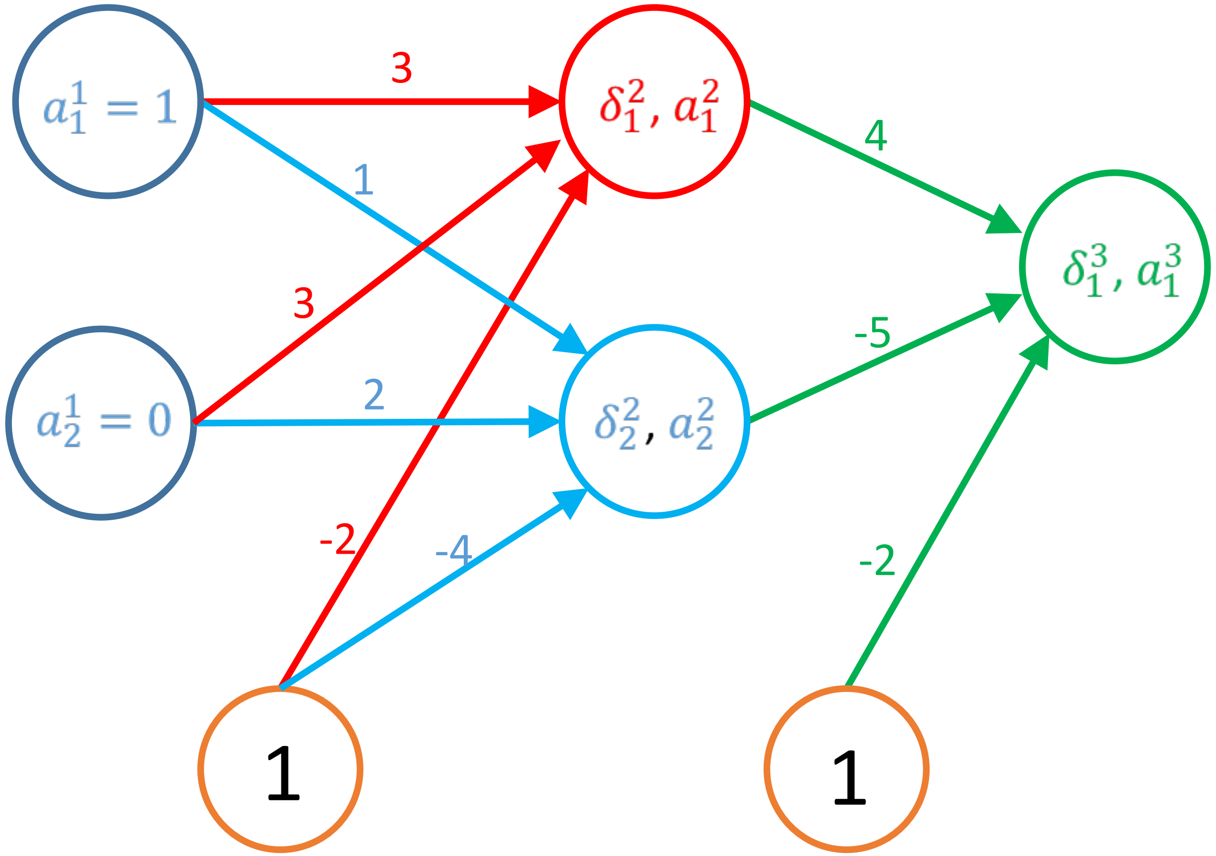
Back-Propagation

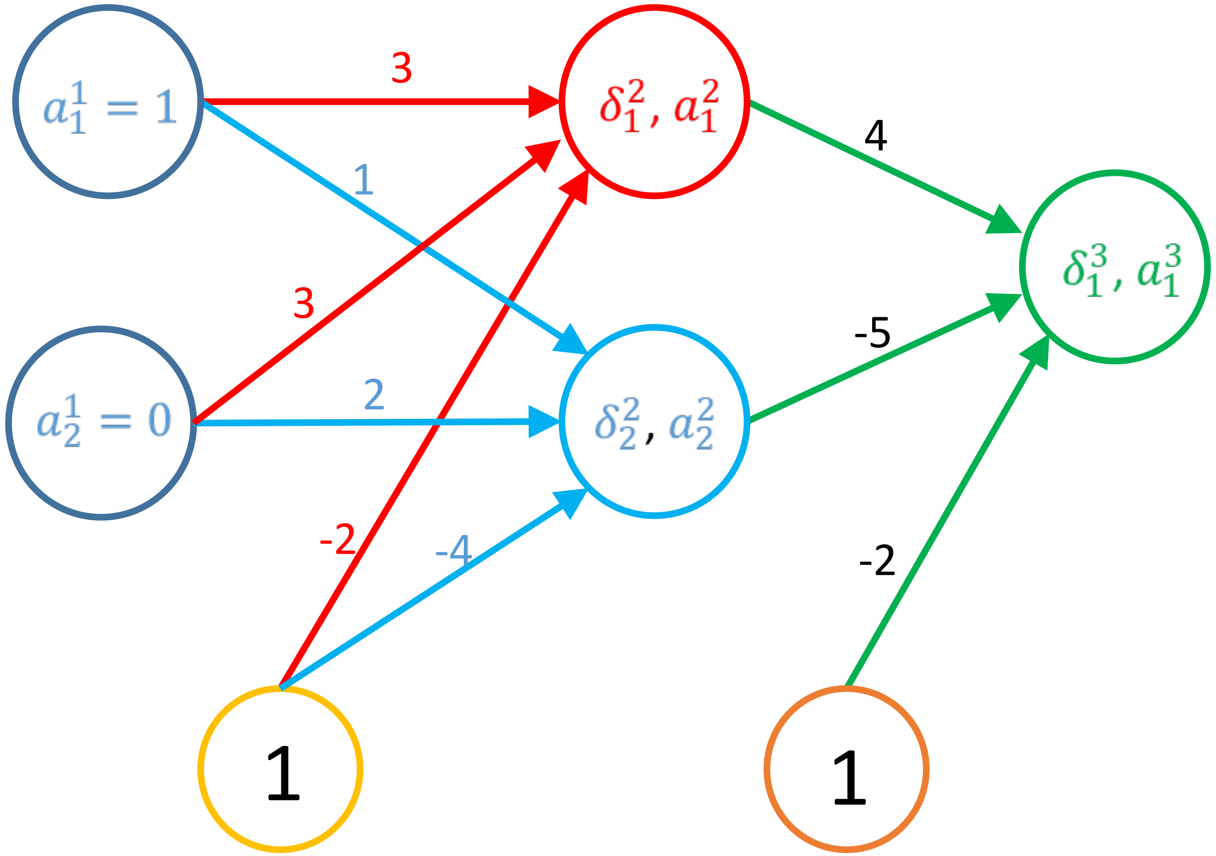
Harpreet Singh (Fall 2023)

Simple Example – Hidden Unit Computations



Forward Pass

