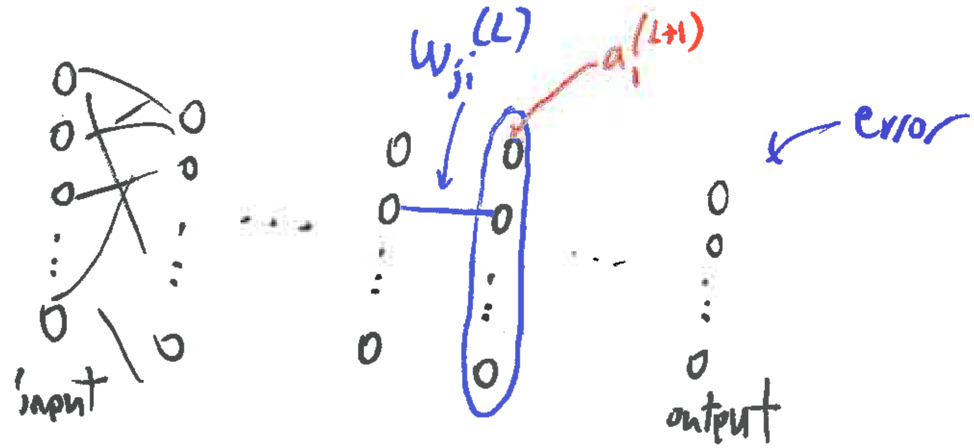


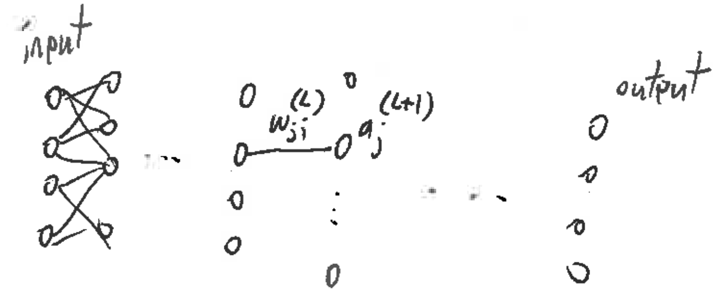
$$E_n(w) = E_n\left(a_1^{(L+1)}, a_2^{(L+1)}, \ldots, a_{D_{L+1}}^{(L+1)}\right)$$

$$\frac{\partial E_n(w)}{\partial w_{ji}^{(L)}} = \frac{\partial E_n}{\partial a_1^{(L+1)}} \frac{\partial a_1^{(L+1)}}{\partial w_{ji}^{(L)}} + \frac{\partial E_n}{\partial a_2^{(L+1)}} \frac{\partial a_2^{(L+1)}}{\partial w_{ji}^{(L)}} + \cdots + \frac{\partial E_n}{\partial a_{D_{L+1}}} \frac{\partial a_{D_{L+1}}}{\partial w_{ji}^{(L)}}$$

$$= \frac{\partial E_n}{\partial a_j^{(L+1)}} \frac{\partial a_j^{(L+1)}}{\partial w_{ji}^{(L)}}$$

