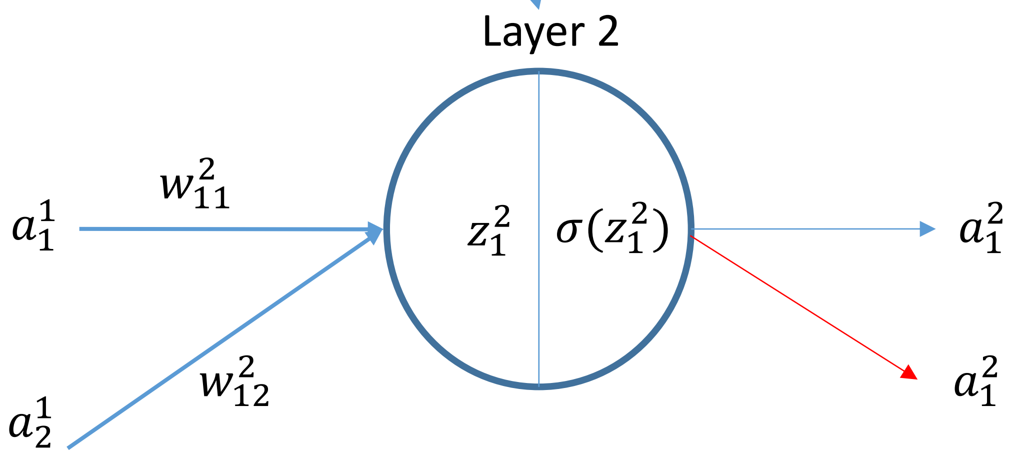
Back-Propagation-Auto-Diff

Harpreet Singh (Fall 2023)

Notation – input/output

w_{jk}^{l} is the weight from k^{th} neuron in $(l-1)^{\mathrm{th}}$ layer to j^{th} neuron in l^{th} layer



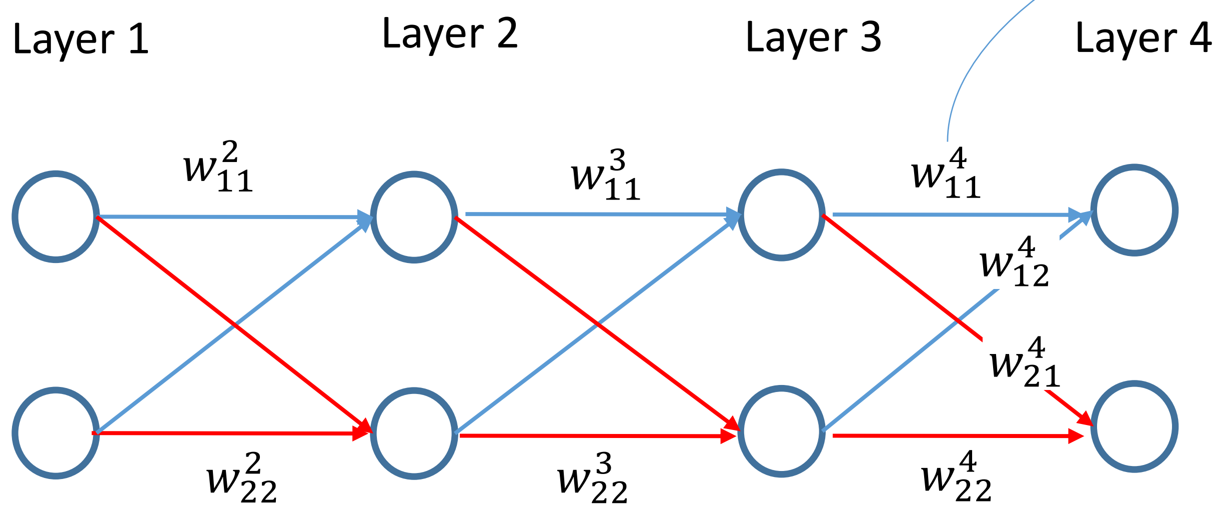


Total Input to neuron $= z_1^2 = w_{11}^2 a_1^1 + w_{12}^2 a_2^1 + b_1^2$

Output from neuron $= a_1^2 = \sigma(z_1^2)$

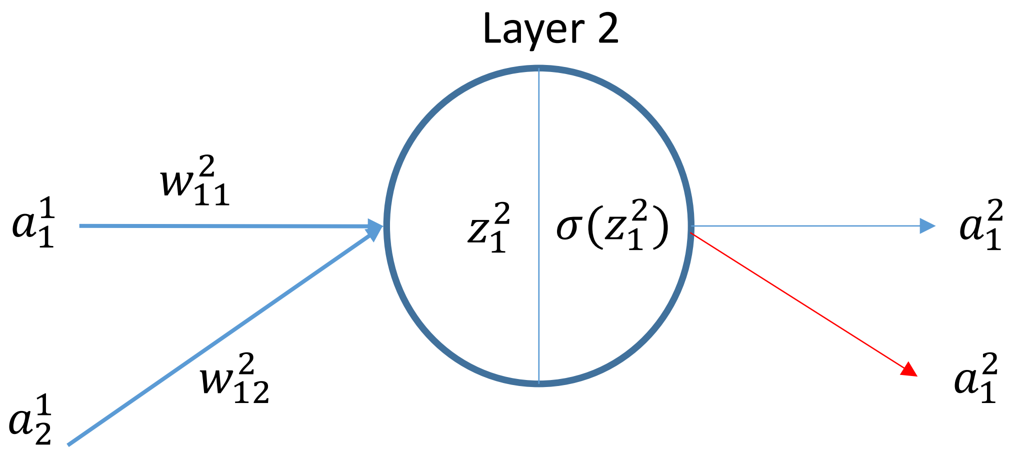
Notation input/output

w_{jk}^{l} is the weight from k^{th} neuron in $(\mathrm{l}-1)^{\mathrm{th}}$ layer to j^{th} neuron in l^{th} layer



