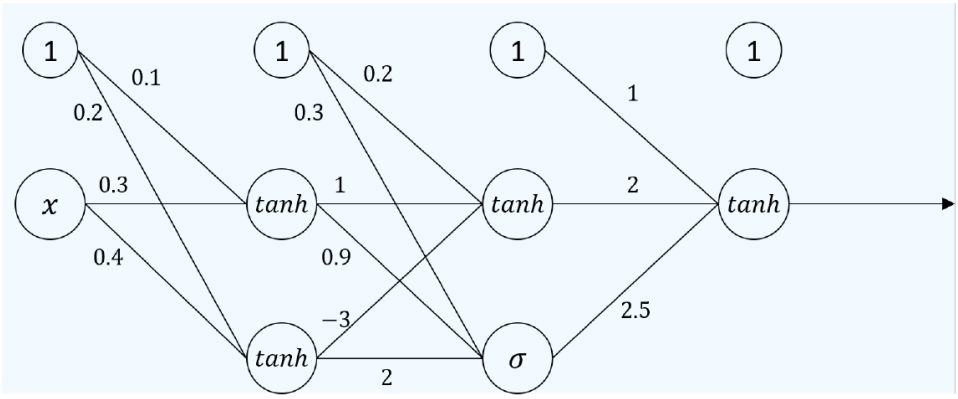

$$\sigma = sigmoid(x) = \frac{1}{1+e^{-x}}$$

(a) (10%) Derive and compute $s^{(1)}, x^{(1)}, s^{(2)}, x^{(2)}, s^{(3)} and\ x^{(3)}$.

$$W^{(1)} = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix} 、 W^{(2)} = \begin{bmatrix} 0.2 & 0.3 \\ 1 & 0.9 \\ -3 & 2 \end{bmatrix} 、 W^{(3)} = \begin{bmatrix} 1 \\ 2 \\ 2.5 \end{bmatrix}$$

Forward propagation

$s^{(1)} = {W^{(1)}}^T \times x^{(0)} = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \end{bmatrix}^T \times \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 1 \end{bmatrix}$
$x^{(1)} = \begin{bmatrix} 1 \\ tanh(0.7) \\ tanh(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0.604368 \\ 0.761594 \end{bmatrix}$
$s^{(2)} = {W^{(2)}}^T \times x^{(1)} = \begin{bmatrix} 0.2 & 0.3 \\ 1 & 0.9 \\ -3 & 2 \end{bmatrix}^T \times \begin{bmatrix} 1 \\ 0.604368 \\ 0.761594 \end{bmatrix} = \begin{bmatrix} -1.48041 \\ 2.36712 \end{bmatrix}$
$x^{(2)} = \begin{bmatrix} 1 \\ tanh(-1.48041) \\ sigmoid(2.36712) \end{bmatrix} = \begin{bmatrix} 1 \\ -0.901546 \\ 0.914285 \end{bmatrix}$
$s^{(3)} = {W^{(3)}}^T \times x^{(2)} = \begin{bmatrix} 1 \\ 2 \\ 2.5 \end{bmatrix}^T \times \begin{bmatrix} 1 \\ -0.901546 \\ 0.914285 \end{bmatrix} = [1.482622]$
$x^{(3)} = tanh(1.482622) = [0.901958]$

(b) (10%)Using the sum square as our error function, derive and compute $\delta^{(3)}, \delta^{(2)}, \delta^{(1)}$.

Error function:

Sum square

$$\delta^{(L)} = 2(x^{(L)} - y)\theta'(S^{(L)})$$

$$\delta^{(3)} = 2\left(x^{(3)} - y\right) \times \left[1 - \left(x^{(3)} \otimes x^{(3)}\right)\right]_1^{d(3)}$$
$$= 2(0.901958 - 1) \times [1 - (0.901958)^2]$$
$$= -0.036564$$

$$\delta^{(2)} = \left[1 - \left(x^{(2)} \otimes x^{(2)}\right)\right] \otimes \left[W^{(3)} \times \delta^{(3)}\right]_1^{d(2)}$$
$$= \begin{bmatrix} [1 - (-0.901546)^2][2 \times (-0.036564)] \\ [(0.914285) \times [1 - (0.914285)]][2.5 \times (-0.036564)] \end{bmatrix}$$
$$= \begin{bmatrix} -0.013691 \\ -0.007164 \end{bmatrix}$$

$$\delta^{(1)} = \left[1 - \left(x^{(1)} \otimes x^{(1)}\right)\right] \otimes \left[W^{(2)} \times \delta^{(2)}\right]_1^{d(1)}$$
$$= \begin{bmatrix} [1 - (0.604368)^2][(-0.0136907) + 0.9 \times (-0.007164)] \\ [1 - (0.761594)^2][-3 \times (-0.0136907) + 2 \times (-0.007164)] \end{bmatrix}$$
$$= \begin{bmatrix} -0.012782 \\ 0.011232 \end{bmatrix}$$

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

$$\frac{\partial e}{\partial W^{(1)}} = x^{(0)} \times (\delta^{(1)})^T$$
$$= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \times \begin{bmatrix} -0.012782 \\ 0.011232 \end{bmatrix}^T = \begin{bmatrix} -0.012782 & 0.011232 \\ -0.025565 & 0.022464 \end{bmatrix}$$

$$\frac{\partial e}{\partial W^{(2)}} = x^{(1)} \times (\delta^{(2)})^T$$
$$= \begin{bmatrix} 1 \\ 0.604368 \\ 0.761594 \end{bmatrix} \times \begin{bmatrix} -0.013691 \\ -0.007164 \end{bmatrix}^T$$
$$= \begin{bmatrix} -0.013690 & -0.007163 \\ -0.008274 & -0.004329 \\ -0.010427 & -0.005456 \end{bmatrix}$$

$$\frac{\partial e}{\partial W^{(3)}} = x^{(2)} \times (\delta^{(3)})^T$$
$$= \begin{bmatrix} 1 \\ -0.901546 \\ 0.914285 \end{bmatrix} \times -0.036564$$
$$= \begin{bmatrix} -0.036564 \\ 0.032964 \\ -0.033430 \end{bmatrix}$$