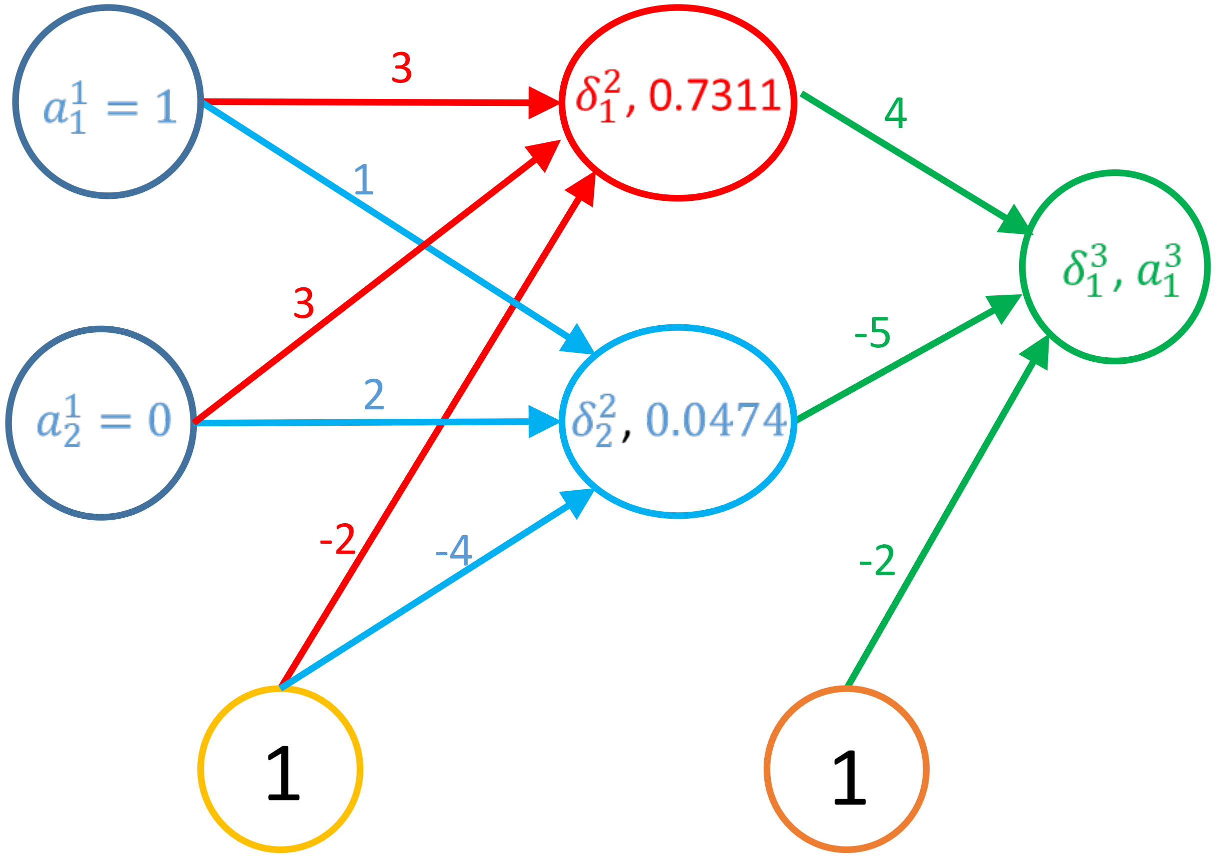
Simple Example – Forward Hidden Layer



$$a_1^2 = \text{sigmoid}(1 * 3 + 0 * 3 - 2)$$
$$= sigmoid(1) = \frac{1}{1 + \mathrm{e}^{-(1)}}$$
$$= 0.7311$$