$$a_j^{(L+1)} = \sum_{i=1}^{D_L} w_{ji}^{(L)} z_i^{(L)} + w_{j0}^{(L)}$$

$$z_j^{(L)} = h\left(a_j^{(L)}\right)$$

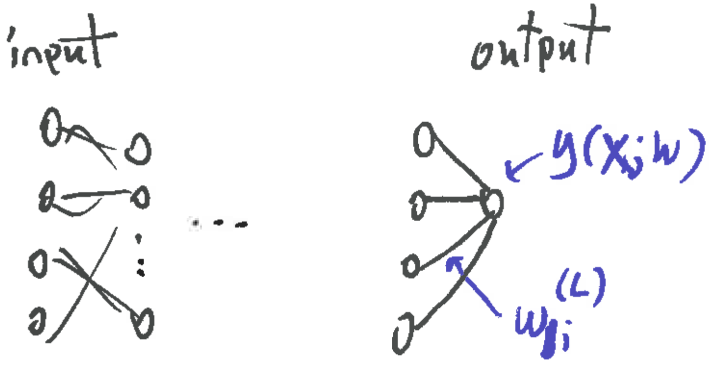


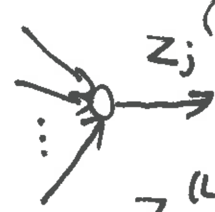
$$\frac{\partial E_n}{\partial w_{ji}^{(L)}} = \frac{\partial E_n}{\partial a_j^{(L+1)}} \frac{\partial a_j^{(L+1)}}{\partial w_{ji}^{(L)}}$$

recursive → define

$$\frac{\partial E_n}{\partial a_j^{(L)}} = \delta_j^{(L)}$$

easy → $$\frac{\partial a_j^{(L+1)}}{\partial w_{ji}^{(L)}} = z_i^{(L)}$$





$z_j^{(L)}$ output of node j in layer L

$$z_j^{(L)} = h(a_j^{(L)})$$

$$a_j^{(L)} = \sum_{i=1}^{D_L} w_{ji}^{(L-1)} z_i^{(L-1)} + w_{j0}^{(L-1)}$$

$$E_n(w) = \frac{1}{2}\left\{ y(x_n;w) - t_n \right\}^2$$

$$= \frac{1}{2}\left\{ h(a_1^{(L+1)}) - t_n \right\}^2$$

$$= \frac{1}{2}\left\{ h\left(\sum_{k=1}^{D_L} w_{1k}^{(L)} z_k^{(L)} + w_{10}^{(L)}\right) - t_n \right\}^2$$

$$\frac{\partial E_n(w)}{\partial w_{1i}^{(L)}} = \left\{ h(a_1^{(L+1)}) - t_n \right\} h'(a_1^{(L+1)}) z_i^{(L)}$$

output layer, e.g. regression $E_n(w) = \frac{1}{2}\{y(x_n; w) - t_n\}^2$

$$\frac{\partial E_n}{\partial w_{ji}^{(L)}} = \{y(x_n; w) - t_n\} \frac{\partial}{\partial w_{ji}^{(L)}} y(x_n; w)$$