Total Input to neuron $= z_j^l = \sum_k w_{jk}^l a_k^{l-1} + b_j^l$

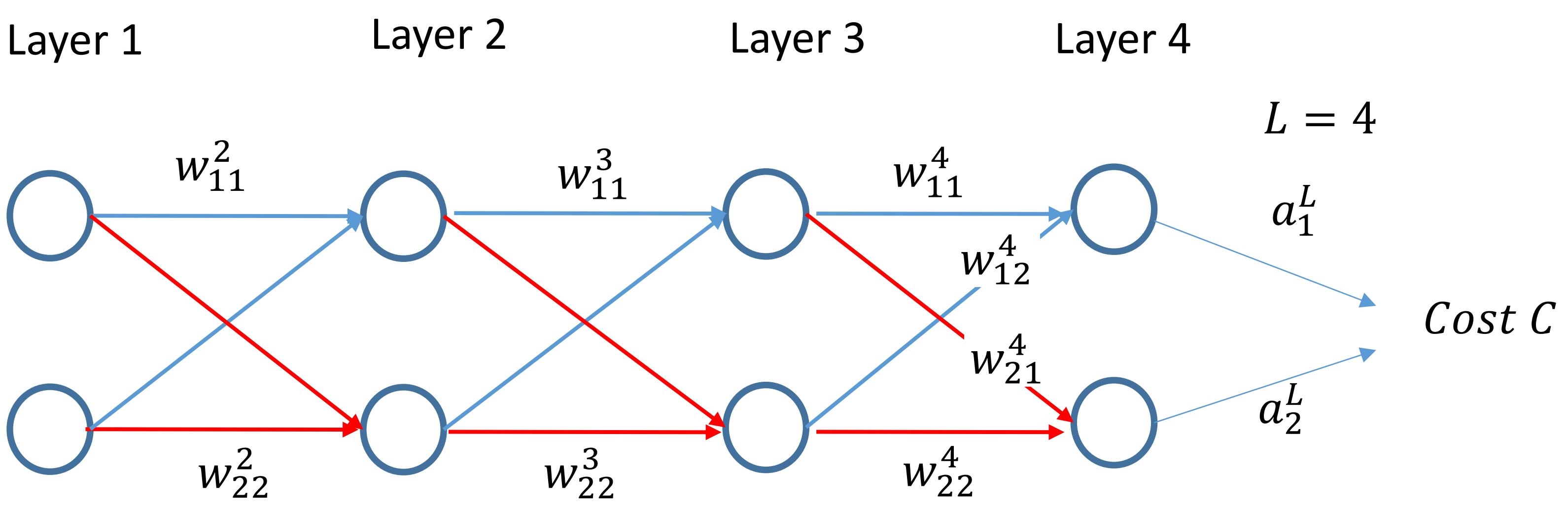
Matrix Notation $\quad z_j^l \equiv w^l a^{l-1} + b_j^l$



Output from neuron$=$ $a_j^l = \sigma(z_j^l)$

Matrix Notation $\quad a^l = \sigma(z^l)$

Notation- Cost



For MSE (mean Squared Error)

$$C = \frac{1}{2}\|y - a^L\|^2 = \frac{1}{2}\sum_j (y_j - a_j^L)^2$$

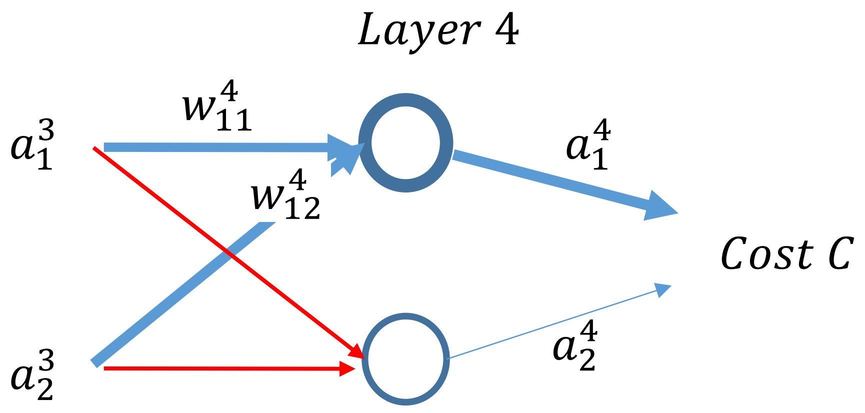
For n observations

$$C = \frac{1}{2n}\sum_x \|y(x) - a^L(x)\|^2$$

Calculations – Last layer (Output Layer)

Objective : $\frac{\partial C}{\partial w_{jk}^L}, \frac{\partial C}{\partial b_j^L}$

