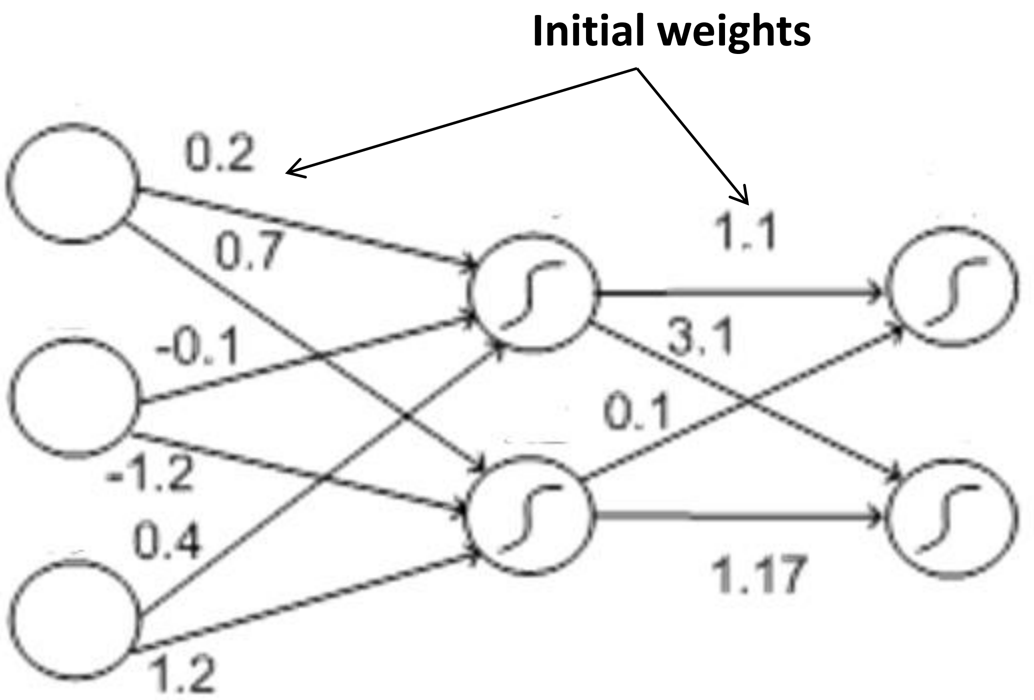
CS409 : Neural Networks (Semester II - 2021/22)

Backpropagation Algorithm – Calculations Example

Dr. Ruwan Nawarathna
Department of Statistics & Computer Science
Faculty of Science
University of Peradeniya

