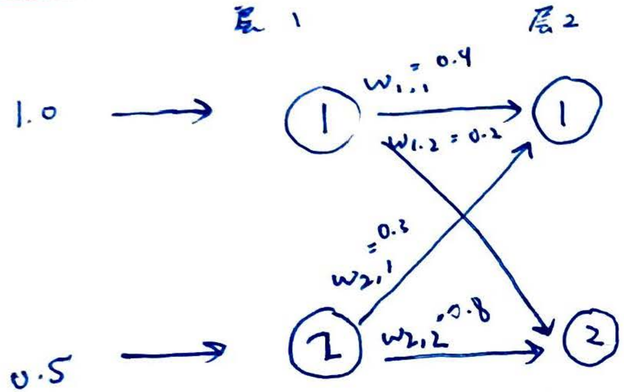


$(0.9 \times 1.0) + (0.3 \times 0.5)$
$= 0.9 + 0.15$
$= 1.05$

$S(x) = \dfrac{1}{1+e^{-1.05}}$

0.6

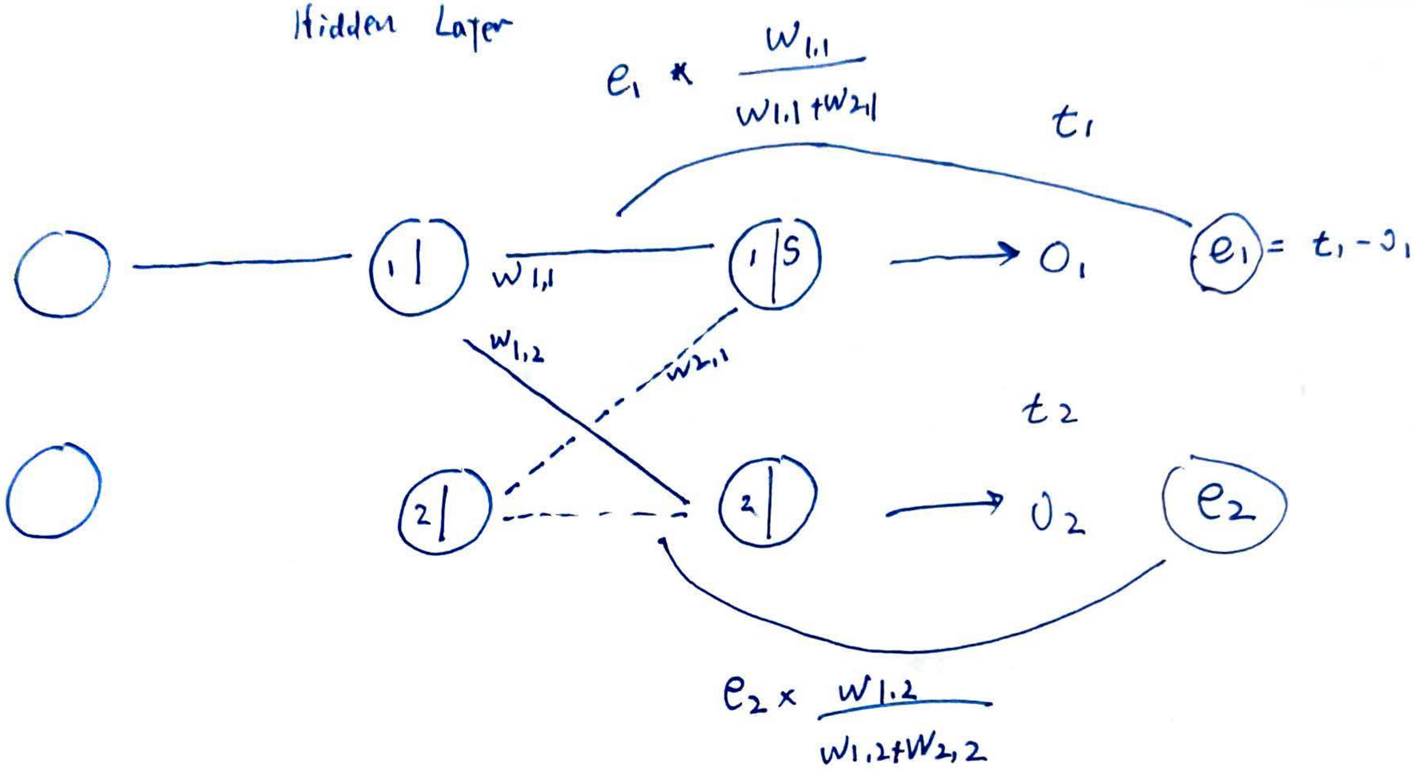
$S(x) = \dfrac{1}{1+e^{-0.6}}$

点乘/内积 dot product / inner product

矩阵乘法

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} (1\times5)+(2\times7) & (1\times6)+(2\times8) \\ (3\times5)+(4\times7) & (3\times6)+(4\times8) \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$$

$$\begin{pmatrix} W1,1 & W2,1 \\ W1,2 & W2,2 \end{pmatrix}\begin{pmatrix} input\,1 \\ input\,2 \end{pmatrix} = \begin{pmatrix} (W1,1 \times in1) + (W2,1 \times in2) \\ (W1,2 \times in1) + (W2,2 \times in2) \end{pmatrix}$$



过程矢量化 —反向传播误差

vectorise the process