For illustration we will use following : $\frac{\partial C}{\partial w_{11}^4}, \frac{\partial C}{\partial b_1^4}$



$$z_1^4 = w_{11}^4 a_1^3 + w_{12}^4 a_2^3 + b_1^4 \qquad a_1^4 = \sigma(z_1^4)$$

$$C = \frac{1}{2}\left[(y_1 - a_1^4)^2 + (y_2 - a_2^4)^2\right]$$

$$\frac{\partial C}{\partial w_{11}^4} = \frac{\partial C}{\partial a_1^4}\frac{\partial a_1^4}{\partial z_1^4}\frac{\partial z_1^4}{\partial w_{11}^4}$$

$$\frac{2}{2}(y_1 - a_1^4) * (-1) = (a_1^4 - y_1)$$

error

$$\sigma'(z_1^4)$$

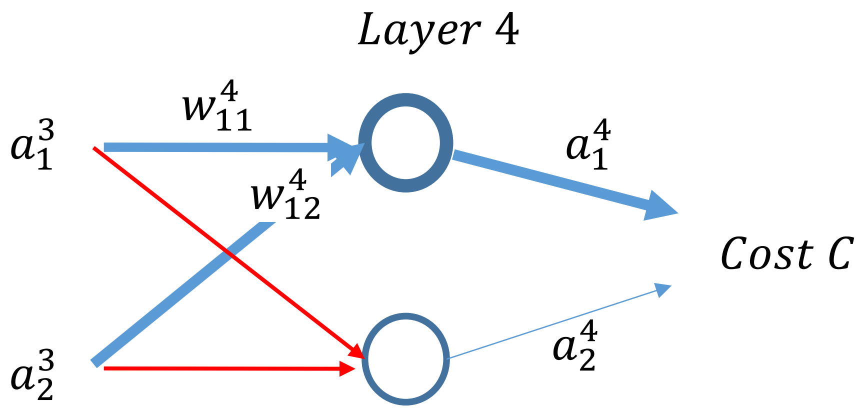
Derivative of activation function

$$a_1^3$$

Calculations – Last layer (Output Layer)

Objective : $\frac{\partial C}{\partial w_{jk}^{L}}, \frac{\partial C}{\partial b_{j}^{L}}$

For illustration we will use following : $\frac{\partial C}{\partial w_{11}^{4}}, \frac{\partial C}{\partial b_{1}^{4}}$



$$z_1^4 = w_{11}^4 a_1^3 + w_{12}^4 a_2^3 + b_1^4$$

$$\frac{\partial C}{\partial w_{11}^4} = \frac{\partial C}{\partial a_1^4}\frac{\partial a_1^4}{\partial z_1^4}\frac{\partial z_1^4}{\partial w_{11}^4} = (a_1^4 - y_1)\,\sigma'(z_1^4)\,a_1^3$$

$$\frac{\partial C}{\partial w_{11}^4} = \frac{\partial C}{\partial z_1^4}\frac{\partial z_1^4}{\partial w_{11}^4} = (a_1^4 - y_1)\,\sigma'(z_1^4)\,a_1^3$$

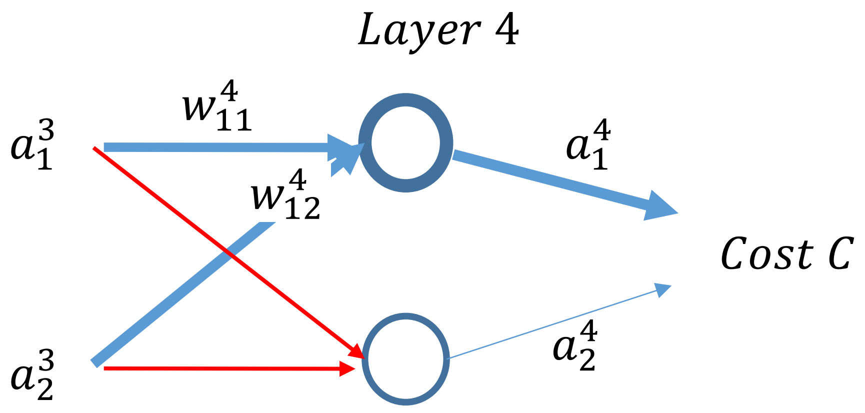
$$\frac{\partial C}{\partial w_{11}^4} = \delta_1^4 a_1^3\text{ , where }\delta_1^4 = \frac{\partial C}{\partial z_1^4} = \frac{\partial C}{\partial a_1^4}\frac{\partial a_1^4}{\partial z_1^4}$$

$$\frac{\partial C}{\partial w_{jk}^4} = \delta_j^L a_k^{L-1}\text{ , where }\delta_j^L = \frac{\partial C}{\partial z_j^L} = \frac{\partial C}{\partial a_1^4}\,\sigma'(z_1^4)$$

Calculations – Last layer (Output Layer)

Objective : $\dfrac{\partial C}{\partial w_{jk}^L}, \dfrac{\partial C}{\partial b_j^L}$

For illustration we will use following : $\dfrac{\partial C}{\partial w_{11}^4}, \dfrac{\partial C}{\partial b_1^4}$



$$z_1^4 = w_{11}^4 a_1^3 + w_{12}^4 a_2^3 + b_1^4$$

$$\frac{\partial C}{\partial b_1^4} = \frac{\partial C}{\partial a_1^4} \frac{\partial a_1^4}{\partial z_1^4} \frac{\partial z_1^4}{\partial b_1^4} = \delta_1^4 * \frac{\partial z_1^4}{\partial b_1^4} = \delta_1^4$$

Summary: $\dfrac{\partial C}{\partial w_{jk}^L} = \delta_j^L a_k^{L-1}, \quad \dfrac{\partial C}{\partial b_l^L} = \delta_j^L$, where $\delta_j^L = \dfrac{\partial C}{\partial z_j^L} = \dfrac{\partial C}{\partial a_j^L} \sigma'\left(z_j^L\right)$