Backpropagation – Example 01



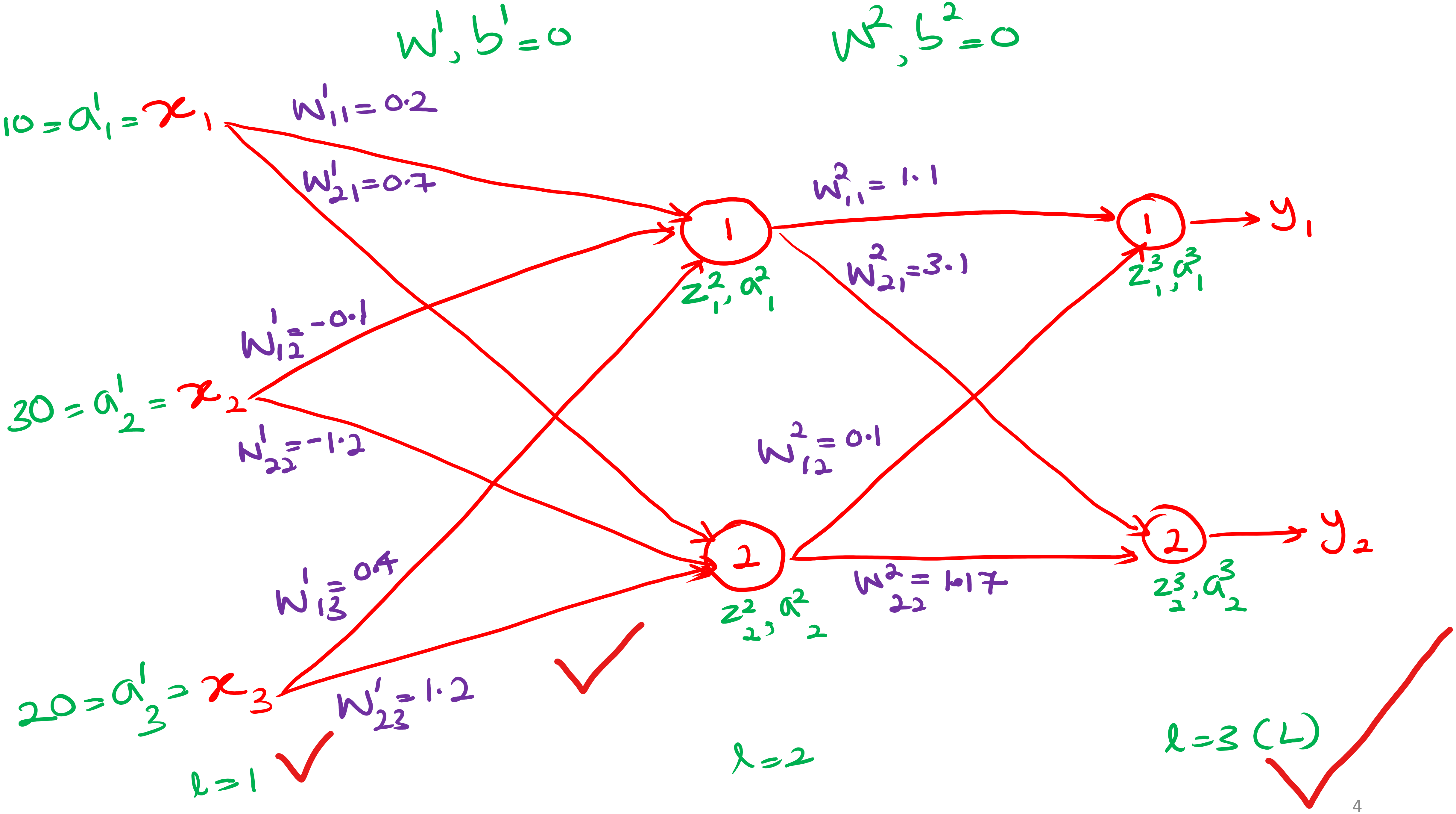
Input Layer with inputs x_1,x_2 and x_3 (no bias units)

Hidden Layer

Output Layer

Backpropagation – Example 01

- Initializations
  - Initialize weights as shown
  - Example (Input) E = (10,30,20)
  - Stipulate that E should have been categorised as O1 (Node 1 of the output layer)
  - Use a learning rate of $\alpha = 0.1$
  - Stochastic gradient descent is used for updating the weights



$W^1, b^1 = 0$
$W^2, b^2 = 0$
$10 = a^1_1 = x_1$
$w^1_{11} = 0.2$
$w^1_{21} = 0.7$
$w^2_{11} = 1.1$
$w^2_{21} = 3.1$
1
1
y_1
z^2_1, a^2_1
z^3_1, a^3_1
$w^1_{12} = -0.1$
$30 = a^1_2 = x_2$
$w^1_{22} = -1.2$
$w^2_{12} = 0.1$
2
2
y_2
$w^1_{13} = 0.4$
z^2_2, a^2_2
$w^2_{22} = 1.17$
z^3_2, a^3_2
$20 = a^1_3 = x_3$
$w^1_{23} = 1.2$
$l = 1$
$l = 2$
$l = 3$ (L)

Forward Propagation - calculating z_j^l and a_j^l of all neurons

$$z_1^2 = 10 \times 0.2 + 30 \times -0.1 + 20 \times 0.4$$
$$= 7$$

$$a_1^2 = \frac{1}{1+e^{-7}}$$

$$\boxed{a_1^2 = 0.999}$$

$$z_2^2 = 10 \times 0.7 + 30 \times -1.2 + 20 \times 1.2$$
$$= 5$$

$$a_2^2 = \frac{1}{1+e^{-5}}$$

$$\boxed{a_2^2 = 0.007}$$

$$z_1^3 = a_1^2 \times 1.1 + a_1^2 \times 0.1$$
$$= (0.999 \times 1.1) + (0.007 \times 0.1)$$
$$= 1.0996$$

