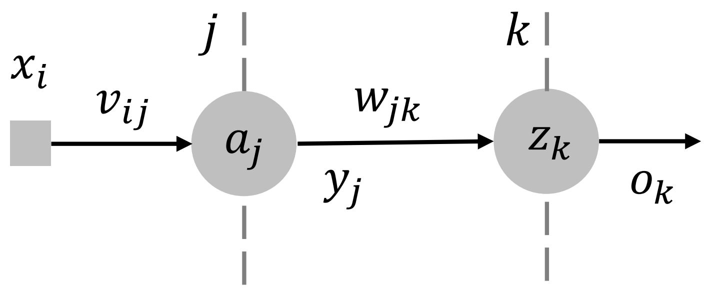
Question 1

- Consider a multilayer perceptron whose **hidden units** use $\mathbf{x^3}$ and whose **output units** use **sin(3x)** as the transfer function, rather than the usual sigmoid or tanh. Using the chain rule, derive the formulas for the weight updates. Your final formulas should be purely algebraic, i.e., they should not contain partial derivatives.



$$a_j = x_i v_{ij}$$

$$y_j = \left(a_j\right)^3$$

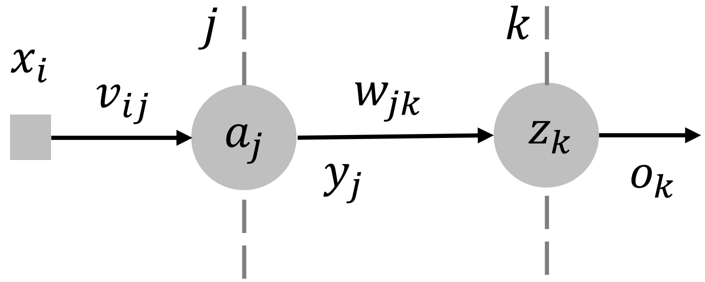
$$z_k = y_j w_{jk}$$

$$o_k = \sin(3z_k)$$

$$e_k = \frac{1}{2}(o_k - t_k)^2$$

Feed forward

Question 1



$$a_j = x_i v_{ij}$$

$$y_j = \left(a_j\right)^3$$

$$z_k = y_j w_{jk}$$

$$o_k = \sin(3z_k)$$

$$e_k = \frac{1}{2}(o_k - t_k)^2$$

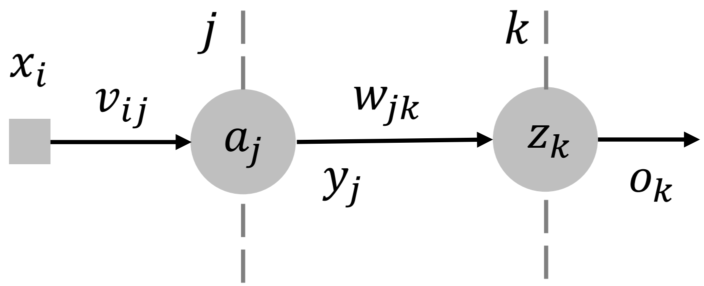
Feed forward

$$e_k = \frac{1}{2}\left(\sin\left(3\left(x_i v_{ij}\right)^3 w_{jk}\right) - t_k\right)^2$$

Given an input x_i and target out t_k

- The error function is variable in v_{ij} and w_{jk}**.** It's required to find set of weights to **minimize** the error function. Which is an **optimization problem**.

Question 1



$$a_j = x_i v_{ij}$$

$$y_j = (a_j)^3$$

$$z_k = y_j w_{jk}$$

$$o_k = \sin(3z_k)$$

$$e_k = \frac{1}{2}(o_k - t_k)^2$$

$$\frac{\partial e}{\partial w_{jk}} = \frac{\partial e}{\partial o_k} \times \frac{\partial o_k}{\partial z_k} \times \frac{\partial z_k}{\partial w_{jk}}$$