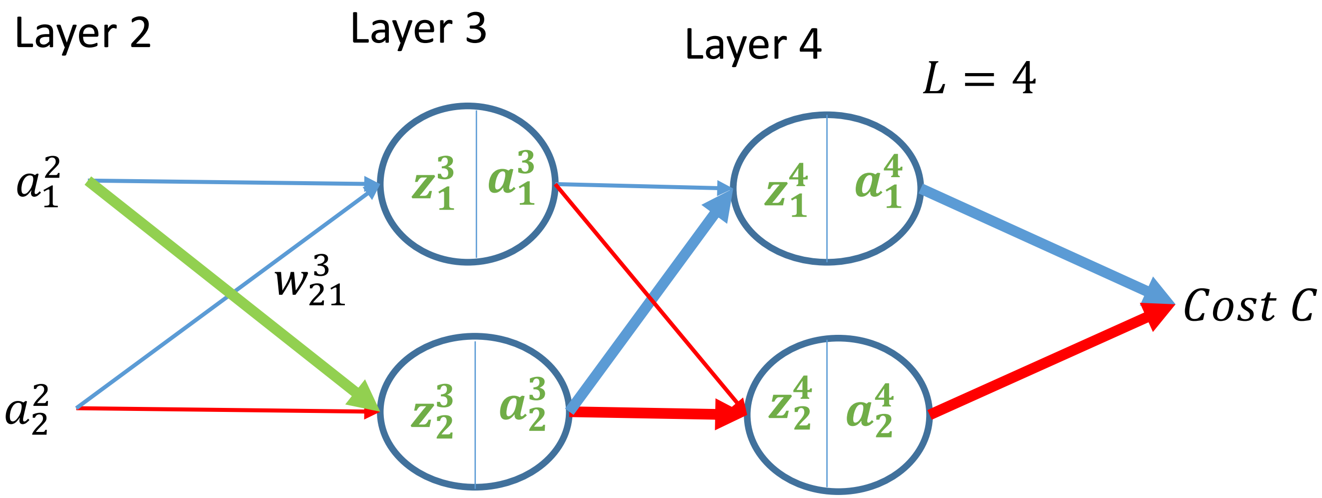
	This Example	Linear Regression	Logistic Regression
δ_j^L	Error * derivative of activation function	Error	Error

Calculations – Hidden layer

For illustration we will use following :

$$\frac{\partial C}{\partial w_{21}^{3}}, \frac{\partial C}{\partial b_{2}^{3}}$$

Objective : $\frac{\partial C}{\partial w_{jk}^{l}}, \frac{\partial C}{\partial b_{j}^{l}}$



$$\frac{\partial C}{\partial w_{21}^{3}} = \frac{\partial C}{\partial a_1^4}\frac{\partial a_1^4}{\partial z_1^4}\frac{\partial z_1^4}{\partial a_2^3}\frac{\partial a_2^3}{\partial z_2^3}\frac{\partial z_2^3}{\partial w_{21}^3} + \frac{\partial C}{\partial a_2^4}\frac{\partial a_2^4}{\partial z_2^4}\frac{\partial z_2^4}{\partial a_2^3}\frac{\partial a_2^3}{\partial z_2^3}\frac{\partial z_2^3}{\partial w_{21}^3}$$

$$\boxed{\frac{\partial C}{\partial a_1^4}\frac{\partial a_1^4}{\partial z_1^4} = \frac{\partial C}{\partial a_1^4}\sigma'(z_1^4) = \delta_1^4}$$

$$\boxed{\frac{\partial C}{\partial a_2^4}\frac{\partial a_2^4}{\partial z_2^4} = \frac{\partial C}{\partial a_2^4}\sigma'(z_2^4) = \delta_2^4}$$

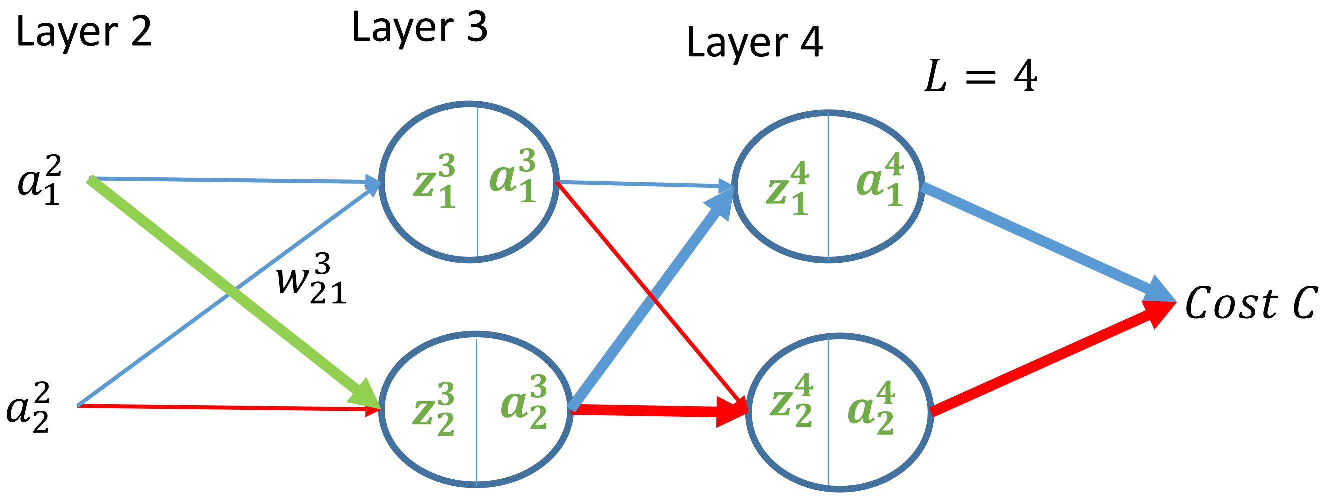
Calculations – Hidden layer

Objective :