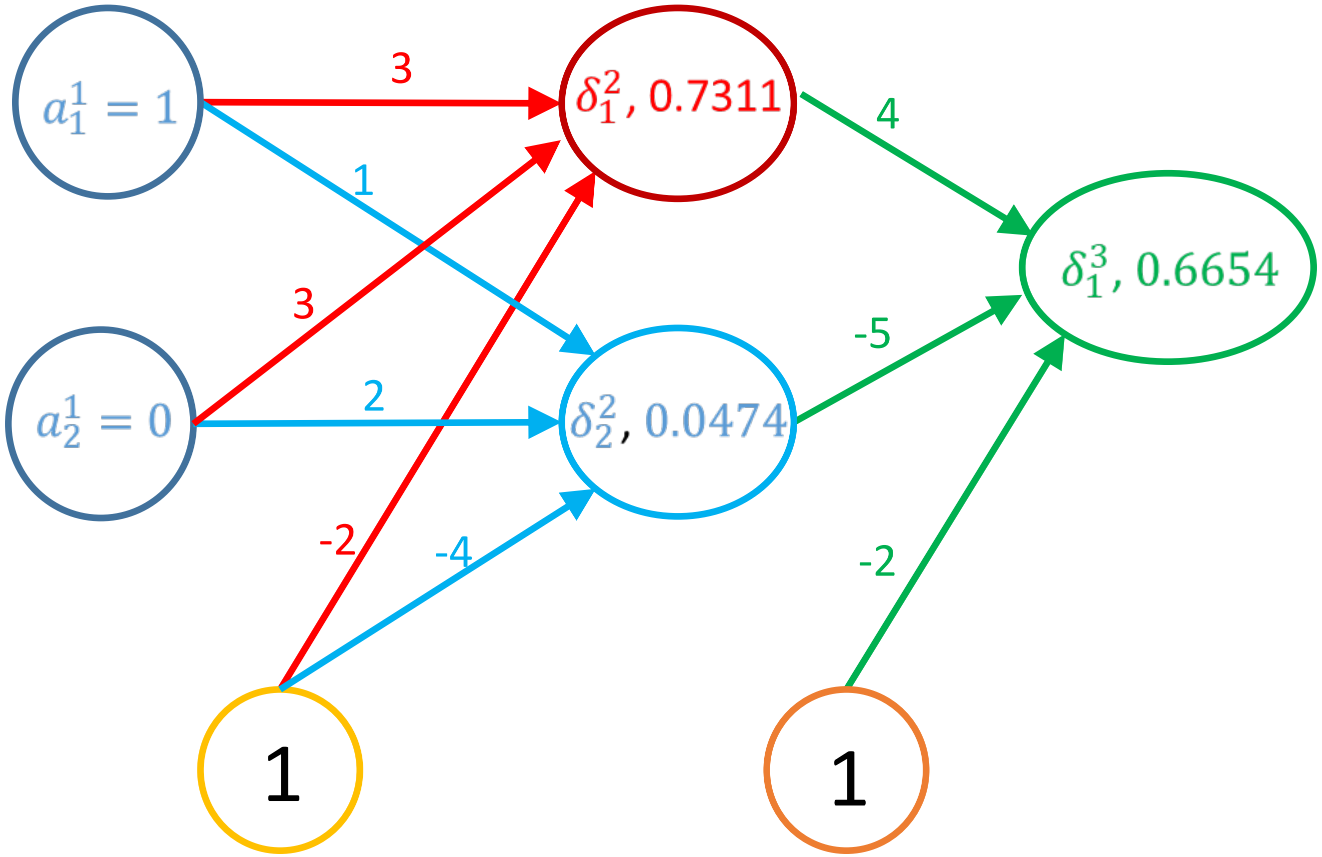
$$a_2^2 = \text{sigmoid}(1 * 1 + 0 * 2 - 4)$$
$$= sigmoid(1) = \frac{1}{1 + \mathrm{e}^{-(-3)}}$$
$$= 0.0474$$

Simple Example – Forward Output Layer



$$a_1^3 = \text{sigmoid}(4 * 0.7311 + (-5) * 0.0474 - 2)$$
$$= sigmoid(0.6874) = \frac{1}{1 + e^{-(0.6874)}} = 0.6654$$