$$a_1^3 = \frac{1}{1+e^{-1.0996}}$$

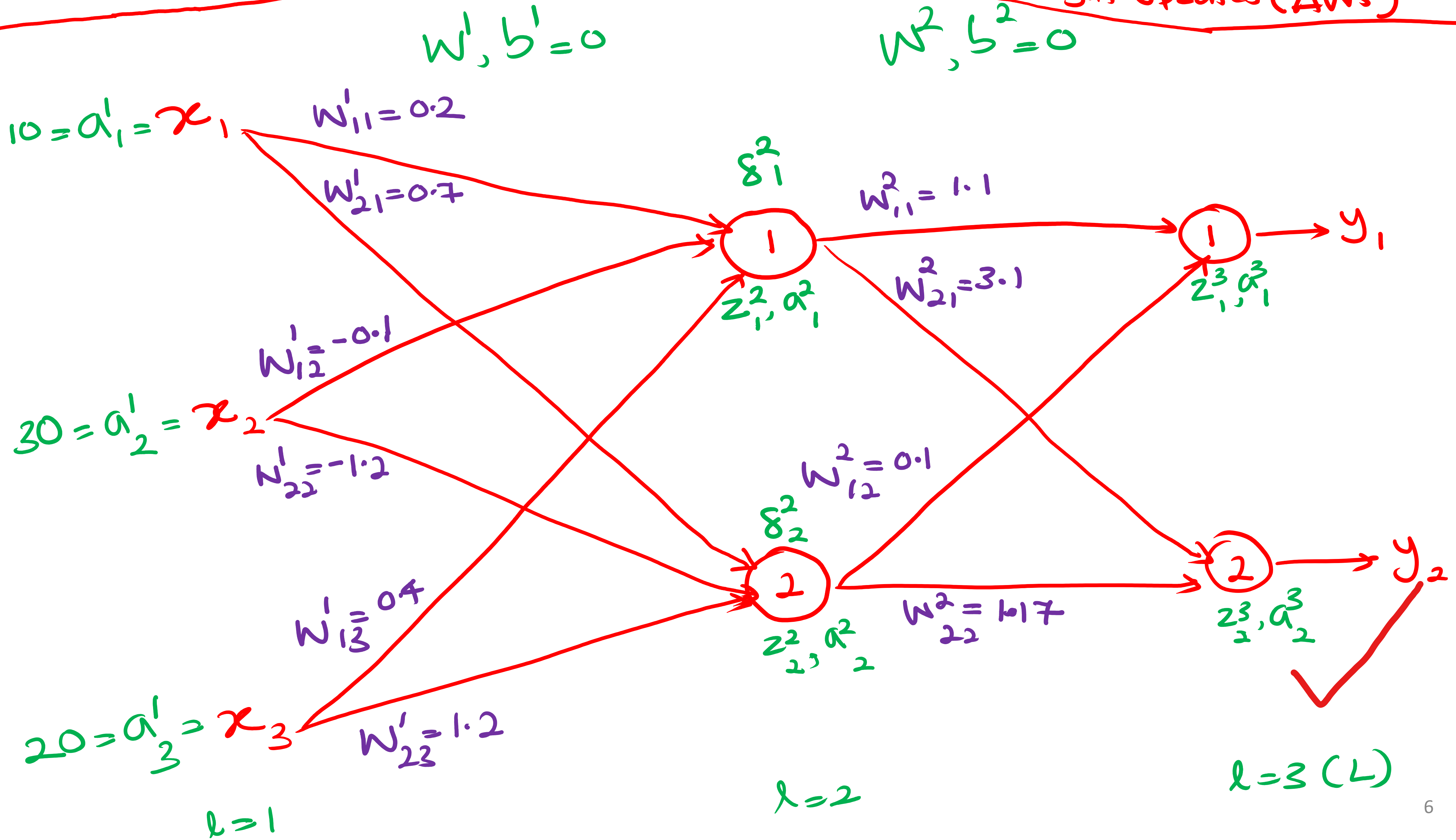
$$\boxed{a_1^3 = 0.750}$$

$$z_2^3 = a_1^2 \times 3.1 + a_2^2 \times 1.17$$
$$= (0.999 \times 3.1) + (0.007 \times 1.17)$$
$$= 3.106$$

$$a_2^3 = \frac{1}{1+e^{-3.106}}$$

$$\boxed{a_2^3 = 0.957}$$

Backward Propagation - Calculating local gradients (δs) and amount of weight updates (ΔWs)