$$\frac{\partial C}{\partial w_{jk}^l}, \frac{\partial C}{\partial b_j^l}$$

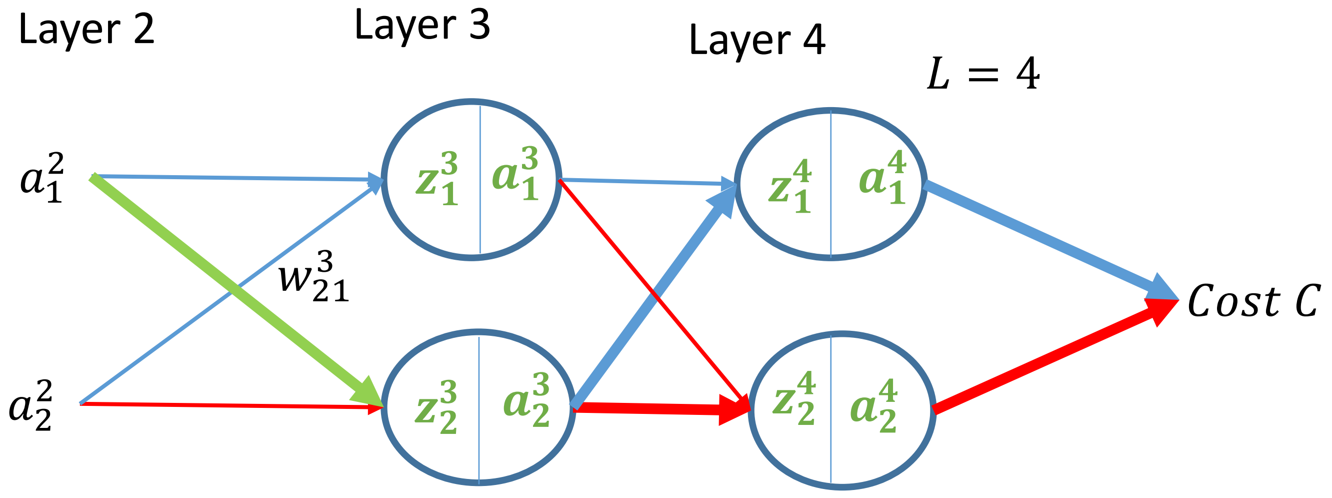
For illustration we will use following :

$$\frac{\partial C}{\partial w_{21}^3}, \frac{\partial C}{\partial b_2^3}$$



$$\frac{\partial C}{\partial w_{21}^3} = \left(\delta_1^4 \frac{\partial z_1^4}{\partial a_2^3} + \delta_2^4 \frac{\partial z_2^4}{\partial a_2^3}\right) \frac{\partial a_2^3}{\partial z_2^3} \frac{\partial z_2^3}{\partial w_{21}^3}$$

Calculations – Hidden layer



Objective :

$$\frac{\partial C}{\partial w_{jk}^l}, \frac{\partial C}{\partial b_j^l}$$

For illustration we will use following :

$$\frac{\partial C}{\partial w_{21}^3}, \frac{\partial C}{\partial b_2^3}$$

$$z_1^4 = w_{11}^4 a_1^3 + w_{12}^4 a_2^3 + b_1^4$$

$$\frac{\partial z_1^4}{\partial a_2^3} = \mathrm{w}_{12}^4$$

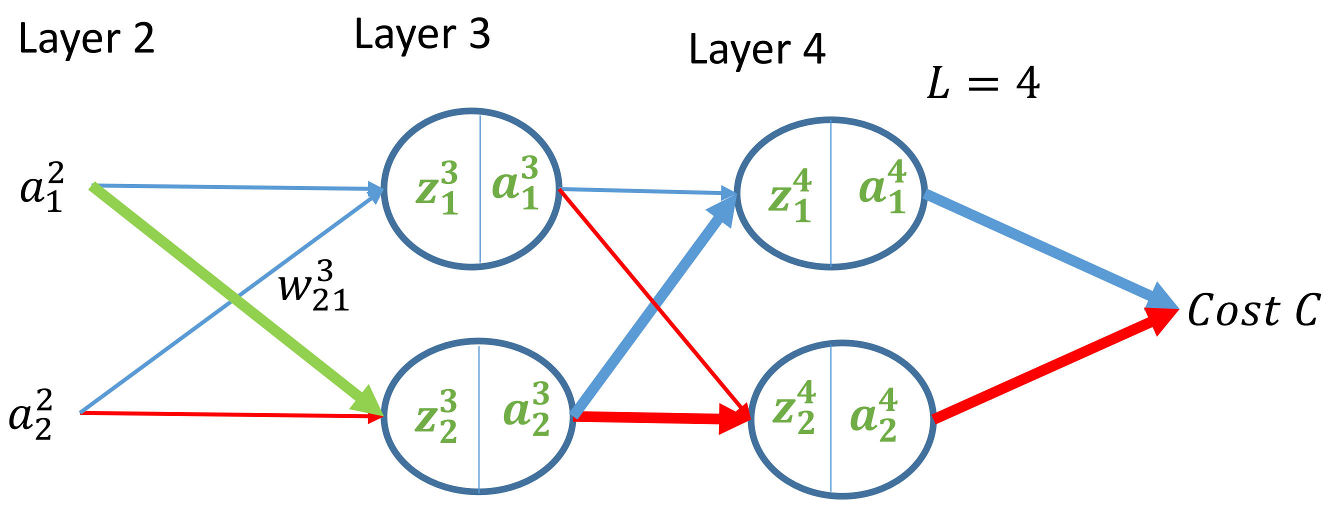
$$z_2^4 = w_{21}^4 a_1^3 + w_{22}^4 a_2^3 + b_2^4$$