$$= \{y(x_n; w) - t_n\} z_i^{(L)}$$

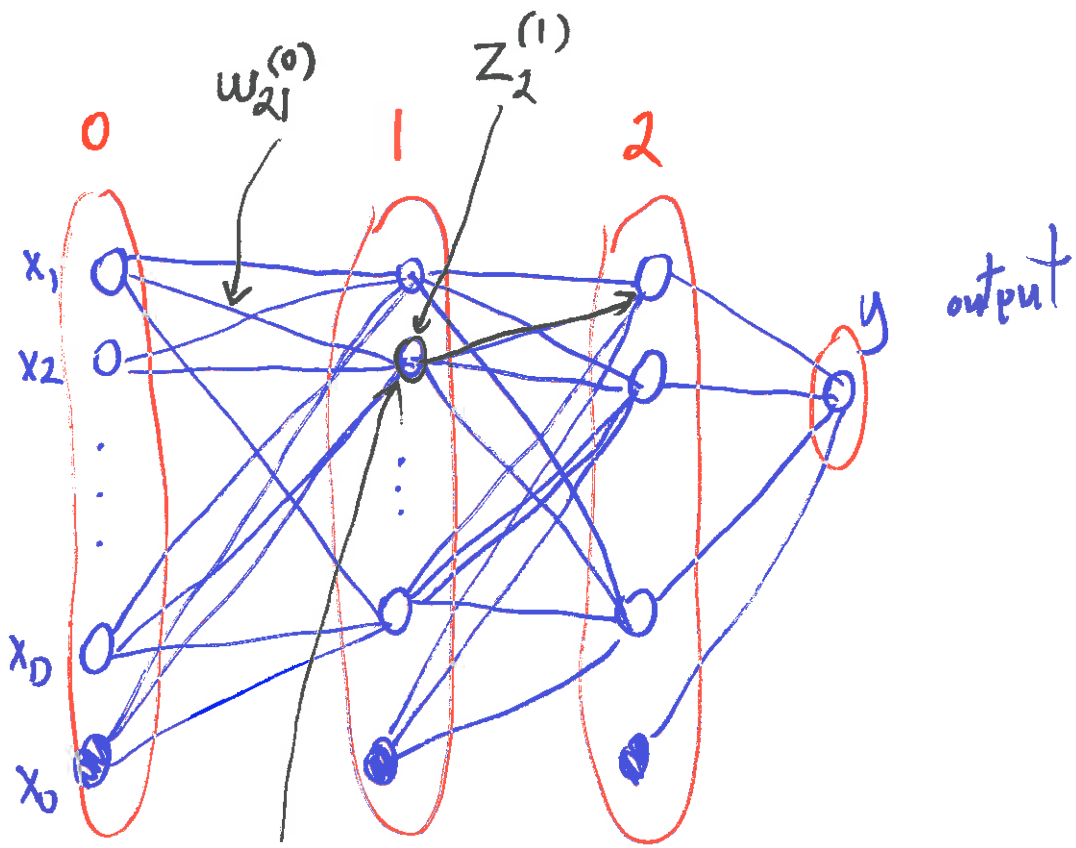
$$E_n(w) = (y(x_n; w) - t_n)^2$$

$$\nabla E_n(w) = ?$$

$$\nabla E_n(w) = \left(\frac{\partial E_n}{\partial w_{10}^{(0)}}, \frac{\partial E_n}{\partial w_{11}^{(0)}}, \ldots, \frac{\partial E_n}{\partial_{KD}^{(L)}} \right)$$

W weights in network 10^8

$$O(W^2) \longrightarrow O(W)$$



$z_j^{(L)}$

(L) → layer

j → node in the layer

output of a node / value of an internal basis function

activation of a node

$$z_i^{(L)} = h\left(a_i^{(L)}\right)$$

$$a_i^{(L)} = \sum_j w_{ij}^{(L-1)} z_j^{(L-1)}$$

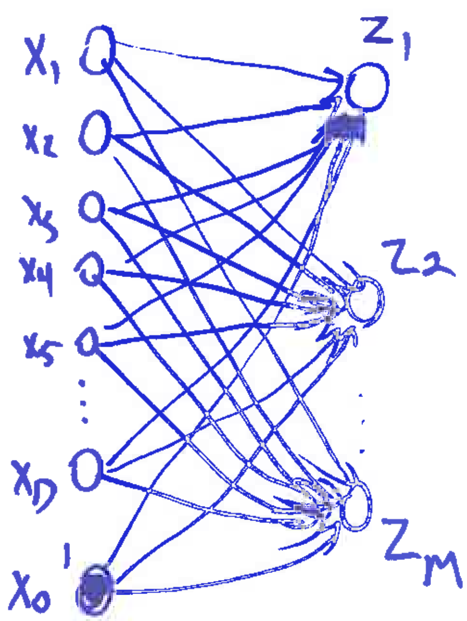
$$f'(x) \approx \frac{f(x+\delta) - f(x-\delta)}{2\delta}$$

$$\frac{\partial E_n(w)}{\partial w_{ji}^{(L)}} \approx \frac{E_n(\vec{w} + (0,0,\dots,0,\delta,0,\dots,0)) - E_n(\vec{w} - (0,\dots,0,\delta,0,\dots,0))}{2\delta}$$

$$x \in \mathbb{R}^D$$

$$y(x;w) \in \mathbb{R}^K$$

$$w^T x$$



$$z_1 = \frac{1}{1+e^{w^T x}}$$