Simple Example – Network after Forward Pass



Backward Pass

Simple Example – Delta Output Layer

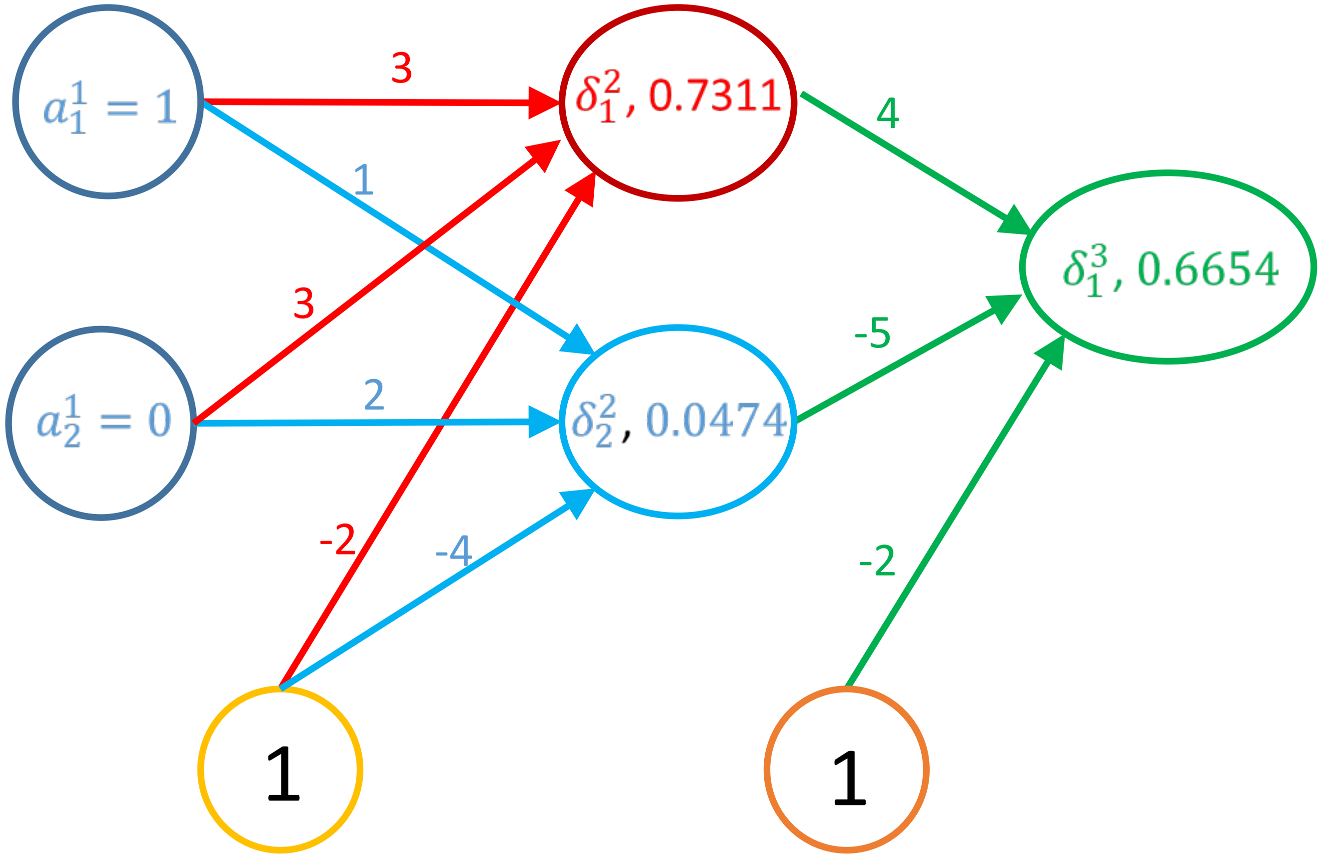
Derivative : sigmoid activation function