$$\frac{\partial z_2^4}{\partial a_2^3} = \mathrm{w}_{22}^4$$

$$\frac{\partial C}{\partial w_{21}^3} = \left(\delta_1^4 \frac{\partial z_1^4}{\partial a_2^3} + \delta_2^4 \frac{\partial z_2^4}{\partial a_2^3}\right) \frac{\partial a_2^3}{\partial z_2^3} \frac{\partial z_2^3}{\partial w_{21}^3}$$

$$= \left(\delta_1^4 \mathbf{w}_{12}^4 + \delta_2^4 \mathrm{w}_{22}^4\right) \frac{\partial a_2^3}{\partial z_2^3} \frac{\partial z_2^3}{\partial w_{21}^3}$$

Calculations – Hidden layer



Objective :

$$\frac{\partial C}{\partial w_{jk}^l}, \frac{\partial C}{\partial b_j^l}$$

For illustration we will use following :

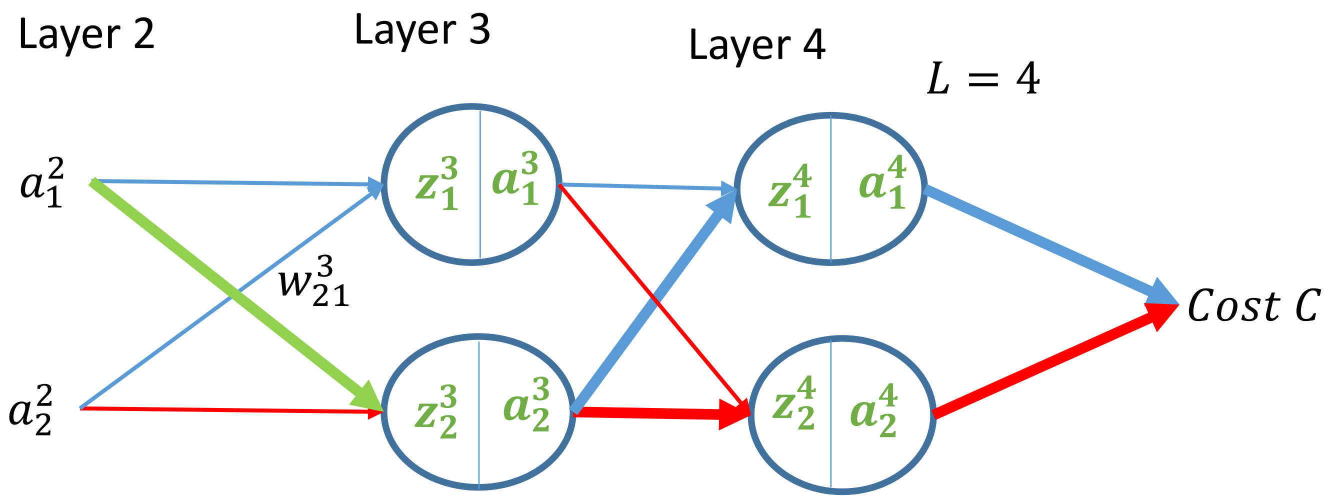
$$\frac{\partial C}{\partial w_{21}^3}, \frac{\partial C}{\partial b_2^3}$$

$$\frac{\partial C}{\partial w_{21}^3} = \left(\delta_1^4 \mathbf{w}_{12}^4 + \delta_2^4 \mathrm{w}_{22}^4\right) \frac{\partial a_2^3}{\partial z_2^3} \frac{\partial z_2^3}{\partial w_{21}^3}$$

$$= \left(\delta_1^4 \mathbf{w}_{12}^4 + \delta_2^4 \mathrm{w}_{22}^4\right) \sigma'(z_2^3) \; \frac{\partial z_2^3}{\partial w_{21}^3}$$

$$\boxed{\begin{aligned} a_2^3 &= \sigma(z_2^3) \\ \frac{\partial a_2^3}{\partial z_2^3} &= \sigma'(z_2^3) \end{aligned}}$$

Calculations – Hidden layer



Objective :

$$\frac{\partial C}{\partial w_{jk}^l}, \frac{\partial C}{\partial b_j^l}$$

For illustration we will use following :

