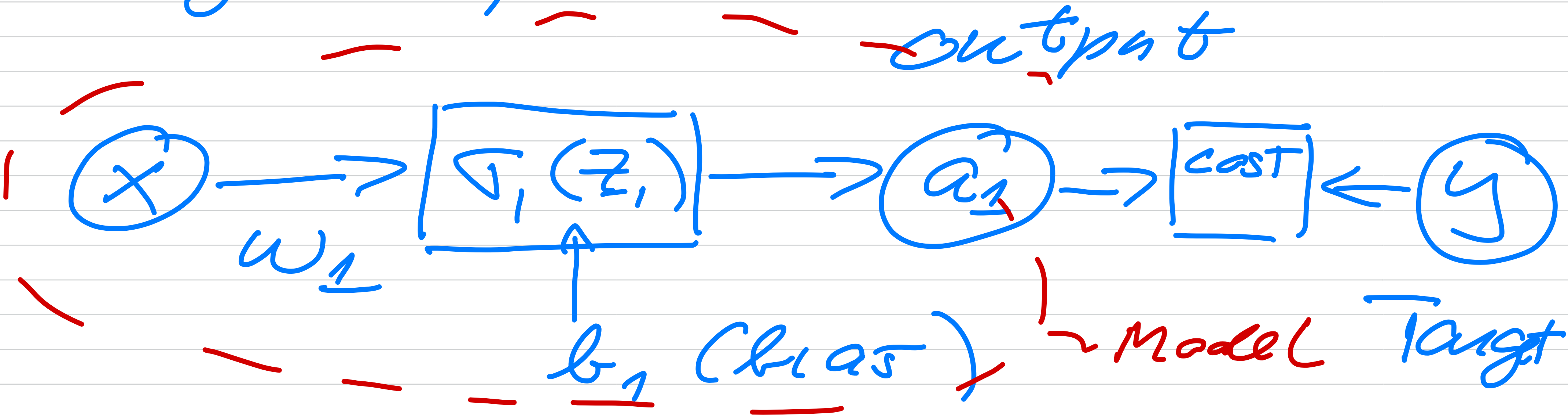
FYS-STK3155/4155
Lecture October 6, 2025

FYS-STK3155/4155 October 6

warm up (all scalars)

one perceptron (no hidden layers)