$$y = sigmoid(x)$$

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

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

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



$$computed\ output = y = a_1^3 = 0.6654$$

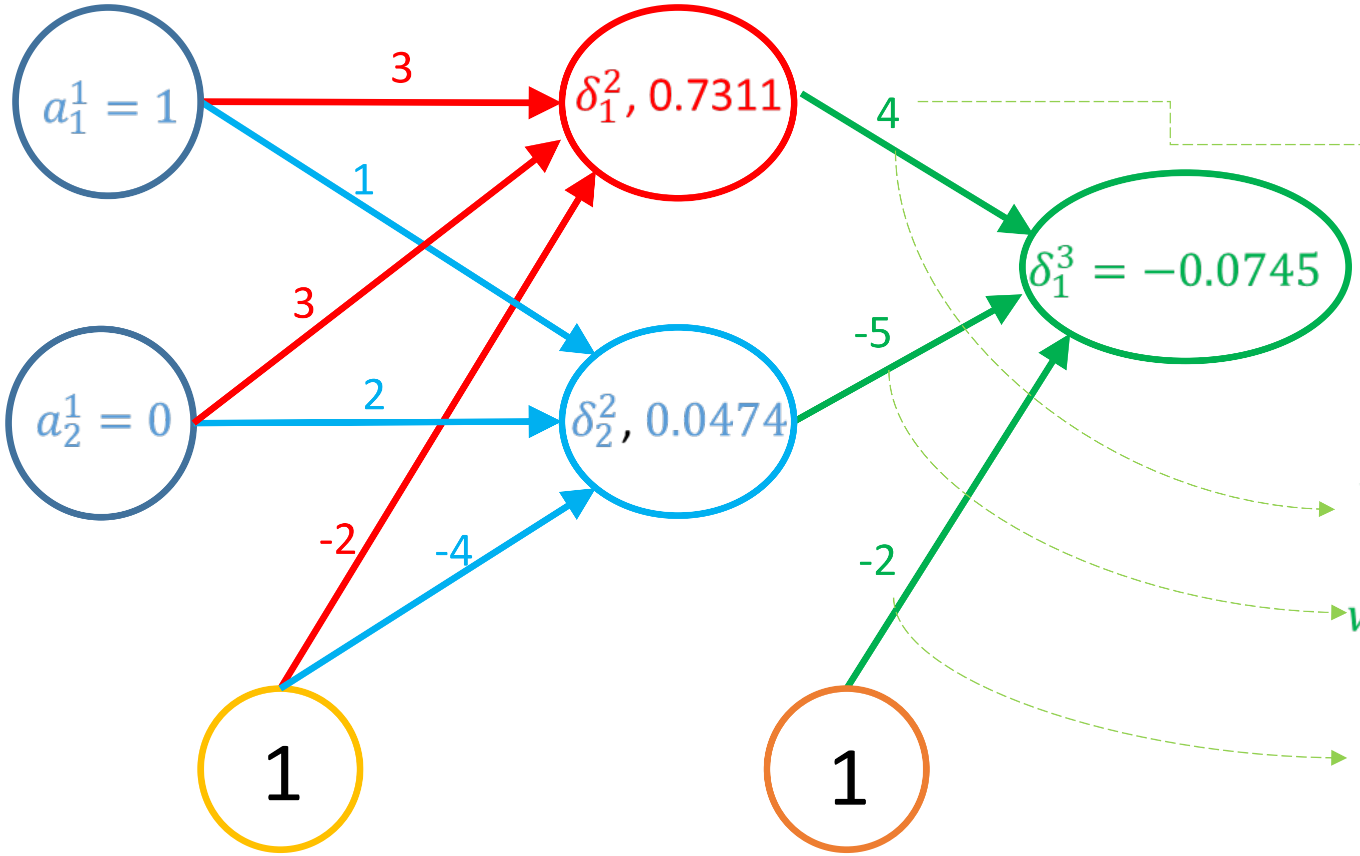
$$correct\ output = t = 1$$

$$error = 0.6654 - 1 = -0.3346$$

