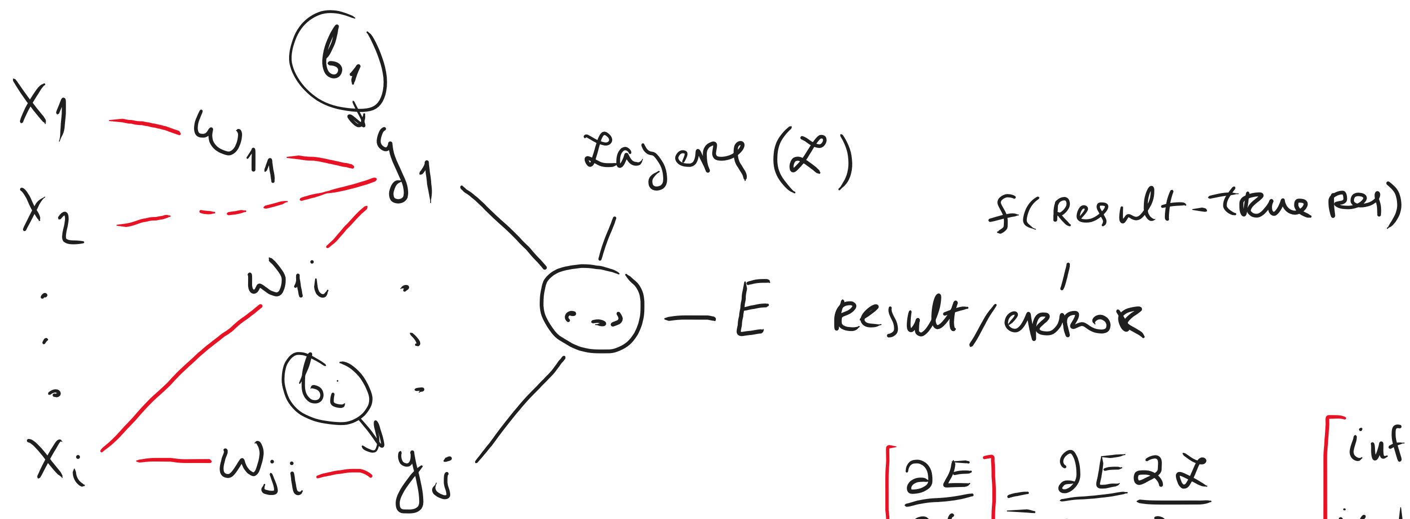

$$\left[\frac{\partial E}{\partial y_j}\right] = \frac{\partial E}{\partial \mathcal{L}}\frac{\partial \mathcal{L}}{\partial y_j}$$

[info of other layers i.e $\{z_1, \ldots, z_j\}$ is hidden in $\frac{\partial E}{\partial \mathcal{L}} = f\left[\frac{\partial E}{\partial z_i}, \frac{\partial z_j}{\partial y_i}, \frac{\partial z_j}{\partial w^{(z)}_{jk}}\right]$]

$\{a_i\}_1^3 \overset{DEF}{=} \{a_1, a_2, a_3\}$; $f(g(x)) \overset{SAME}{=} f[g(x)]$

$$E = E[\mathcal{L}(y_1 \cdots y_j)] = E[\mathcal{L}(\{y_k\}_{k=1}^{k=j})]$$

$$y_j = y_j\left[\{w_{ji}\}_{i=1}^{i}, \{x_i\}_{i=1}^{i}, \{b_i\}\right]$$

$$\left[\frac{\partial E}{\partial x_i}\right] = \sum_j \frac{\partial E}{\partial \mathcal{L}}\frac{\partial \mathcal{L}}{\partial y_j} \cdot \frac{\partial y_j}{\partial x_i} = \sum_j \left[\frac{\partial E}{\partial y_j}\right] \cdot \frac{\partial y_j}{\partial x_i}$$

$$\frac{\partial E}{\partial w_{ji}} = \left[\frac{\partial E}{\partial y_j}\right] \cdot \frac{\partial y_j}{\partial w_{ji}} \qquad \frac{\partial E}{\partial b_j} = \frac{\partial E}{\partial y_j}\underbrace{\frac{\partial y_j}{\partial b_j}}_{1} = \frac{\partial E}{\partial y_j}$$

$$\vec{y} = [y_1 \cdots y_j]^T$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} w_{11} & \cdots & w_{1i} \\ \vdots & \ddots & \\ w_{j1} & & w_{ji} \end{bmatrix}}_{j \times i} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix}_{i \times 1}$$

