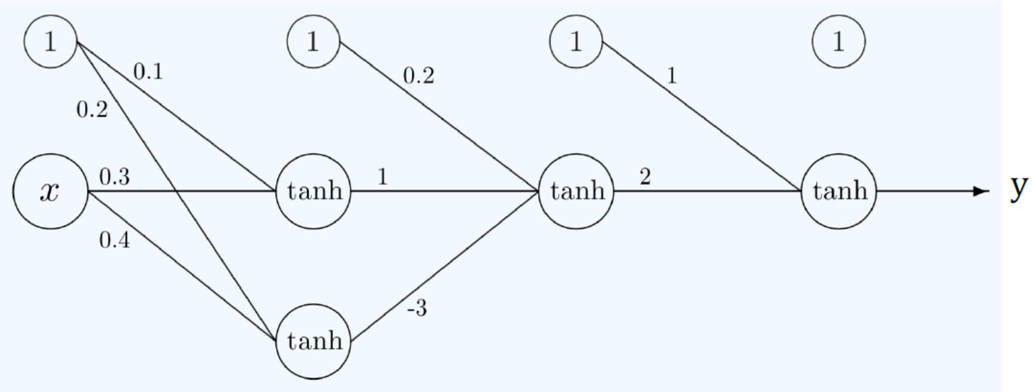

(a) (10%) Using the **half** of the sum square as our error function, derive and compute $\delta^{(3)}, \delta^{(2)}, \delta^{(1)}$。

$$W^{(1)} = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix} 、 W^{(2)} = \begin{bmatrix} 0.2 \\ 1 \\ -3 \end{bmatrix} 、 W^{(3)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$x^{(0)}$	$s^{(1)}$	$x^{(1)}$	$s^{(2)}$	$x^{(2)}$	$s^{(3)}$	$x^{(3)}$
$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 0.7 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0.604368 \\ 0.761594 \end{bmatrix}$	$[-1.48041]$	$\begin{bmatrix} 1 \\ -0.901546 \end{bmatrix}$	$[-0.803091]$	$[-0.665761]$

Error function:

Half of the sum square

$$\delta^{(L)} = (x^{(L)} - y)\theta'(S^{(L)})$$

$\delta^{(3)} = \left(x^{(3)} - y\right) \times [1 - \left(x^{(3)} \otimes x^{(3)}\right)]_1^{d(3)}$ $= (-0.665761 - 1) \times [1 - (-0.665761)^2]$ $= -0.927432$
$\delta^{(2)} = [1 - \left(x^{(2)} \otimes x^{(2)}\right)] \otimes [W^{(3)} \times \delta^{(3)}]_1^{d(2)}$ $= [1 - (-0.901546)^2][2 \times (-0.927432)]$ $= -0.347259$
$\delta^{(1)} = [1 - \left(x^{(1)} \otimes x^{(1)}\right)] \otimes [W^{(2)} \times \delta^{(2)}]_1^{d(1)}$ $= \begin{bmatrix} [1 - (0.604368)^2][1 \times (-0.347259)] \\ [1 - (0.761594)^2][-3 \times (-0.347259)] \end{bmatrix}$ $= \begin{bmatrix} -0.220419 \\ 0.43752 \end{bmatrix}$

(b) (10%) Compute $\frac{\partial e}{\partial W^{(1)}}, \frac{\partial e}{\partial W^{(2)}}, \frac{\partial e}{\partial W^{(3)}}$.

$\frac{\partial e}{\partial W^{(1)}} = x^{(0)} \times (\delta^{(1)})^T$ $= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \times \begin{bmatrix} -0.220419 \\ 0.43752 \end{bmatrix}^T = \begin{bmatrix} -0.220419 & 0.43752 \\ -0.440838 & 0.87504 \end{bmatrix}$
$\frac{\partial e}{\partial W^{(2)}} = x^{(1)} \times (\delta^{(2)})^T$ $= \begin{bmatrix} 1 \\ 0.604368 \\ 0.761594 \end{bmatrix} \times -0.347259$ $= \begin{bmatrix} -0.347259 \\ -0.209872 \\ -0.264471 \end{bmatrix}$
$\frac{\partial e}{\partial W^{(3)}} = x^{(2)} \times (\delta^{(3)})^T$ $= \begin{bmatrix} 1 \\ -0.901546 \end{bmatrix} \times -0.927432$ $= \begin{bmatrix} -0.927432 \\ 0.836122 \end{bmatrix}$

(c) (20%) Update the weight matrices using this single datapoint with a learning rate $\eta = 0.5$, repeat the forward propagation and compute $s^{(1)}, x^{(1)}, s^{(2)}, x^{(2)}, s^{(3)} and\ x^{(3)}$.

Update the weight matrices

$W^{(1)}(2) = W^{(1)}(1) - \eta \frac{\partial e}{\partial W^{(1)}}$ $= \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix} - 0.5 \begin{bmatrix} -0.220419 & 0.43752 \\ -0.440838 & 0.87504 \end{bmatrix} = \begin{bmatrix} 0.21021 & -0.01876 \\ 0.520419 & -0.0375199 \end{bmatrix}$
$W^{(2)}(2) = W^{(2)}(1) - \eta \frac{\partial e}{\partial W^{(2)}}$ $= \begin{bmatrix} 0.2 \\ 1 \\ -3 \end{bmatrix} - 0.5 \begin{bmatrix} -0.347259 \\ -0.209872 \\ -0.264471 \end{bmatrix} = \begin{bmatrix} 0.37363 \\ 1.10494 \\ -2.86776 \end{bmatrix}$

$W^{(3)}(2) = W^{(3)}(1) - \eta \frac{\partial e}{\partial W^{(3)}}$ $= \begin{bmatrix} 1 \\ 2 \end{bmatrix} - 0.5 \begin{bmatrix} -0.927432 \\ 0.836122 \end{bmatrix} = \begin{bmatrix} 1.46372 \\ 1.58194 \end{bmatrix}$

Forward propagation

$s^{(1)} = W^{(1)}(2)^T \times x^{(0)} = \begin{bmatrix} 0.21021 & -0.01876 \\ 0.520419 & -0.0375199 \end{bmatrix}^T \times \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.25105 \\ -0.0937998 \end{bmatrix}$
$x^{(1)} = \begin{bmatrix} 1 \\ tanh(1.25105) \\ tanh(-0.0937998) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.848577 \\ -0.0935257 \end{bmatrix}$
$s^{(2)} = W^{(2)}(2)^T \times x^{(1)} = \begin{bmatrix} 0.37363 \\ 1.10494 \\ -2.86776 \end{bmatrix}^T \times \begin{bmatrix} 1 \\ 0.848577 \\ -0.0935257 \end{bmatrix} = [1.57946]$
$x^{(2)} = \begin{bmatrix} 1 \\ tanh(1.57946) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.918518 \end{bmatrix}$
$s^{(3)} = W^{(3)}(2)^T \times x^{(2)} = \begin{bmatrix} 1.46372 \\ 1.58194 \end{bmatrix}^T \times \begin{bmatrix} 1 \\ 0.918518 \end{bmatrix} = [2.91676]$
$x^{(3)} = tanh(2.91676) = [0.994162]$