$$\begin{aligned} &\delta_1^3 \\ &= error \\ &* derivative\ of\ activation\ function \\ &= error\ * y(1-y) \\ &= (-0.3346) * (0.6654) * (1 - 0.6654) \\ &= -0.0745 \end{aligned}$$

Simple Example – Weight Update Output layer

$$\frac{\partial C}{\partial w} =$$



$$w_{11}^3 = w_{11}^3 - learning\,rate \; * gradient$$

$$learning\,rate = 10$$

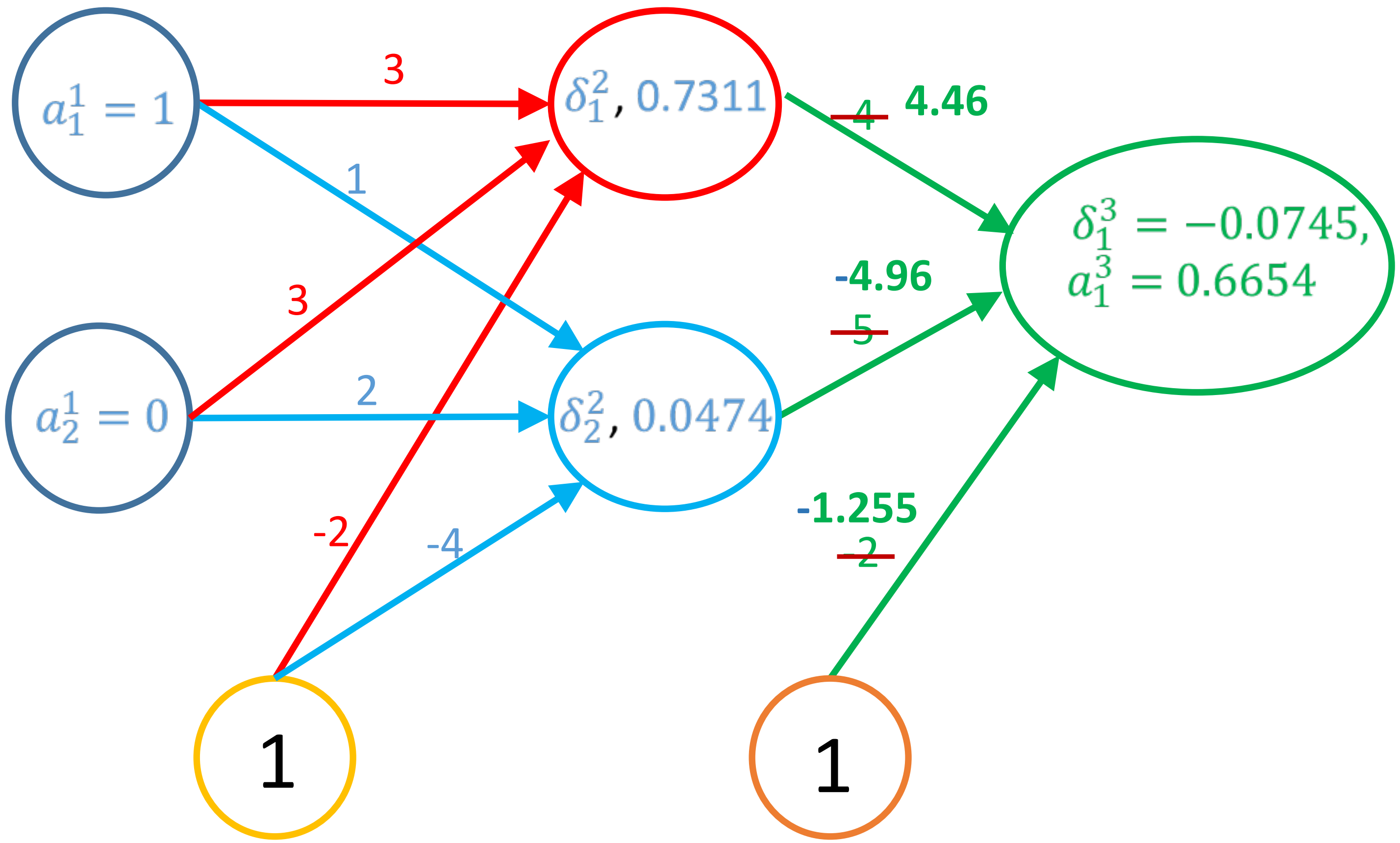
$$gradient \; = 0.7311 * (-0.0745) = -0.05467$$