Lets stick to column vectors and $j \times i$ weights matrix

$$\frac{\partial E}{\partial \vec{x}} = \left[\frac{\partial E}{\partial x_1} \cdots \frac{\partial E}{\partial x_i}\right] \quad \text{formal definition}$$

$$y_{j1} = \sum_i w_{ji} x_{i1} + b_{i1} \qquad \frac{\partial y_j}{\partial x_i} = w_{ji} = (W)_{j,i}$$

$$\left(\frac{\partial E}{\partial \vec{x}}\right)_{1,i} = \sum_j \left(\frac{\partial E}{\partial \vec{y}}\right)_{1,j}\left(\frac{\partial \vec{y}}{\partial \vec{x}}\right)_{j,i} = \sum_j \left(\frac{\partial E}{\partial \vec{y}}\right)_{1,j}(W)_{j,i} \Leftrightarrow \frac{\partial E}{\partial \vec{x}}_{col} = \text{dot}\left(\frac{\partial E}{\partial \vec{y}}_{row}, W_{col}\right)$$

$$\Rightarrow \frac{\partial E}{\partial \vec{x}} = \frac{\partial E}{\partial \vec{y}} \cdot W$$

formally $\frac{\partial E}{\partial \vec{x}}$ is a row vec. but we keep it as a col. vec.

$$\Rightarrow \left[\begin{bmatrix} \frac{\partial E}{\partial x_1} \\ \frac{\partial E}{\partial x_2} \end{bmatrix} = W^T \begin{bmatrix} \frac{\partial E}{\partial y_1} \\ \frac{\partial E}{\partial y_2} \end{bmatrix}\right]$$

$$\left(\frac{\partial E}{\partial \vec{x}}\right)^T = W^T\left(\frac{\partial E}{\partial \vec{y}}\right)^T$$

$$\left(\frac{\partial E}{\partial w_{ji}}\right) = \left(\frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial w_{ji}}\right) = \left(\frac{\partial E}{\partial y_j} x_i\right) \quad \text{need } j \times i \text{ matrix}$$

$$= \underbrace{\begin{bmatrix} \frac{\partial E}{\partial y_1} \\ \frac{\partial E}{\partial y_j} \end{bmatrix}}_{j \times 1} \otimes \underbrace{[x_1 \cdots x_i]}_{1 \times i} = \begin{bmatrix} x_1\frac{\partial E}{\partial y_1} & \cdots & \\ & \ddots & \\ & & x_i\frac{\partial E}{\partial y_j} \end{bmatrix} = \left(\frac{\partial E}{\partial \vec{y}}\right)^T \otimes x^T$$

$$\left[\left(\frac{\partial E}{\partial w_{ji}}\right) = \left(\frac{\partial E}{\partial \vec{y}}\right)^T \otimes x^T\right]$$

update weights $W = W - \underset{\text{learning rate}}{L.R.}\left(\frac{\partial E}{\partial \vec{y}}\right)^T \otimes x^T$

$$\left[\left(\frac{\partial E}{\partial b_j}\right)^T = \left(\frac{\partial E}{\partial y_j}\right)^T\right]$$

update biases the same way

$$\left[\left(\frac{\partial E}{\partial \vec{x}}\right)^T = W^T\left(\frac{\partial E}{\partial \vec{y}}\right)^T\right]$$

pass this as $\frac{\partial E}{\partial \mathcal{L}}$ further