$$e_{hidden} = \begin{pmatrix} \frac{w_{1,1}}{w_{1,1}+w_{2,1}}, & \frac{w_{1,2}}{w_{1,2}+w_{2,2}} \\ \frac{w_{2,1}}{w_{2,1}+w_{1,1}}, & \frac{w_{2,2}}{w_{1,2}+w_{2,2}} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} e_1 * \frac{w_{1,1}}{w_{1,1}+w_{2,1}} + e_2 * \frac{w_{1,2}}{w_{1,2}+w_{2,2}} \\ \frac{e_1 * w_{2,1}}{w_{2,1}+w_{1,1}}, \frac{e_2 * w_{2,2}}{w_{1,2}+w_{2,2}} \end{pmatrix}$$

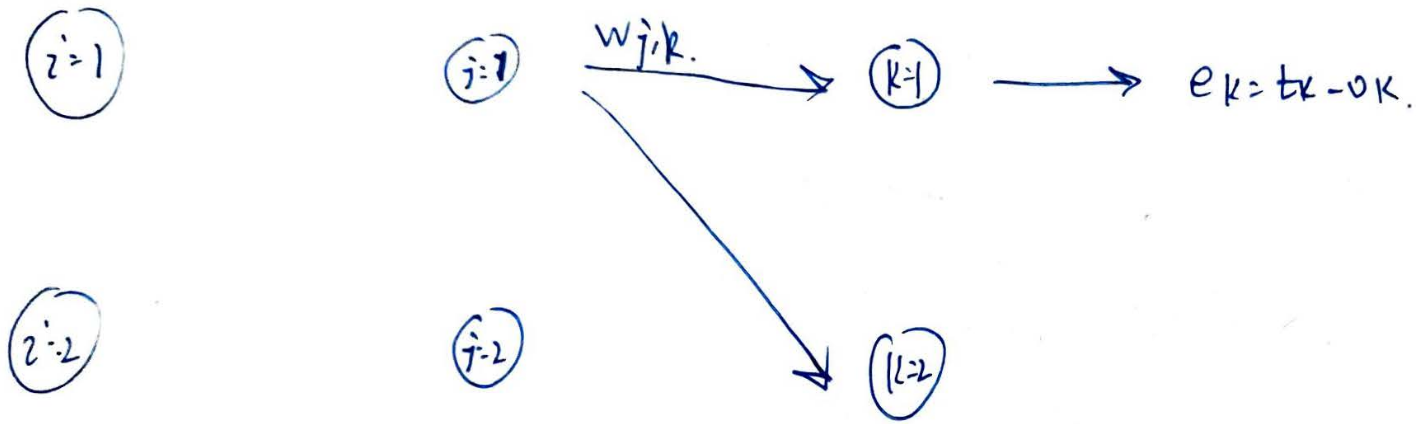
忽略归一化分母

$$e_{hidden} = \begin{pmatrix} w_{1.1}, & w_{1.2} \\ w_{2.1}, & w_{2.2} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} e_1 \cdot w_{1.1} + e_2 w_{1.2} \\ e_1 \cdot w_{2.1}, e_2 \cdot w_{2.2} \end{pmatrix}.$$

$$= W_{h-o}^T \cdot e_o$$

梯度下降 更新权重

Gradient Descent



$$\frac{\partial E}{\partial w_{j,k}} = \frac{\partial}{\partial w_{j,k}} \sum_n (t_n - o_n)^2$$