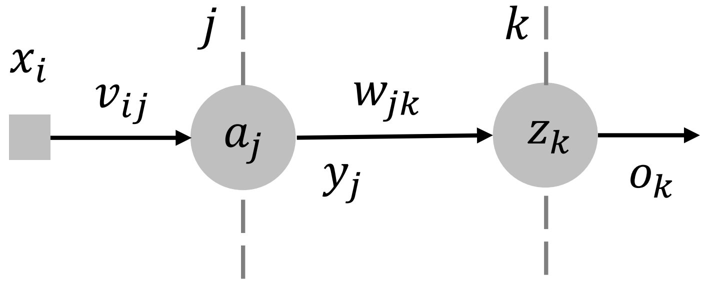
$$\frac{\partial e}{\partial v_{ij}} = \frac{\partial e}{\partial o_k} \times \frac{\partial o_k}{\partial z_k} \times \frac{\partial z_k}{\partial y_j} \times \frac{\partial y_j}{\partial a_j} \times \frac{\partial a_j}{\partial v_{ij}}$$

$$\frac{\partial e}{\partial w_{jk}} \qquad \frac{\partial e}{\partial v_{ij}}$$

$$w_{jk}^{n+1} = w_{jk}^{n} - \eta \frac{\partial e}{\partial w_{jk}}$$

$$v_{ij}^{n+1} = v_{ij}^{n} - \eta \frac{\partial e}{\partial v_{ij}}$$

Question 1



$$a_j = x_i v_{ij}$$

$$y_j = (a_j)^3$$

$$z_k = y_j w_{jk}$$

$$o_k = \sin(3z_k)$$

$$e_k = \frac{1}{2}(o_k - t_k)^2$$

$$\frac{\partial e}{\partial w_{jk}} = \frac{\partial e}{\partial o_k} \times \frac{\partial o_k}{\partial z_k} \times \frac{\partial z_k}{\partial w_{jk}}$$

$$\frac{\partial e}{\partial w_{jk}} = (o_k - t_k) \times 3\cos(3z_k) \times y_j$$

$$\frac{\partial e}{\partial w_{jk}}$$

$$w_{jk}^{n+1} = w_{jk}^{n} - \eta \frac{\partial e}{\partial w_{jk}}$$

$$\frac{\partial e}{\partial o_k} = (o_k - t_k)$$

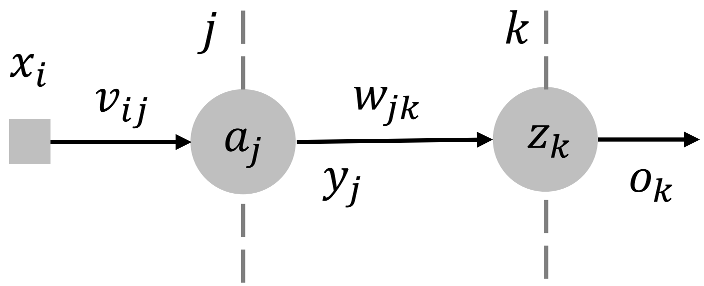
$$\frac{\partial o_k}{\partial z_k} = 3\cos(3z_k)$$

$$\frac{\partial z_k}{\partial w_{jk}} = y_j$$

$$w_{jk}^{n+1} = w_{jk}^{n} - \eta((o_k - t_k) \times 3\cos(3z_k) \times y_j)$$

Learning Rule

Question 1



$$a_j = x_i v_{ij}$$

$$y_j = (a_j)^3$$

$$z_k = y_j w_{jk}$$

$$o_k = \sin(3z_k)$$

$$e_k = \frac{1}{2}(o_k - t_k)^2$$

$$\frac{\partial e}{\partial v_{ij}} = \frac{\partial e}{\partial o_k} \times \frac{\partial o_k}{\partial z_k} \times \frac{\partial z_k}{\partial y_j} \times \frac{\partial y_j}{\partial a_j} \times \frac{\partial a_j}{\partial v_{ij}}$$

$$\frac{\partial e}{\partial v_{ij}} = (o_k - t_k) \times 3\cos(3z_k) \times w_{jk} \times 3(a_j)^2 \times x_i$$

$$v_{ij}^{n+1} = v_{ij}^{n} - \eta \frac{\partial e}{\partial v_{ij}}$$ **Learning Rule**

