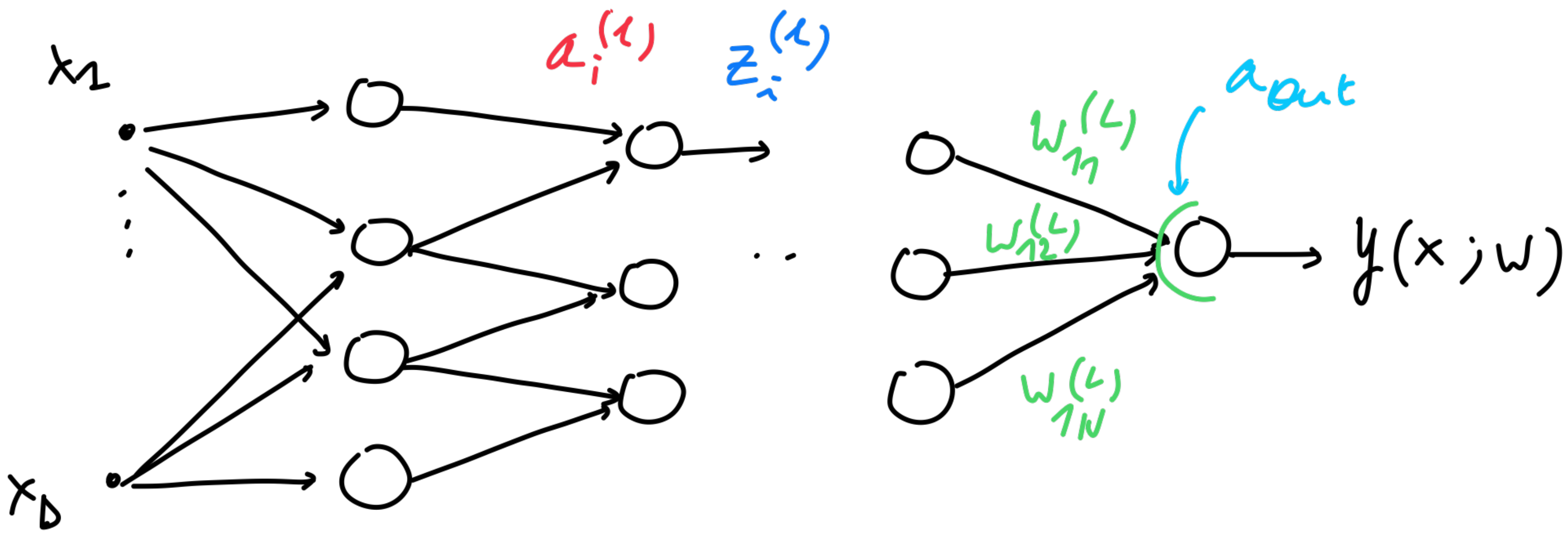


$a_i^{(l)}$ = preactivation of unit i in layer l

$$a_i^{(l)} = \sum_{j=1}^{N_{L-1}} w_{ij}^{(l)} z_j^{(l-1)} + w_{i0}^{(l)} \longleftarrow \frac{\partial a_i^{(l)}}{\partial w_{ij}^{(l)}} = z_j^{(l-1)}$$

$z_i^{(l)}$ = postactivation for unit i in layer l

$$z_i^{(l)} = \sigma(a_i^{(l)}) = \sigma\left(\sum_{j=1}^{N_{L-1}} w_{ij}^{(l)} z_j^{(l-1)} + w_{i0}^{(l)} \right)$$

$$p(t=0|x) = y(x;w)$$

$$p(t=1|x) = 1 - y(x;w)$$

$$p(\{t(x^{(i)}) = t^{(i)}\}_{i=1}^{N} | x) = \prod_{i=1}^{N} p(t(x^{(i)}) = t^{(i)} | x)$$

$$= \prod_{i=1}^{N} y(x^{(i)};w)^{1-t^{(i)}} (1 - y(x^{(i)};w))^{t^{(i)}}$$

→ taking the log

$$\boxed{L(w) = -(1-t^{(i)}) \log y(x^{(i)};w) - t^{(i)} \log (1 - y(x^{(i)};w))}$$