节点输出只取决于所连接链接

$$= \frac{\partial}{\partial w_{i,k}} (t_k - o_k)^2$$

链式法则

$$\frac{\partial E}{\partial w_{j,k}} = \frac{\partial E}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_{j,k}}$$

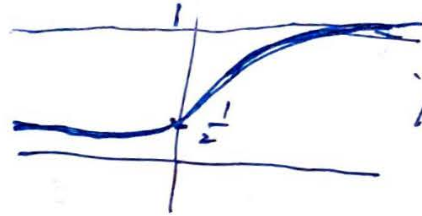
$$\frac{\partial E}{\partial w_{j.k}} = \frac{\partial E}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_{j.k}}$$

$$= \frac{(t_k - o_k)^2}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_{j.k}}$$

$$= \frac{t_k^2 - 2t_k o_k + o_k^2}{\partial o_k} \cdot \frac{\partial o_k}{\partial w_{j.k}}$$

$$= -2(t_k - o_k) \cdot \frac{\partial o_k}{\partial w_{j.k}}$$

Sigmoid function



$$y = \frac{1}{1+e^{-x}}$$

到 Ok 权重.

隐藏层输出

$$= -2(t_k - o_k) \cdot \frac{\partial}{\partial w_{j.k}} \text{sigmoid}\left(\textstyle\sum_j w_{j.k} \cdot o_j\right)$$

$$\frac{\partial}{\partial x} \frac{1}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right)$$

可去

$$= -2(t_k - o_k) \cdot S\left(\textstyle\sum_j w_{j.k} \cdot o_j\right)\left(1 - S\left(\textstyle\sum_j w_{j.k} \cdot o_j\right)\right) \cdot \frac{\partial}{\partial w_{j.k}}\left(\textstyle\sum_j w_{j.k} \cdot o_j\right)$$

$$= -(t_k - o_k) \cdot S\left(\textstyle\sum_j w_{j.k} \cdot o_j\right)\left(1 - S\left(\textstyle\sum_j w_{j.k} \cdot o_j\right)\right) \cdot o_j$$

下去

$$= -(e_j) \cdot S\left(\textstyle\sum_i w_{i.j} \cdot o_i\right)\left(1 - S\left(\textstyle\sum_i w_{i.j} \cdot o_i\right)\right) \cdot o_i$$

权重改变方向与梯度方向相反.

$$\text{new } w_{j.k} = \text{old } w_{j.k} - \alpha \cdot \frac{\partial E}{\partial w_{j.k}}$$

学习率.

$$\begin{pmatrix} \Delta W_{1,1}, & \Delta W_{1,2} \\ \Delta W_{2,1}, & \Delta W_{2,2} \end{pmatrix} = \begin{pmatrix} E_1 * S_1(1-S_1) \\ E_2 * S_2(1-S_2) \end{pmatrix} \cdot \begin{pmatrix} O_1 & O_2 \end{pmatrix}$$

变化矩阵.

$$\Delta W_{j,k} = \alpha \cdot \left(E_k \cdot O_k (1 - O_k) \right) \cdot O_j^T.$$

$$\frac{\partial E}{\partial w_{j,k}} = -(t_k - O_k) \cdot \text{sigmoid}\left(\Sigma_j W_{j,k} \cdot O_j\right)\left(1 - \text{sigmoid}\left(\Sigma_j w_{j,k} \cdot O_j\right)\right) \cdot O_j.$$

$$\text{new } W_{j,k} = \text{old } W_{j,k} - \alpha \cdot \frac{\partial E}{\partial w_{j,k}}.$$