$$\frac{\partial e}{\partial v_{ij}}$$

$$v_{ij}^{n+1} = v_{ij}^{n} - \eta \frac{\partial e}{\partial v_{ij}}$$

$$\frac{\partial e}{\partial o_k} = (o_k - t_k)$$

$$\frac{\partial o_k}{\partial z_k} = 3\cos(3z_k)$$

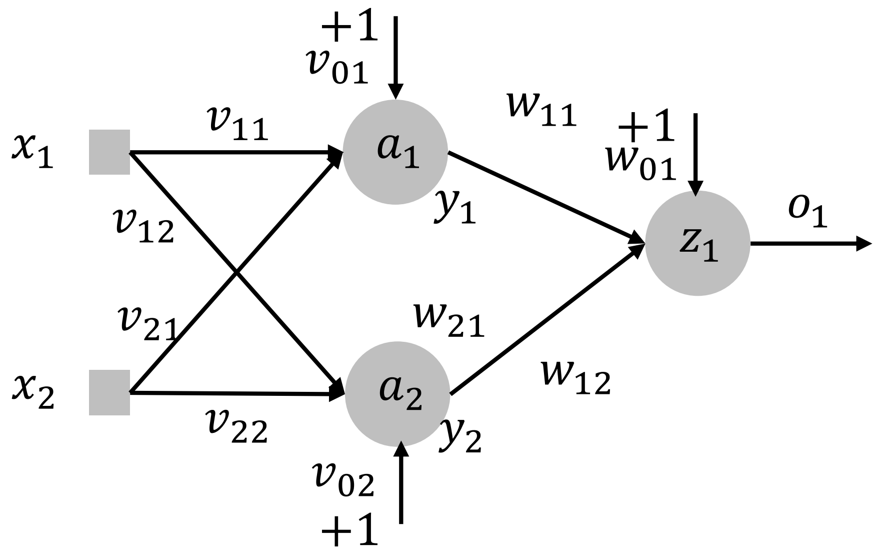
$$\frac{\partial z_k}{\partial y_j} = w_{jk}$$

$$\frac{\partial y_j}{\partial a_j} = 3(a_j)^2$$

$$\frac{\partial a_j}{\partial v_{ij}} = x_i$$

Question 2

Consider a multilayer feedforward network, all the neurons of which operate in their linear regions. Justify the statement that such a network is equivalent to a single-layer feedforward network.



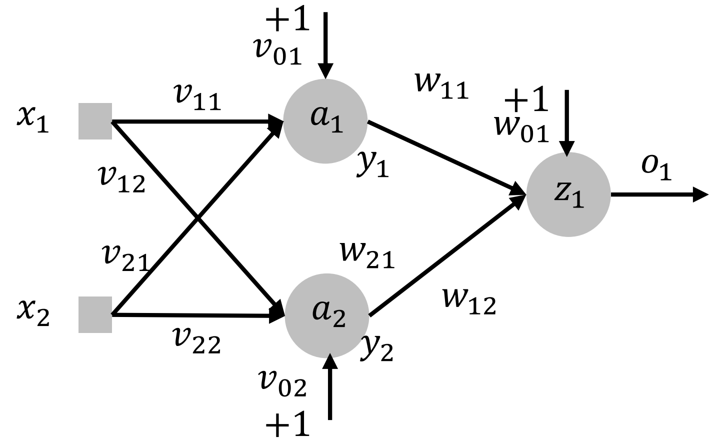
Question 2

Feed forward

$$a_1 = x_1 v_{11} + x_2 v_{21} + v_{01} \qquad y_1 = c_1(a_1)$$

$$a_2 = x_1 v_{21} + x_2 v_{22} + v_{02} \qquad y_2 = c_2(a_2)$$

$$z_1 = y_1 w_{11} + y_2 w_{21} + w_{01} \qquad o_1 = c_3(z_1)$$



$$o_1 = c_3(y_1 w_{11} + y_2 w_{21} + w_{01})$$

$$o_1 = c_3(c_1 w_{11}(a_1) + c_2 w_{21}(a_2) + w_{01})$$

$$o_1 = c_3(c_1 w_{11}(x_1 v_{11} + x_2 v_{21} + v_{01}) + c_2 w_{21}(x_1 v_{21} + x_2 v_{22} + v_{02}) + w_{01})$$

$$o_1 = c_3(\,(x_1 c_1 w_{11} v_{11} + x_2 c_1 w_{11} v_{21} + c_1 w_{11} v_{01}) + (x_1 c_2 w_{21} v_{21} + x_2 c_2 w_{21} v_{22} + c_2 w_{21} v_{02}) + w_{01})$$

$$o_1 = (x_1 c_1 c_3 w_{11} v_{11} + x_2 c_1 c_3 w_{11} v_{21} + c_1 c_3 w_{11} v_{01}) + (x_1 c_2 c_3 w_{21} v_{21} + x_2 c_2 c_3 w_{21} v_{22} + c_2 c_3 w_{21} v_{02}) + c_3 w_{01}$$