$$w_{11}^3 = 4 \; - 10 * (-0.054667) = 4.45667$$

$$w_{12}^3 = -5 \; - 10 * (0.0474 * (-0.0745)) = -4.9647$$

$$b_1^3 = -2 \; - 10 * (-0.0745) = -1.255$$

Simple Example – Weight Update Output layer

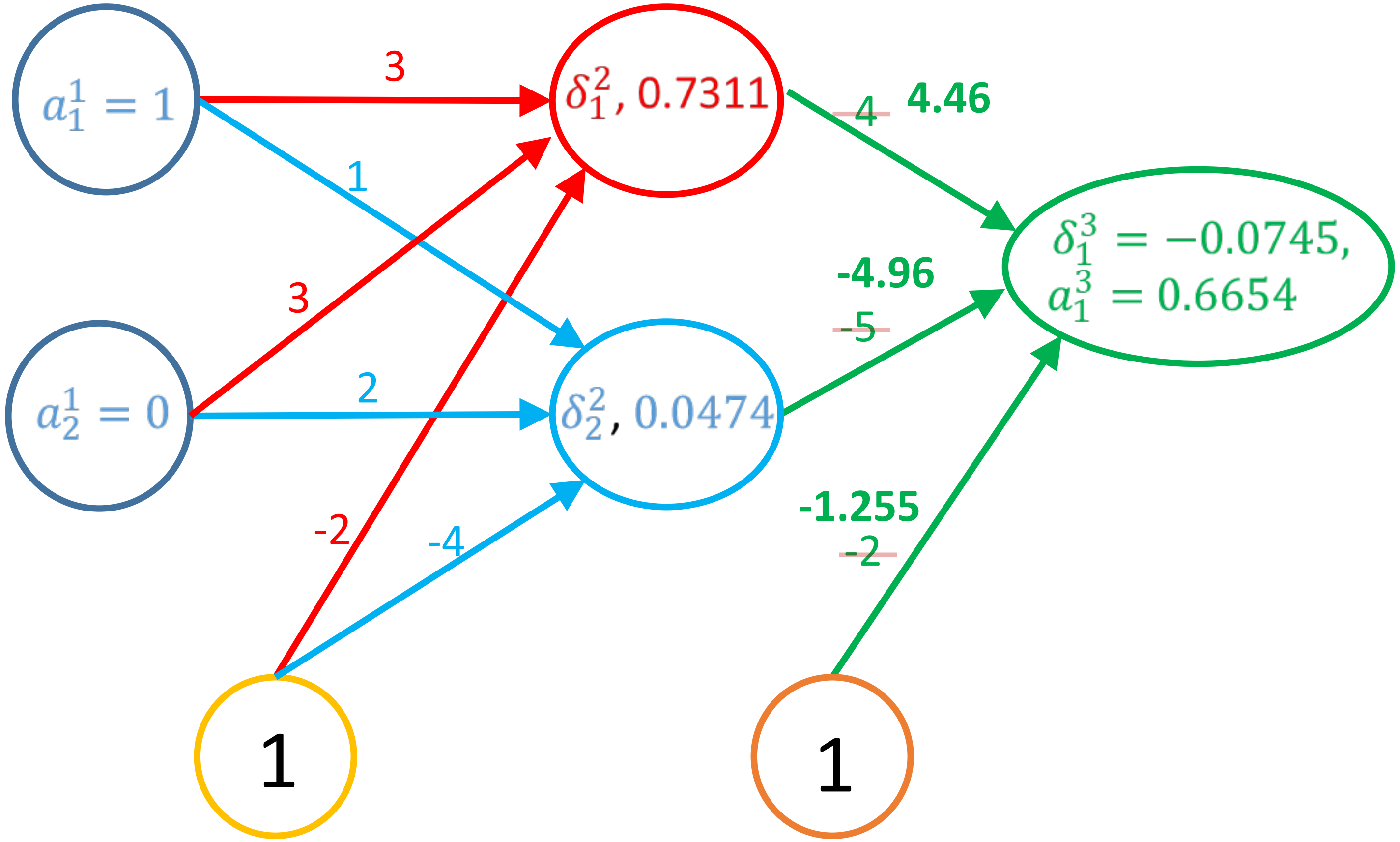


Simple Example – Delta for Hidden Layer

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

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

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



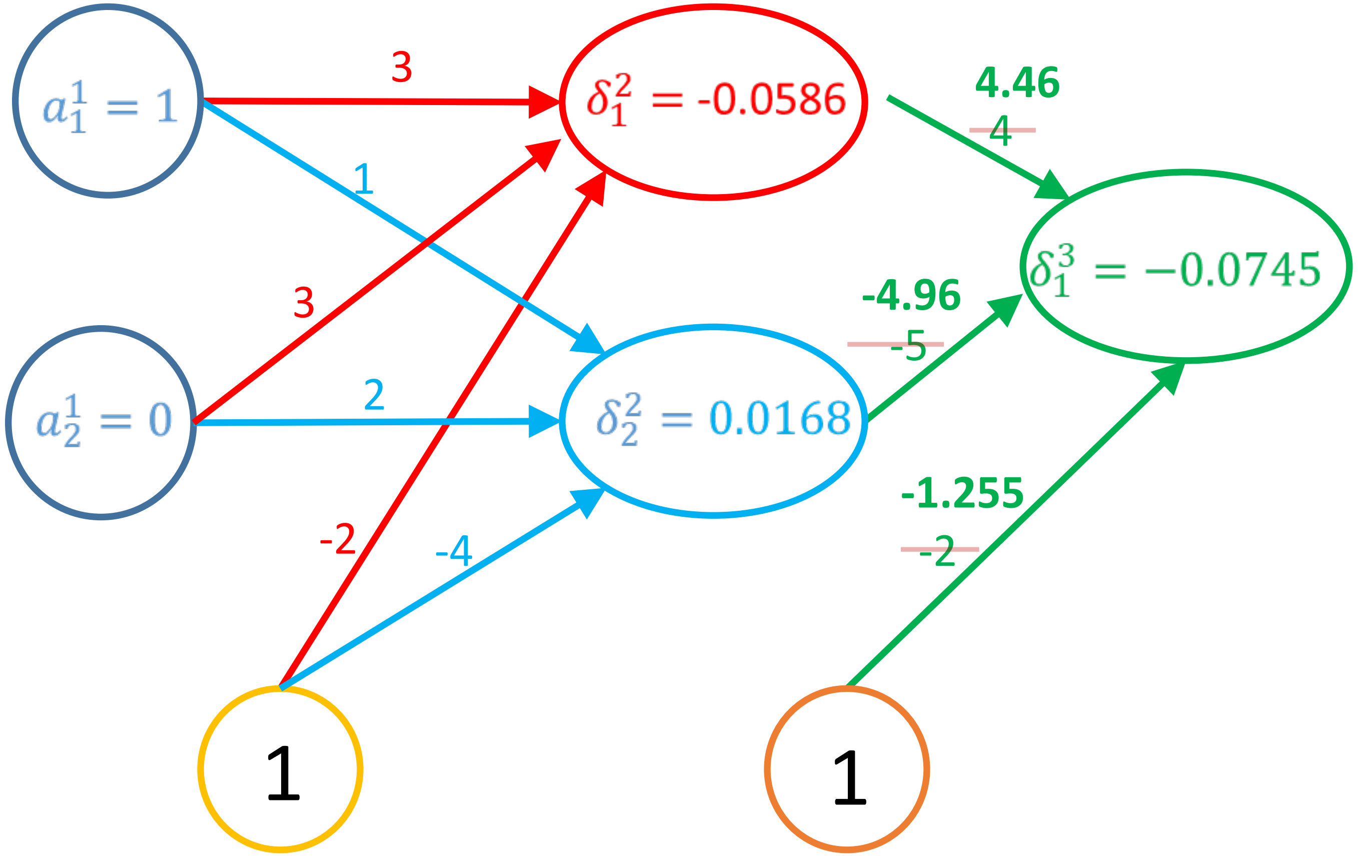
δ_1^2 = *weighted sum of deltas from previous layer ∗ derivative of activation function*