1) We want $$\frac{\partial L}{\partial w_{ij}^{(l)}} = \underbrace{\frac{\partial L}{\partial a_i^{(l)}}}_{\delta_i^{(l)}} \cdot \underbrace{\frac{\partial a_i^{(l)}}{\partial w_{ij}^{(l)}}}_{z_j^{(l-1)}}$$

$$L \leftarrow a_i^{(l)}, \dots \leftarrow w_{ij}^{(l)}$$

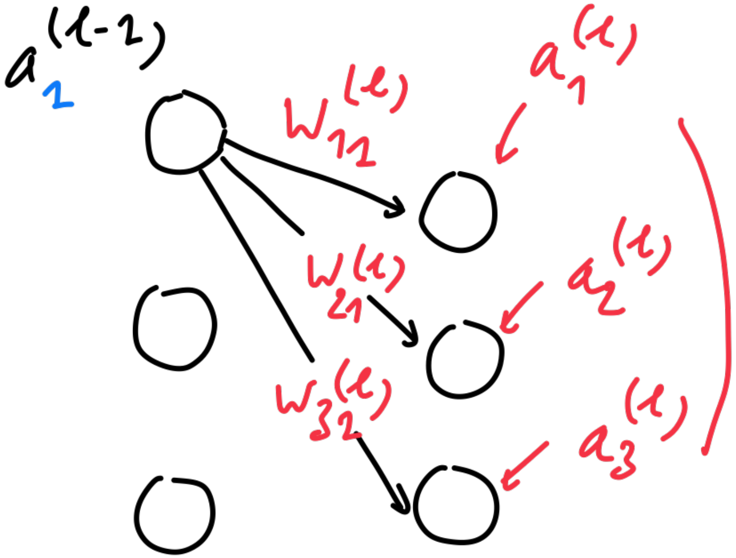
2) $$\delta_{out} = \frac{\partial L}{\partial a_{out}} \rightarrow \frac{\partial\left(-(1-t^{(i)}) \log \sigma(a_{out}) - t^{(i)} \log(1-\sigma(a_{out}))\right)}{\partial a_{out}}$$

$$= -(1-t^{(i)}) \frac{\sigma'}{\sigma(a_{out})} + t^{(i)} \frac{\sigma'}{1-\sigma(a_{out})}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$= -(1-t^{(i)})(1-\sigma(a_{out})) + t^{(i)} \sigma(a_{out})$$

$$\boxed{\delta_{out} = \sigma(a_{out}) - (1-t^{(i)})}$$



$$L\left(a_1^{(\ell)} \quad a_2^{(\ell)} \quad a_3^{(\ell)}\right)$$

$$\frac{\partial L}{\partial a_1^{(\ell-1)}} \leftarrow L\left(a_1^{(\ell)}(a_1^{(\ell-1)}) \quad a_2^{(\ell)}(a_1^{(\ell-1)})\right)$$

$$\frac{\partial L}{\partial a_i^{(\ell-1)}} = \sum_{j=1}^{N_L} \frac{\partial L}{\partial a_j^{(\ell)}} \cdot \frac{\partial a_j^{(\ell)}}{\partial a_i^{(\ell-1)}}$$

$$\delta_i^{(\ell-1)} = \sum_{j=1}^{N_L} \delta_j^{(\ell)}$$

$$\frac{\partial a_j^{(\ell)}}{\partial a_i^{(\ell-1)}} \rightarrow a_j^{(\ell)} = \sum_{i=1}^{N_{L-1}} w_{ji}^{(\ell)} \cdot \sigma(a_i^{(\ell-1)}) + w_{j0}^{(\ell)}$$

$$\frac{\partial a_j^{(\ell)}}{\partial a_i^{(\ell-1)}} = w_{ji}^{(\ell)} \, \sigma'(a_i^{(\ell-1)})$$

Backpropagation

1) → Forward propagat $x^{(i)}$ through the network, and get all $a_i^{(l)}$, $z_i^{(l)}$

2) you get $\delta_{out} = \frac{\partial L}{\partial a_{out}} = \sigma(a_{out}) - (1 - t^{(i)})$

3) Backpropagat the δ_{out}

$$\delta_i^{(l-1)} = \sum_{j=1}^{N_L} \delta_j^{(l)} W_{ji}^{(l)} \cdot \sigma'(a_i^{(l-1)})$$

4) Get the gradient as

$$\frac{\partial L}{\partial W_{ij}^{(l)}} = \delta_i^{(l)} \cdot z_j^{(l-1)}$$

$$\frac{\partial L}{\partial W_{i0}^{(l)}} = \delta_i^{(l)}$$