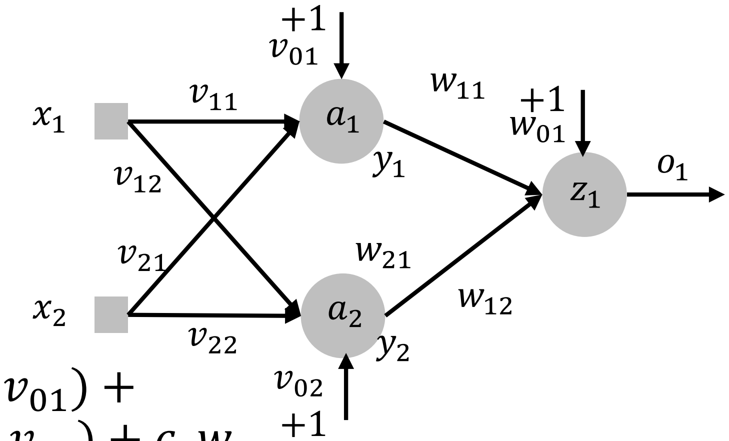
Question 2

Feed forward

$$a_1 = x_1 v_{11} + x_2 v_{21} + v_{01} \qquad y_1 = c_1(a_1)$$

$$a_2 = x_1 v_{21} + x_2 v_{22} + v_{02} \qquad y_2 = c_2(a_2)$$

$$z_1 = y_1 w_{11} + y_2 w_{21} + w_{01} \qquad o_1 = c_3(z_1)$$



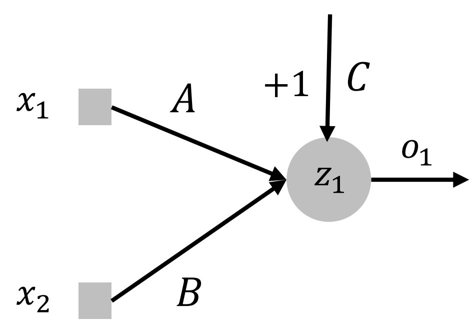
$$o_1 = \left(x_1 c_1 c_3 w_{11} v_{11} + x_2 c_1 c_3 w_{11} v_{21} + c_1 c_3 w_{11} v_{01}\right) + \left(x_1 c_2 c_3 w_{21} v_{21} + x_2 c_2 c_3 w_{21} v_{22} + c_2 c_3 w_{21} v_{02}\right) + c_3 w_{01}$$

$$o_1 = \left(x_1 A + x_2 B + C\right)$$

$$A = c_1 c_3 w_{11} v_{11} + c_2 c_3 w_{21} v_{21}$$

$$B = c_1 c_3 w_{11} v_{21} + c_2 c_3 w_{21} v_{22}$$

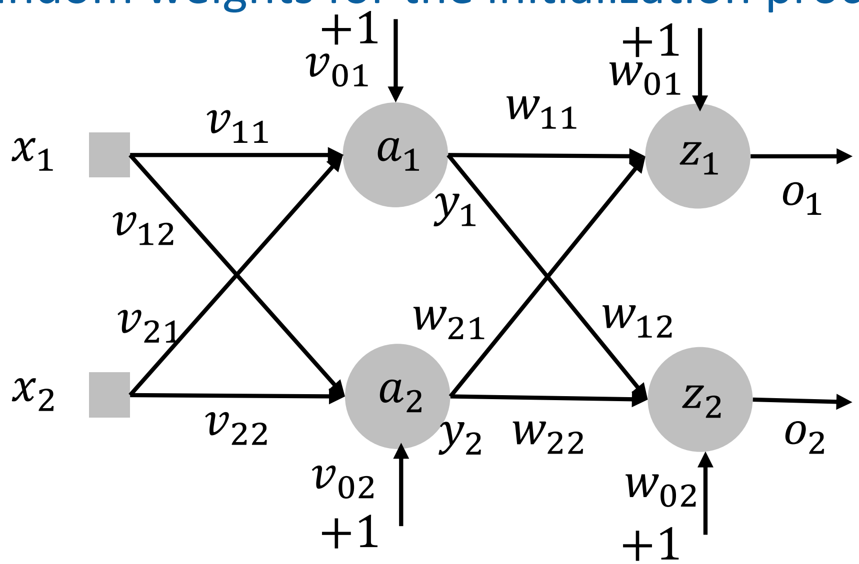
$$C = c_1 c_3 w_{11} v_{01} + c_2 c_3 w_{21} v_{02} + c_3 w_{01}$$



A multilayer network with only linear activation function is equivalent to a single-layer feedforward network.