Backward Propagation - calculating δ^l_i and ΔW^l_{ij}

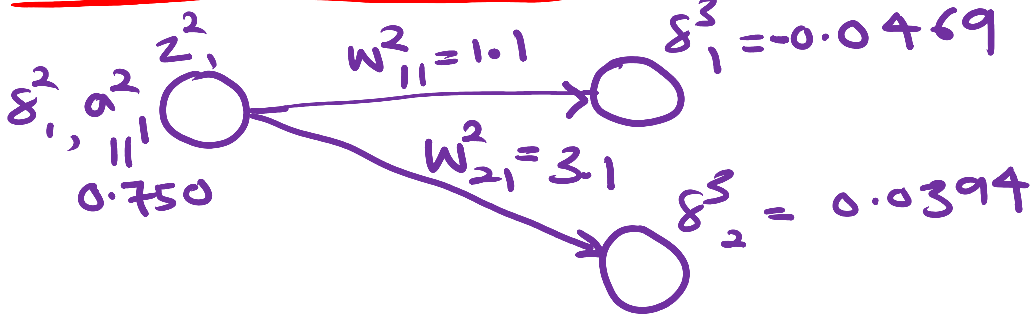
Output Neurons

$$\delta^3_1 = (a^3_1 - y_1)\, g'(z^3_1)$$
$$= (a^3_1 - y_1)\, \sigma'(z^3_1)$$
$$= (a^3_1 - y_1)\, \sigma(z^3_1) \times (1 - \sigma(z^3_1))$$
$$= (a^3_1 - y_1)\,(a^3_1(1 - a^3_1))$$
$$= (0.750 - 1)\,[0.750\,(1 - 0.750)]$$
$$= -0.0469$$

Similarly,

$$\delta^3_2 = (a^3_2 - y_2)\,(a^3_2(1 - a^3_2))$$
$$= (0.957 - 0)\,[0.957(1 - 0.957)]$$
$$= 0.0394$$