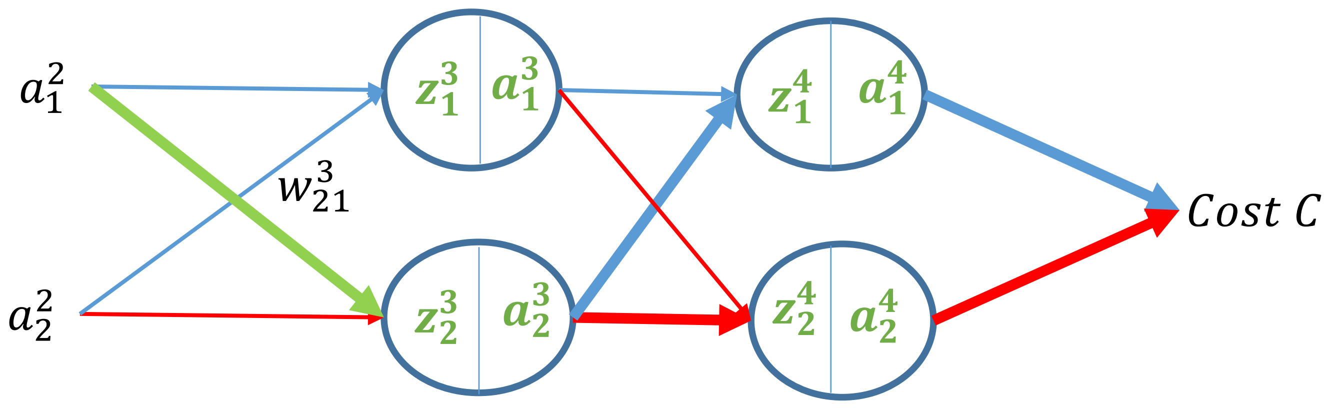
$$\frac{\partial C}{\partial w_{21}^3}, \frac{\partial C}{\partial b_2^3}$$

$$z_2^3 = w_{21}^3 a_1^2 + w_{22}^3 a_2^2 + b_2^3$$

$$\frac{\partial z_2^3}{\partial w_{21}^3} = a_1^2$$

$$\frac{\partial C}{\partial w_{21}^3} = \underbrace{\left(\delta_1^4 w_{12}^4 + \delta_2^4 w_{22}^4\right) \sigma'(z_2^3)}_{\frac{\partial c}{\partial z_2^3} = \delta_2^3} \frac{\partial z_2^3}{\partial w_{21}^3}$$

Calculations – Hidden layer



Objective :

$$\frac{\partial C}{\partial w_{jk}^l}, \frac{\partial C}{\partial b_j^l}$$

For illustration we will use following :

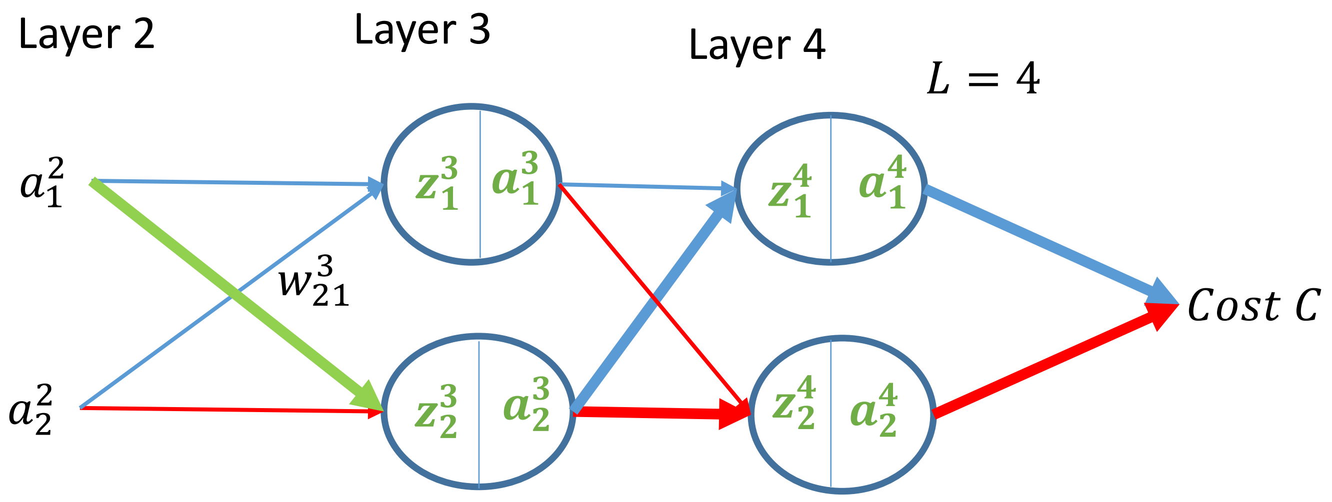
$$\frac{\partial C}{\partial w_{21}^3}, \frac{\partial C}{\partial b_2^3}$$

$$\frac{\partial C}{\partial w_{21}^3} = \underbrace{\left(\delta_1^4 \mathbf{w}_{12}^4 + \delta_2^4 \mathrm{w}_{22}^4\right) \sigma'(z_2^3)}_{\frac{\partial c}{\partial z_2^3} = \delta_2^3} \frac{\partial z_2^3}{\partial w_{21}^3}$$

$$\delta_j^l = \sum_k w_{kj}^{l+1} \delta_k^{l+1} \sigma'(z_j^l).$$

$$\frac{\partial c}{\partial z_2^3} = \delta_2^3$$ = weighted sum of deltas from previous layer * derivative of activation function

Calculations – Hidden layer



Objective :

$$\frac{\partial C}{\partial w_{jk}^l}, \frac{\partial C}{\partial b_j^l}$$

For illustration we will use following :

$$\frac{\partial C}{\partial w_{21}^3}, \frac{\partial C}{\partial b_2^3}$$

$$z_2^3 = w_{21}^3 a_1^2 + w_{22}^3 a_2^2 + b_2^3$$

$$\frac{\partial z_2^3}{\partial w_{21}^3} = a_1^2 \qquad \frac{\partial z_2^3}{\partial b_2^3} = 1$$