Question 3

- Given the neural network shown below, carry out one iteration using the back-propagation algorithm using a learning rate of **0.5** and a **sigmoid activation function**. Your network is supposed to learn a pattern **x1=1** and **x2=0.1** with a desired output **o1=0.6** and **o2=0.01**. Assume random weights for the initialization process.



$$x_1 = 1$$

$$x_2 = 0.1$$

$$t_1 = 0.6$$

$$t_2 = 0.01$$

$$\eta = 0.5$$

$$\sigma(x) = \frac{1}{1 - e^{-x}}$$

$$\sigma'(x) = \sigma(x) \times (1 - \sigma(x))$$

Question 3

Feed forward

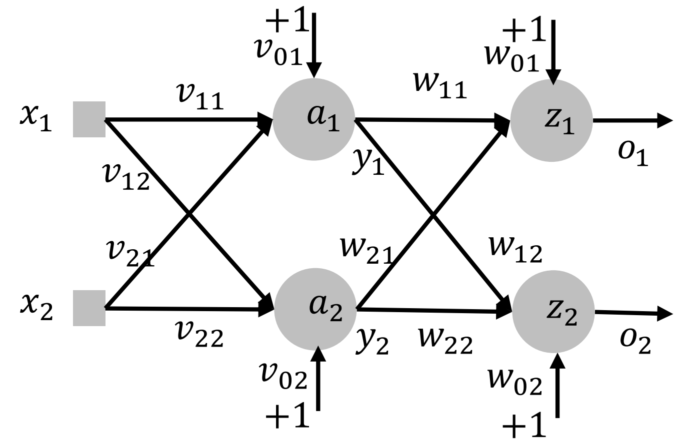
$$a_1 = x_1 v_{11} + x_2 v_{21} + v_{01} \qquad y_1 = \sigma(a_1)$$

$$a_2 = x_1 v_{21} + x_2 v_{22} + v_{02} \qquad y_2 = \sigma(a_2)$$

$$z_1 = y_1 w_{11} + y_2 w_{21} + w_{01} \qquad o_1 = \sigma(z_1)$$

$$z_2 = y_1 w_{12} + y_2 w_{22} + w_{02} \qquad o_2 = \sigma(z_2)$$

$$e_1 = 0.5(o_1 - t_1)^2 \qquad e_2 = 0.5(o_2 - t_2)^2$$



$$x_1 = 1 \qquad x_2 = 0.1$$

$$t_1 = 0.6 \qquad t_2 = 0.01$$

$$v_{12} = v_{21} = 1 \quad , v_{11} = v_{22} = v_{01} = v_{02} = 0.5$$

$$w_{12} = w_{21} = 1 \quad , w_{11} = w_{22} = w_{01} = w_{02} = 0.5$$

$$a_1 = 1.1 \qquad y_1 = 0.75026 \qquad z_1 = 1.7 \qquad o_1 = 0.84554$$

$$a_2 = 1.55 \qquad y_2 = 0.82491 \qquad a_2 = 1.663 \qquad o_2 = 0.84060$$

$$e_1 = 0.5(0.84554 - 0.6)^2 = 0.0301$$

$$e_2 = 0.5(0.84060 - 0.01)^2 = 0.34495$$

Question 3

Backward Propagation

$e_1 = 0.5(o_1 - t_1)^2 \qquad e_2 = 0.5(o_2 - t_2)^2$

$\partial e_1/\partial o_1 = o_1 - t_1 \qquad \partial e_2/\partial o_2 = o_2 - t_2$

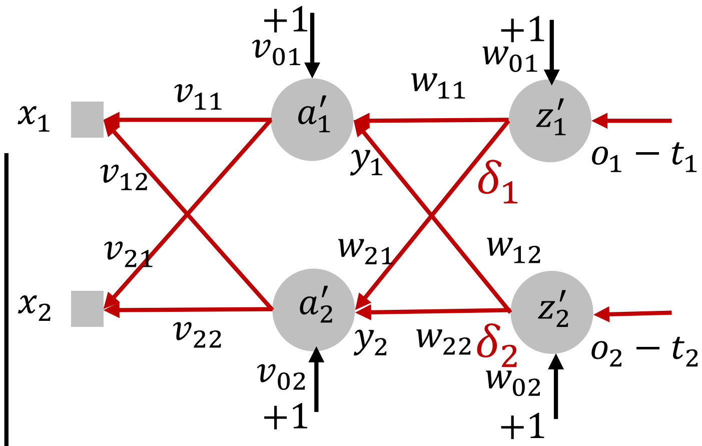
$\delta_1 = (o_1 - t_1) \times (\partial o_1/\partial z_1)$