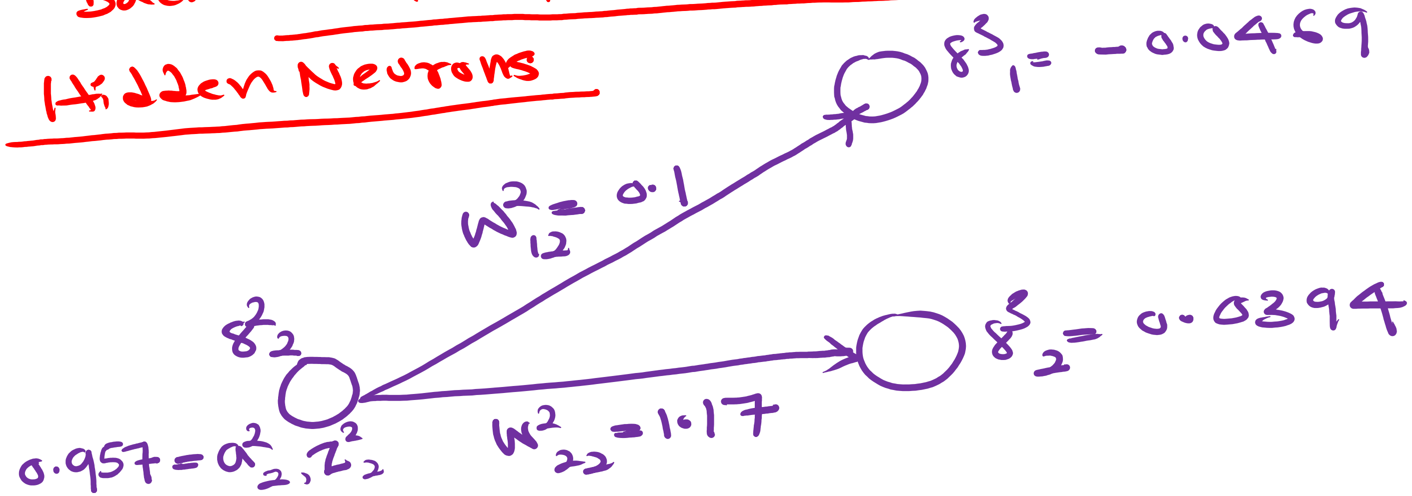
Hidden Neurons



$$\delta^2_1 = [W^2_{11} \times \delta^3_1 + W^2_{21} \times \delta^3_2]\, g'(z^2_1)$$
$$= [W^2_{11} \times \delta^3_1 + W^2_{21} \times \delta^3_2]\, \sigma'(z^2_1)$$
$$= [W^2_{11} \times \delta^3_1 + W^2_{21} \times \delta^3_2]\,[a^2_1(1 - a^2_1)]$$
$$= [1.1 \times -0.0469 + 3.1 \times 0.0394] \times [0.750\,(1 - 0.750)]$$
$$= 0.0132$$

Backward Propagation - Calculating (δ_j^l and ΔW_{ij}^l)