$$\delta_1^2 = 4 * (-0.0745) * 0.7311 * (1 - 0.7311) = -0.0586$$

$$\delta_2^2 = -5 * (-0.0745) * 0.0474 * (1 - 0.0474) = 0.01682$$

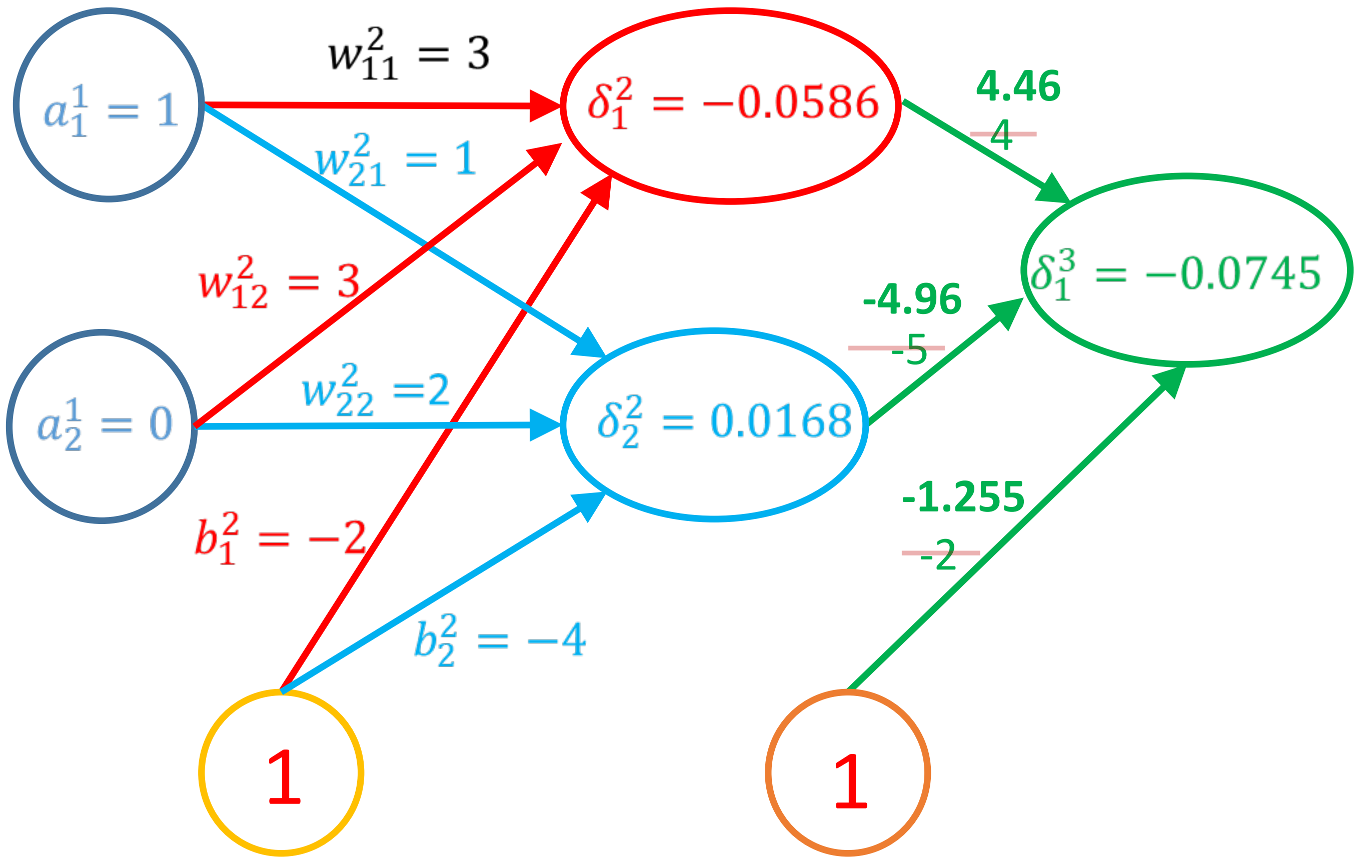
NOTE – We are not using the updated weights to calculate delta for hidden layer. All the calculations are done based on weights from previous iteration.

Simple Example – Delta for Hidden Layer



Simple Example – Weight update for Hidden Layer

gradient = $\frac{\partial C}{\partial w} =$



$$weight = weight - learning\,rate * gradient$$

$$w_{11}^2 = 3 - 10 * (-0.0586 * 1) = 3.586$$

$$w_{12}^2 = 3 - 10 * (-0.0586 * 0) = 3$$

$$b_1^2 = -2 - 10 * (-0.0586) = -1.414$$

$$w_{21}^2 = 1 - 10 * (0.0168 * 1) = 0.832$$

$$w_{22}^2 = 2 - 10 * (0.0168 * 0) = 2$$

$$b_2^2 = -4 - 10 * (0.0168) = -4.168$$

NOTE – We are not using the updated weights to calculate delta for hidden layer. All the calculations are done based on weights from previous iteration.

Simple Example – Network after one iteration