$\delta_1 = (o_1 - t_1) \times o_1 \times (1 - o_1)$

$\delta_2 = (o_2 - t_2) \times o_2 \times (1 - o_2)$

$$\frac{\partial e_1}{\partial w_{11}} = \delta_1 \times y_1 \quad \frac{\partial e_1}{\partial w_{21}} = \delta_1 \times y_2 \quad \frac{\partial e_1}{\partial w_{01}} = \delta_1 \times 1$$

$$\frac{\partial e_2}{\partial w_{12}} = \delta_2 \times y_1 \quad \frac{\partial e_2}{\partial w_{22}} = \delta_2 \times y_2 \quad \frac{\partial e_2}{\partial w_{02}} = \delta_2 \times 1$$



$x_1 = 1 \qquad x_2 = 0.1$

$t_1 = 0.6 \qquad t_2 = 0.01$

$\sigma'(x) = \sigma(x) \times (1 - \sigma(x))$

$o_1 = \sigma(z_1) \qquad o_2 = \sigma(z_2)$

$z_1 = y_1 w_{11} + y_2 w_{21} + w_{01}$

$z_2 = y_1 w_{12} + y_2 w_{22} + w_{02}$

$y_1 = \sigma(a_1) \qquad y_2 = \sigma(a_2)$

$a_1 = x_1 v_{11} + x_2 v_{21} + v_{01}$

$a_2 = x_1 v_{21} + x_2 v_{22} + v_{02}$

Question 3

Backward Propagation

$$\frac{\partial e_1}{\partial w_{11}} = \delta_1 \times y_1 \quad \frac{\partial e_1}{\partial w_{21}} = \delta_1 \times y_2 \quad \frac{\partial e_1}{\partial w_{01}} = \delta_1 \times 1$$

$$\frac{\partial e_2}{\partial w_{12}} = \delta_2 \times y_1 \quad \frac{\partial e_2}{\partial w_{22}} = \delta_2 \times y_2 \quad \frac{\partial e_2}{\partial w_{02}} = \delta_2 \times 1$$

$$\delta_1 = 0.0321 \quad y_1 = 0.75026$$

$$\delta_2 = 0.1113 \quad y_2 = 0.82491$$

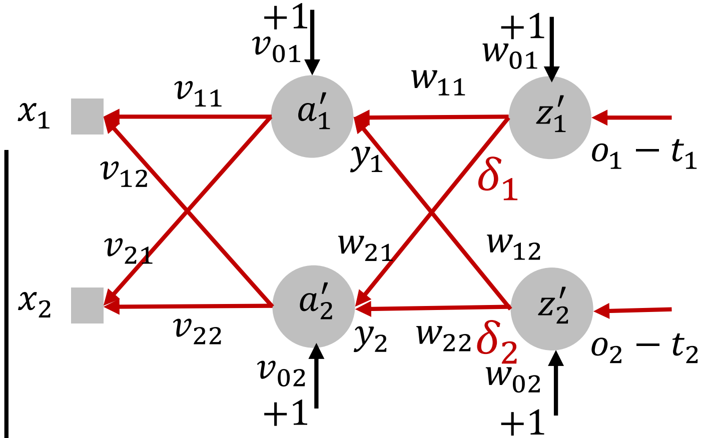
$$w_{11}^2 = w_{11}^1 - 0.5\frac{\partial e_1}{\partial w_{11}}$$