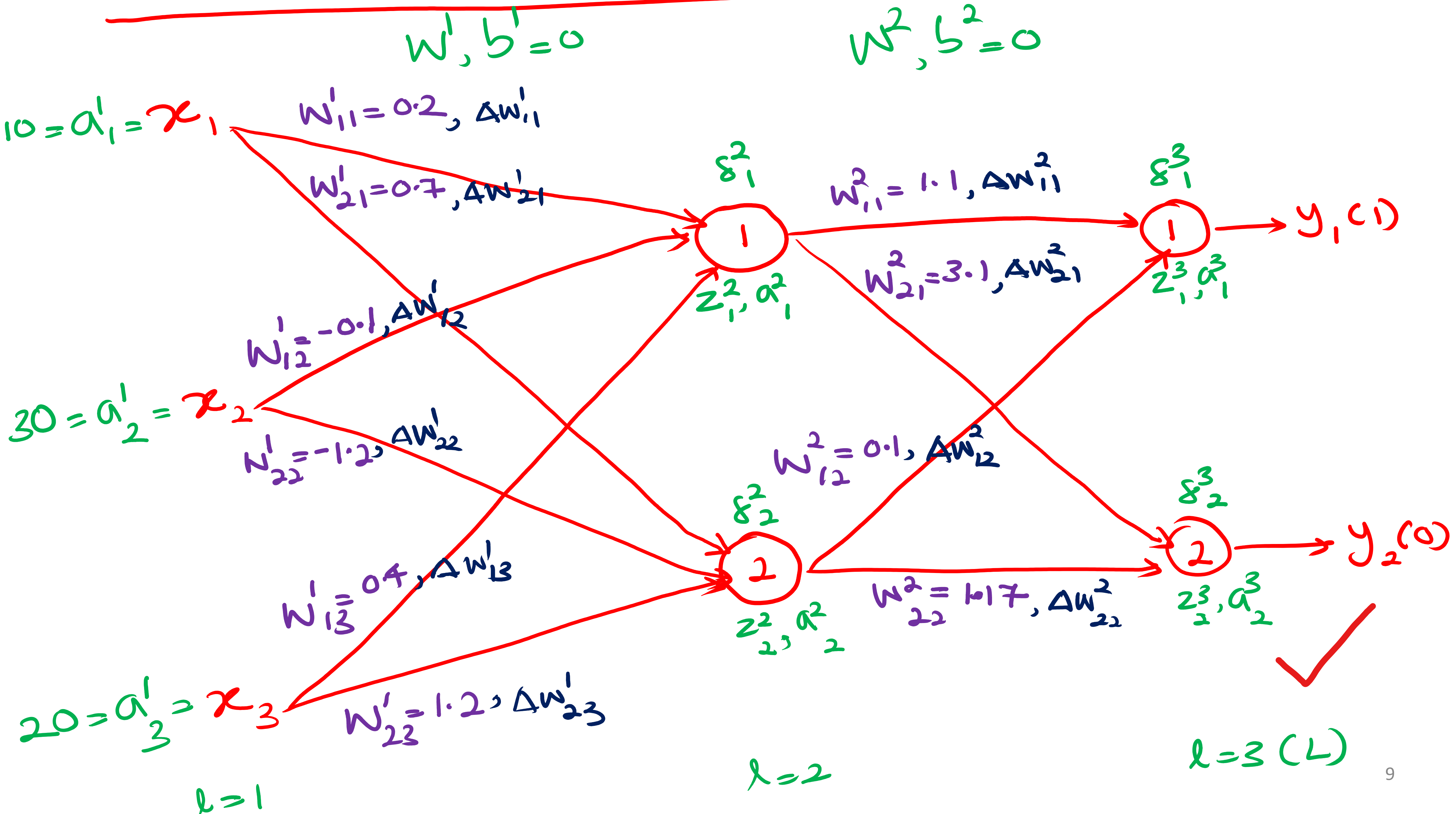
Hidden Neurons



$$\delta_2^2 = [0.1\times -0.0469 + 1.17\times 0.0394][0.957(1-0.957)]$$

$$= 0.0132$$

Calculating Δw^l_{ij} (the amount of weight update for w^l_{ij})



Δw^1_{11}

$$10 = a^1_1 = x_1 \quad (1) \xrightarrow{w^1_{11} = 0.2} (1) \quad \delta^2_1 = 0.0132$$

$$\Delta w^1_{11} = a^1_1 \times \delta^2_1$$
$$= 10 \times 0.0132$$
$$= 0.13200$$

Similarly, the other Δws can be calculated.

Δw^2_{21}

$$0.990 = a^2_1 \quad (1) \xrightarrow{w^2_{21} = 3.1} (2) \quad \delta^3_2 = 0.0394$$

$$\Delta w^2_{12} = a^2_1 \times \delta^3_2$$
$$= 0.990 \times 0.0394$$
$$= 0.03936$$

Amount of weight updates ($\Delta \boldsymbol{W_{ij}^l}$)

Amount of weight Update (ΔW_{ij}^l)	$\Delta W_{ij}^l = a_j^l \delta_i^{l+1}$	
ΔW_{11}^1	10*0.0132	= 0.13200
ΔW_{21}^1	10*0.0017	= 0.01700
ΔW_{12}^1	30*0.0132	= 0.39600
ΔW_{22}^1	30*0.0017	= 0.05100
ΔW_{13}^1	20*0.0132	= 0.26400
ΔW_{23}^1	20*0.0017	= 0.03400
ΔW_{11}^2	0.999*-0.0469	= -0.04685
ΔW_{21}^2	0.999*0.0394	= 0.03936
ΔW_{12}^2	0.007*-0.0469	= -0.00033
ΔW_{22}^2	0.007*0.0394	= 0.00028

Updated weights with stochastic gradient approach ($\boldsymbol{W_{ij}^{l}}$)

If the stochastic gradient descent is used to update the weights, all weights are updated after each example.

e.g.

$$W'_{11} = W'_{11} - \alpha \Delta W'_{11} \qquad (\alpha = 0.1)$$

$$= 0.2 - (0.1 \times 0.132)$$

$$= 0.1868$$

Likewise, all weights should be updated.

Updated weights with stochastic gradient approach ($\boldsymbol{W_{ij}^{l}}$)

Updated weights with stochastic gradient approach (W_{ij}^{l})	$W_{ij}^{l} = W_{ij}^{l} - \alpha \Delta W_{ij}^{l}$	
W_{11}^{1}	0.2 - (0.1*0.13200)	= 0.1868
W_{21}^{1}	0.7 - (0.1*0.01700)	= 0.6983
W_{12}^{1}	(-0.1) - (0.1*0.39600)	= -0.1396
W_{22}^{1}	(-1.2) - (0.1*0.05100)	= -1.2051
W_{13}^{1}	0.4 - (0.1*0.26400)	= 0.3736
W_{23}^{1}	1.2 - (0.1*0.03400)	= 1.2034
W_{11}^{2}	1.1 - (0.1*-0.04685)	= 1.104685
W_{21}^{2}	3.1 - (0.1*0.03936)	= 3.096064
W_{12}^{2}	0.1 - (0.1*-0.00033)	= 0.100033
W_{22}^{2}	1.17 - (0.1*0.00028)	= 1.169972