Blackboard 2.2a : BackProp A

output

$$\hat{y}_i = g^{(3)}\left[\underline{a_i^{(3)}}\right] \quad ; \quad \boxed{a_i^{(n)} = \sum_j w_{ij}^{(n)} x_j^{(n-1)}}$$

$$= g^{(3)}\left[\sum_j w_{ij}^{(3)} \cdot x_j^{(2)}\right]$$

$$= g^{(3)}\left[\sum_j w_{ij}^{(3)} \, g^{(2)}\left(\underline{a_j^{(2)}}\right)\right]$$

$$= g^{(3)}\left[\sum_j w_{ij}^{(3)} \, g^{(2)}\left[\sum_k w_{jk}^{(2)} x_k^{(1)}\right]\right]$$

$$= g^{(3)}\left[\sum_j w_{ij}^{(3)} \, g^{(2)}\left[\sum_k w_{jk}^{(2)} \, g^{(1)}\left(\underline{a_k^{(1)}}\right)\right]\right]$$

$$\uparrow$$

$$\sum_l w_{kl}^{(1)} x_l^{(0)}$$

$$\uparrow$$

input

gradient $\frac{\partial E}{\partial w_{23}^{(1)}}$ of $E = \frac{1}{2} \sum_i \sum_\mu \left[t_i^\mu - \hat{y}_i^\mu \right]^2$

$= \frac{1}{2} \sum_i \sum_\mu \tilde{E}(\mu, i)$

Step 1: identify intermediate variables

- $a_i^{(n)}$ = activation/drive of a neuron
- $x_i^{(n)}$ = neuron output
- $\delta_k^{(n)} = \frac{\partial E}{\partial a_k^{(n)}}$ definition!

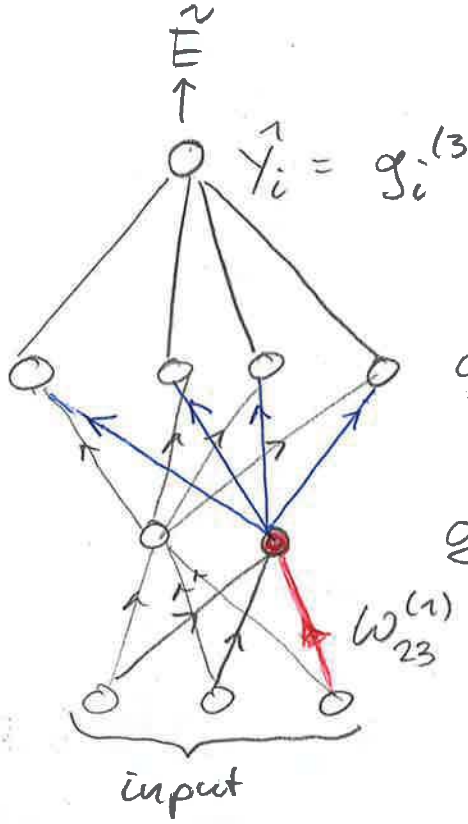
Step 2: write weight update with these variables

- $\Delta w_{23}^{(1)} = -\eta \cdot \frac{\partial E}{\partial w_{23}^{(1)}} = -\eta \frac{\partial E}{\partial a_2^{(1)}} \cdot \frac{\partial a_2^{(1)}}{\partial w_{23}^{(1)}}$

$\overset{(*)}{=} -\eta \cdot \delta_2^{(1)} \cdot x_3^{(0)}$

- analogous for all weights / layers

Step 3: analyze dependency graph / chain rule



$\hat{Y}_i = g_i^{(3)}[a_i^{(3)}]$

$g^{(2)}(a_j^{(2)})$

$g^{(1)}(a_2^{(1)})$

how much does $\hat{E}$ change, if I change $a_2^{(1)}$?

$$\frac{\partial E}{\partial a_2^{(1)}} = \sum_j \frac{\partial E}{\partial a_j^{(2)}} \cdot \frac{\partial a_j^{(2)}}{\partial a_2^{(1)}}$$

chain rule

all units of previous layer

$$\delta_2^{(1)} = \sum_j \delta_j^{(2)} \cdot w_{i2} \cdot g'^{(n-1)} \ldots\ldots$$

from (*): $a_i^{(n)} = \sum_j w_{ij} \, x_j^{(n-1)}$

$$a_i^{(n)} = \sum_j w_{ij} \, g^{(n-1)}(a_j^{(n-1)})$$

$$\boxed{\frac{\partial a_i^{(n)}}{\partial a_2^{(n-1)}} = w_{i2} \cdot g'^{(n-1)}}$$