$$w_{11}^2 = 0.5 - 0.5(0.0241) = 0.488$$

$$w_{21}^2 = 1 - 0.5(0.0265) = 0.9868$$

$$w_{01}^2 = 0.5 - 0.5(0.0321) = 0.484$$

$$w_{12}^2 = 0.958 \quad w_{22}^2 = 0.4541 \quad w_{02}^2 = 0.4443$$



$$x_1 = 1 \qquad x_2 = 0.1$$

$$t_1 = 0.6 \qquad t_2 = 0.01$$

$$\sigma'(x) = \sigma(x) \times (1 - \sigma(x))$$

$$o_1 = \sigma(z_1) \qquad o_2 = \sigma(z_2)$$

$$z_1 = y_1 w_{11} + y_2 w_{21} + w_{01}$$

$$z_2 = y_1 w_{12} + y_2 w_{22} + w_{02}$$

$$y_1 = \sigma(a_1) \qquad y_2 = \sigma(a_2)$$

$$a_1 = x_1 v_{11} + x_2 v_{21} + v_{01}$$

$$a_2 = x_1 v_{21} + x_2 v_{22} + v_{02}$$

Question 3

Backward Propagation

$\delta_1 = 0.0321 \quad y_1 = 0.75026$

$\delta_2 = 0.1113 \quad y_2 = 0.82491$

$$\frac{\partial e}{\partial y_1} = \delta_1 \times w_{11} + \delta_2 \times w_{12} = 0.1273$$

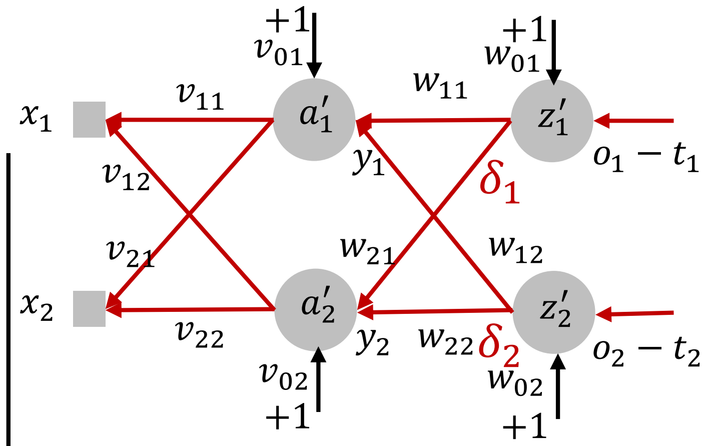
$$\frac{\partial e}{\partial y_2} = \delta_1 \times w_{21} + \delta_2 \times w_{22} = 0.0877$$

$$\frac{\partial e}{\partial a_1} = \frac{\partial e}{\partial y_1} \times y_1 \times (1 - y_1) = 0.02385$$

$$\frac{\partial e}{\partial a_2} = \frac{\partial e}{\partial y_2} \times y_2 \times (1 - y_2) = 0.01267$$

$$\frac{\partial e}{\partial v_{11}} = \frac{\partial e}{\partial a_1} \times x_1 \qquad \frac{\partial e}{\partial v_{21}} = \frac{\partial e}{\partial a_1} \times x_2 \qquad \frac{\partial e}{\partial v_{01}} = \frac{\partial e}{\partial a_1} \times 1$$

$$\frac{\partial e}{\partial v_{12}} = \frac{\partial e}{\partial a_2} \times x_1 \qquad \frac{\partial e}{\partial v_{22}} = \frac{\partial e}{\partial a_2} \times x_2 \qquad \frac{\partial e}{\partial v_{02}} = \frac{\partial e}{\partial a_2} \times 1$$



