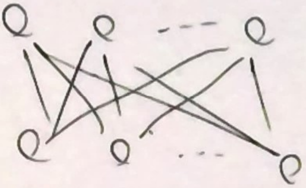
Forward Pass for 1 layer:

layer [l]

layer [l-1]



linear fn

$$x_j^{[l]} = \sum_{i=1}^{n^{[l-1]}} w_{ji} y_i^{[l-1]} + b_j^{[l]}$$

activation

$$y_j^{[l]} = \frac{1}{1 + e^{-x_j^{[l]}}}$$

total Error

network output

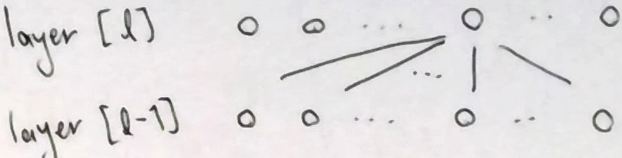
$$E = \frac{1}{2} \sum_{c} \sum_{j} \left(y_{j,c} - d_{j,c} \right)^2 .$$

training examples

output units

target

Backprop for 1 layer:



Suppose we have $\dfrac{\partial E}{\partial y_j^{[l]}}$, for $j = 1, \dots, n^{[l]}$.

we wish to find $\dfrac{\partial E}{\partial w_{ji}^{[l]}}$, $\dfrac{\partial E}{\partial b_j^{[l]}}$, and $\dfrac{\partial E}{\partial y_i^{[l-1]}}$.

I) $$\frac{\partial E}{\partial w_{ji}^{[l]}} = \frac{\partial E}{\partial y_j^{[l]}} \cdot \frac{\partial y_j^{[l]}}{\partial x_j^{[l]}} \cdot \frac{\partial x_j^{[l]}}{\partial w_{ji}^{[l]}};$$

the first term is given.

The second term:

sigmoid function

$$\frac{\partial y_j^{[l]}}{\partial x_j^{[l]}} = \frac{\partial}{\partial x_j^{[l]}} \left[\frac{1}{1 + e^{-x_j^{[l]}}} \right]$$

$$= y_j^{[l]} \left(1 - y_j^{[l]} \right)$$

the third term:

$$\frac{\partial x_j^{[l]}}{\partial w_{ji}^{[l]}} = \frac{\partial}{\partial w_{ji}^{[l]}} \left[\sum_{i=1}^{n^{[l-1]}} w_{ji}^{[l]} y_i^{[l-1]} + b_j^{[l]} \right]$$

$$= y_i^{[l-1]}.$$

Putting all three terms together,

$$\boxed{\begin{gathered} \frac{\partial E}{\partial w_{ji}^{[l]}} = \frac{\partial E}{\partial y_j^{[l]}} \cdot y_j^{[l]} \left(1 - y_j^{[l]}\right) \cdot y_i^{[l-1]}; \\ \text{for } j = 1, \dots, n^{[l]} \\ \text{for } i = 1, \dots, n^{[l-1]} \end{gathered}}$$

II) $$\frac{\partial E}{\partial b_j^{[l]}} = \frac{\partial E}{\partial y_j^{[l]}} \cdot \frac{\partial y_j^{[l]}}{\partial x_j^{[l]}} \cdot \frac{\partial x_j^{[l]}}{\partial b_j^{[l]}}$$

First two terms are calculated for case (I)

third term :

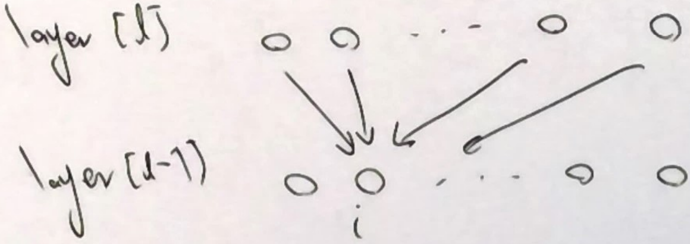
$$\frac{\partial x_j^{[l]}}{\partial b_j^{[l]}} = \frac{\partial}{\partial b_j^{[l]}} \left(\sum_{i=1}^{n^{[l-1]}} w_{ji} \, y_i^{[l-1]} + b_j^{[l]} \right)$$

$$= 1 .$$

Putting all three together,

$$\boxed{\frac{\partial E}{\partial b_j^{[l]}} = \frac{\partial E}{\partial y_j^{[l]}} \cdot y_j^{[l]} \cdot (1-y_j^{[l]}) \cdot 1, \quad \text{for } j = 1, \ldots, n^{[l]}}$$

III) $$\frac{\partial E}{\partial y_i^{[l-1]}} = \sum_{j=1}^{n^{[l]}} \frac{\partial E}{\partial x_j^{[l]}} \frac{\partial x_j^{[l]}}{\partial y_i^{[l-1]}}$$



$\frac{\partial E}{\partial y_i^{[l-1]}}$ is a linear combination of all $\frac{\partial E}{\partial x_j^{[l]}}$ and contribution of a particular $x_j^{[l]}$ is $\frac{\partial x_j^{[l]}}{\partial y_i^{[l-1]}}$

$$\frac{\partial E}{\partial x_j^{[l]}} = \frac{\partial E}{\partial y_j^{[l]}} \cdot \frac{\partial y_j^{[l]}}{\partial x_j^{[l]}} = \frac{\partial E}{\partial y_j^{[l]}} \cdot y_j^{[l]} (1-y_j^{[l]})$$

and

$$\frac{\partial x_j^{[l]}}{\partial y_i^{[l-1]}} = \frac{\partial}{\partial y_i^{[l-1]}} \left[\sum_{i=1}^{n^{[l-1]}} w_{ji}^{[l]} y_i^{[l-1]} + b_j^{[l]} \right]$$

$$= w_{ji}^{[l]}$$

Putting all terms together,

$$\boxed{\begin{aligned} \frac{\partial E}{\partial y_i^{[l-1]}} &= \sum_{j=1}^{n^{[l]}} \frac{\partial E}{\partial y_j^{[l]}} \cdot y_j^{[l]} (1 - y_j^{[l]}) \cdot w_{ji}^{[l]} ; \\ &\text{for } i = 1, \cdots, n^{[l-1]} . \end{aligned}}$$

If training over C items:

$$\frac{\partial E}{\partial w_{ji}^{[l]}} = \sum_{c} \frac{\partial E}{\partial w_{ji}^{[l](c)}}$$

$$\frac{\partial E}{\partial b_{j}^{[l]}} = \sum_{c} \frac{\partial E}{\partial b_{j}^{[l](c)}}$$

$$\left(\frac{\partial E}{\partial y_{i}^{[l-1]}} \right)^{(c)} = \frac{\partial E}{\partial y_{i}^{[l-1](c)}}$$