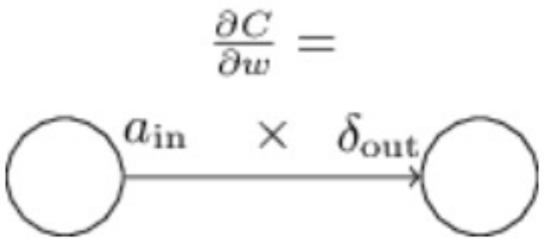
$$\frac{\partial C}{\partial w_{21}^3} = \delta_2^3 \frac{\partial z_2^3}{\partial w_{21}^3} = \delta_2^3 a_1^2, \quad \frac{\partial C}{\partial b_2^3} = \delta_2^3 \frac{\partial z_2^3}{\partial b_2^3} = \delta_2^3$$

$$\frac{\partial C}{\partial w_{jk}^l} = \delta_j^l a_k^{l-1}, \frac{\partial C}{\partial b_j^l} = \delta_j^L$$

Gradient Update Rule

$$\frac{\partial C}{\partial w_{jk}^{l}} = \delta_j^l a_k^{l-1}, \frac{\partial C}{\partial b_j^l} = \delta_j^L$$

$$\frac{\partial C}{\partial w} = a_{\rm in}\delta_{\rm out},$$



Delta

$$\delta_j^l \equiv \frac{\partial C}{\partial z_j^l}.$$

Hidden Layer $$\delta_j^l = \sum_k w_{kj}^{l+1}\delta_k^{l+1}\sigma'(z_j^l).$$

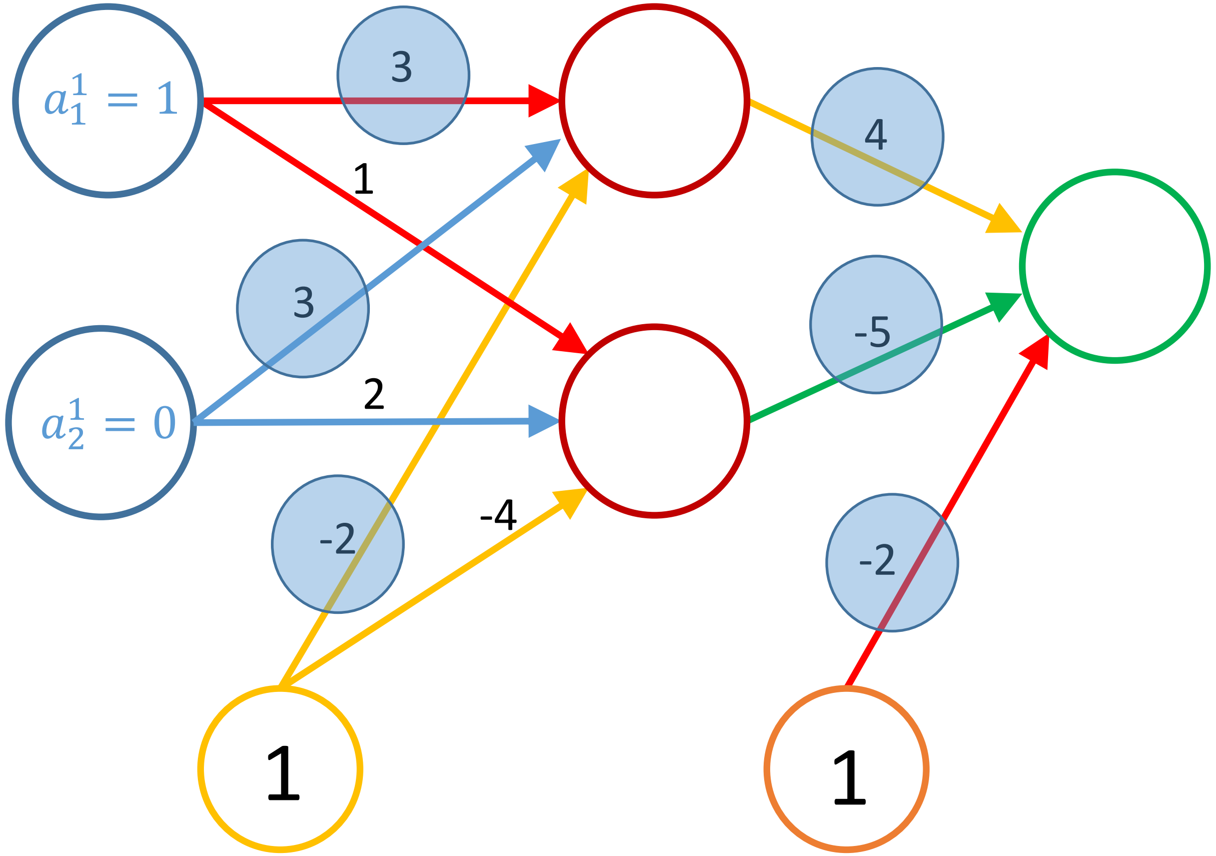
= weighted sum of deltas from previous layer

* derivative of activation function

Output Layer $$\delta_j^L = \frac{\partial C}{\partial a_j^L}\sigma'\left(z_j^L\right)$$

	This Example	Linear Regression	Logistic Regression
δ_j^L	Error * derivative of activation function	Error	Error

Simple Neural Network



Implement One pass of forward
Implement backward pass the update the circled values

Use sigmoid activation function for both hidden layer and output layer

Correct Output = t = 1.0

Useful Formulae:

$$y = sigmoid(x) = 1/(1 + e^{-x})$$

$$\frac{\partial y}{\partial x} = \frac{\partial sigmoid(x)}{\partial y}$$

$$= \mathrm{sigmoid(x)(1 - sigmoid(x)}$$

$$= \mathrm{y(1 - y)}$$