$x_1 = 1 \qquad x_2 = 0.1$

$t_1 = 0.6 \qquad t_2 = 0.01$

$$\sigma'(x) = \sigma(x) \times (1 - \sigma(x))$$

$o_1 = \sigma(z_1) \qquad o_2 = \sigma(z_2)$

$z_1 = y_1 w_{11} + y_2 w_{21} + w_{01}$

$z_2 = y_1 w_{12} + y_2 w_{22} + w_{02}$

$y_1 = \sigma(a_1) \qquad y_2 = \sigma(a_2)$

$a_1 = x_1 v_{11} + x_2 v_{21} + v_{01}$

$a_2 = x_1 v_{21} + x_2 v_{22} + v_{02}$

Question 3

Backward Propagation

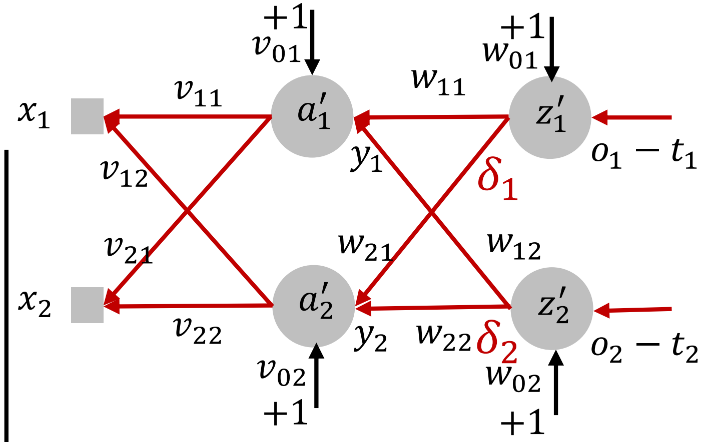
$$\frac{\partial e}{\partial v_{11}} = \frac{\partial e}{\partial a_1} \times x_1 \qquad \frac{\partial e}{\partial v_{21}} = \frac{\partial e}{\partial a_1} \times x_2 \qquad \frac{\partial e}{\partial v_{01}} = \frac{\partial e}{\partial a_1} \times 1$$

$$\frac{\partial e}{\partial v_{12}} = \frac{\partial e}{\partial a_2} \times x_1 \qquad \frac{\partial e}{\partial v_{22}} = \frac{\partial e}{\partial a_2} \times x_2 \qquad \frac{\partial e}{\partial v_{02}} = \frac{\partial e}{\partial a_2} \times 1$$

$$v_{11}^2 = v_{11}^1 - \eta \frac{\partial e}{\partial v_{11}}$$

$$v_{11}^2 = 0.4881 \qquad v_{12}^2 = 0.9937 \qquad v_{10}^2 = 0.4881$$

$$v_{21}^2 = 0.9988 \qquad v_{22}^2 = 0.4994 \qquad v_{20}^2 = 0.4973$$



$$x_1 = 1 \qquad x_2 = 0.1$$

$$t_1 = 0.6 \qquad t_2 = 0.01$$

$$\sigma'(x) = \sigma(x) \times (1 - \sigma(x))$$

$$o_1 = \sigma(z_1) \qquad o_2 = \sigma(z_2)$$

$$z_1 = y_1 w_{11} + y_2 w_{21} + w_{01}$$

$$z_2 = y_1 w_{12} + y_2 w_{22} + w_{02}$$

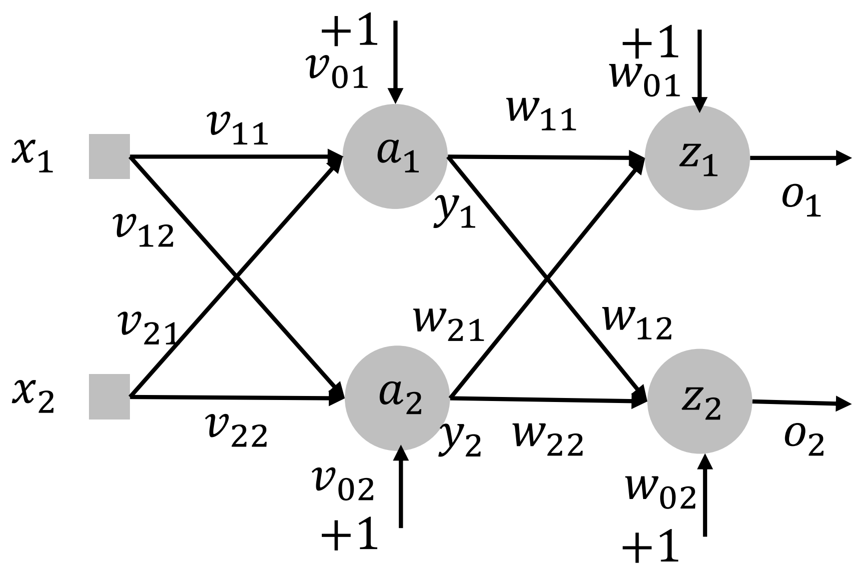
$$y_1 = \sigma(a_1) \qquad y_2 = \sigma(a_2)$$

$$a_1 = x_1 v_{11} + x_2 v_{21} + v_{01}$$

$$a_2 = x_1 v_{21} + x_2 v_{22} + v_{02}$$

Question 4

4- Derive the learning rule for the network of problem 3 but using a tanh activation function.



$x_1 = 1$

$x_2 = 0.1$

$t_1 = 0.6$

$t_2 = 0.01$

$\eta = 0.5$

$y = \tanh(x)$

$\tanh'(x) = (1 - y^2)$