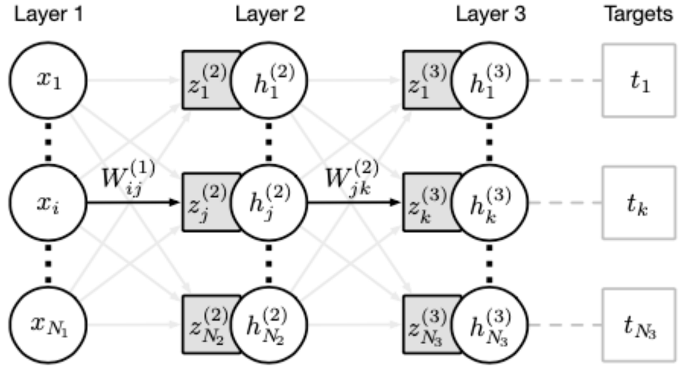
Ex 2: Backprop by Hand



a) $H^{(2)} = \sigma(X \cdot W^{(1)})$

Note that $H^{(2)} \in \mathbb{R}^{D \times N_2}$

b) $H^{(3)} = \sigma\left(H^{(2)} \cdot W^{(2)}\right) \in \mathbb{R}^{D \times N_3}$

c) $$\nabla_{Z^{(3)}} E = \begin{array}{cc} & i \\ d & \begin{bmatrix} \frac{\partial E}{\partial Z_{11}^{(3)}} & \vdots \\ \cdots & \frac{\partial E}{\partial Z_{di}^{(3)}} \end{bmatrix} \end{array}$$

First, $$\nabla_{H^{(3)}} E = \begin{array}{cc} & i \\ d & \begin{bmatrix} \frac{\partial E}{\partial H_{11}^{(3)}} & \downarrow \\ \longrightarrow & \frac{\partial E}{\partial H_{di}^{(3)}} \end{bmatrix} \end{array}$$ where $\frac{\partial E}{\partial H_{di}^{(3)}} = \frac{\partial L}{\partial H_{di}^{(3)}}$

Then $\nabla_{Z^{(3)}} E = \sigma'(Z^{(3)}) \odot \nabla_{H^{(3)}} E$

d) $\nabla_{W^{(2)}} E$

$[\nabla_{W^{(2)}} E]_{jk} = \nabla_{W_{jk}^{(2)}} E$ But $W_{jk}^{(2)}$ appears in all D terms of $E = \frac{1}{D}\sum_{d=1}^{D} L(h_d^{(3)}, t_d)$

$\therefore \nabla_{W_{jk}^{(2)}} E = \frac{1}{D} \sum_{d=1}^{D} \nabla_{W_{jk}^{(2)}} L(h_d^{(3)}, t_d)$

$= \frac{1}{D} \sum^{D} \frac{\partial L}{\ } \quad \frac{\partial Z^{(3)}}{\ }$

$$= \frac{1}{D} \sum_{d=1} \frac{\partial L}{\partial Z^{(3)}_{d,k}} \underbrace{\frac{\partial Z^{(3)}_{d,k}}{\partial W_{jk}}}_{H^{(2)}_{d,j}}$$

$$= \frac{1}{D} \underbrace{\begin{bmatrix} H^{(2)}_{1,j} & \cdots & H^{(2)}_{D,j} \end{bmatrix}}_{\text{This is the transpose of the } j^{th} \text{ col of } H^{(2)}} \begin{bmatrix} \frac{\partial L}{\partial Z^{(3)}_{1,k}} \\ \vdots \\ \frac{\partial L}{\partial Z^{(3)}_{D,k}} \end{bmatrix}$$

This is the k^{th} col of $\nabla_{Z^{(3)}} E$.

$$\boxed{\therefore \nabla_{W^{(2)}} E = \left(H^{(2)}\right)^T \cdot \nabla_{Z^{(3)}} E}$$