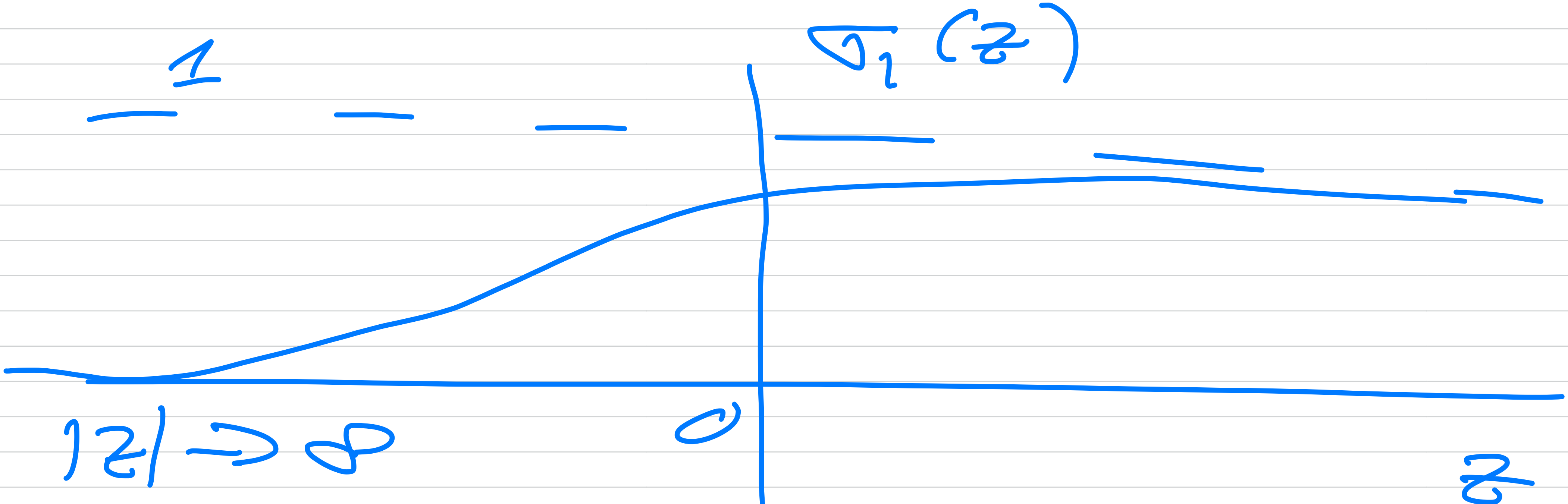


$$C = C(a_1, y, \Theta)$$

$$\Theta = \{w_1, b_1\}$$

$\sigma_1(z_1) \quad \wedge \quad z_1 = x\underline{W_1} + \underline{b_1}$

$$\sigma_1(z) = \frac{1}{1+e^{-z}}$$



$|z| \Rightarrow \infty$

$\sigma_1(z) \Rightarrow \text{const}$

(Preparing for later & Automatic differentiation)

$$\frac{\partial C}{\partial w_1} = 0 \quad \wedge \quad \frac{\partial C}{\partial b_1} = 0$$

Here

$$C(a_1, y, \theta) = \frac{1}{2}(a_1 - y)^2$$

$$a_1 = \sigma_1(z_1)$$

$$\frac{\partial C}{\partial w_1} = \frac{\partial C}{\partial a_1} \frac{\partial a_1}{\partial z_1} \frac{\partial z_1}{\partial w_1}$$

x

chain rule

$$C = \frac{1}{2}(a_1 - y)^2$$

$$z_1 = x w_1 + b_1$$

$$a_1 = \sigma_1(z_1)$$

$$\frac{d\sigma_1}{dz_1} = \sigma_1'$$

$$\frac{\partial C}{\partial a_1} = (a_1 - y)$$

$$\frac{\partial C}{\partial w_1} = \underbrace{\underbrace{(a_1 - y)}_{\frac{\partial C}{\partial a_1}} \underbrace{\sigma_1'}_{\frac{\partial a_1}{\partial z_1}}}_{\delta_1} \underbrace{x}_{\frac{\partial z_1}{\partial w_1}}$$

$$\boxed{\frac{\partial C}{\partial w_1} = \delta_1 x} \wedge \boxed{\frac{\partial C}{\partial b_1} = \delta_1}$$

Gradient descent

$$w_1 \leftarrow w_1 - \eta \frac{\partial C}{\partial w_1}$$

$$b_1 \leftarrow b_1 - \eta \frac{\partial C}{\partial b_1}$$

hidden layer (all scalar)

$$
\begin{tikzcd}
x \arrow[r, "w_1"] & \boxed{\sigma_1} \arrow[r] & a_1 \arrow[r] & \boxed{\sigma_2} \arrow[r] & a_2 \arrow[d] \\
& b_1 \arrow[u] & & b_2 \arrow[u] & \boxed{C} \\
& & & & y \arrow[u]
\end{tikzcd}
$$

hidden layer

output layer

$z_1 = w_1 x + b_1 \wedge a_1 = \sigma_1(z_1)$

$z_2 = w_2 a_1 + b_2 \wedge a_2 = \sigma_2(z_2)$

$$C = \frac{1}{2}(a_2 - y)^2$$

$$\frac{\partial C}{\partial w_2} = \underbrace{\frac{\partial C}{\partial a_2}}_{a_2 - y} \underbrace{\frac{\partial a_2}{\partial z_2}}_{\sigma_2'} \underbrace{\frac{\partial z_2}{\partial w_2}}_{a_1}$$

$$= \underbrace{(a_2 - y)\sigma_2'}_{\delta_2} a_1$$

$$\frac{\partial C}{\partial b_2} = \delta_2$$

gradient descent update of w_2 and b_2 (output layer parameters)

$$w_2 \leftarrow w_2 - \eta\, \delta_2 a_1$$

$$b_2 \leftarrow b_2 - \eta\, \delta_2$$

$$\delta_2 = (a_2 - y)\, \sigma_2'$$

$$\frac{\partial C}{\partial w_1} = 0 \quad \wedge \quad \frac{\partial C}{\partial b_1} = 0$$

$$\frac{\partial C}{\partial w_1} = \overbrace{\underbrace{\frac{\partial C}{\partial a_2}\frac{\partial a_2}{\partial z_2}}_{\delta_2}\underbrace{\frac{\partial z_2}{\partial a_1}}_{w_1}\underbrace{\frac{\partial a_1}{\partial z_1}}_{\sigma_1'}}^{\delta_1}\underbrace{\frac{\partial z_1}{\partial w_1}}_{x}$$

$$\left(\begin{aligned} z_2 &= a_1 w_2 + b_2 \\ z_1 &= w_1 x + b_1 \end{aligned}\right)$$

$$= \delta_1 x$$

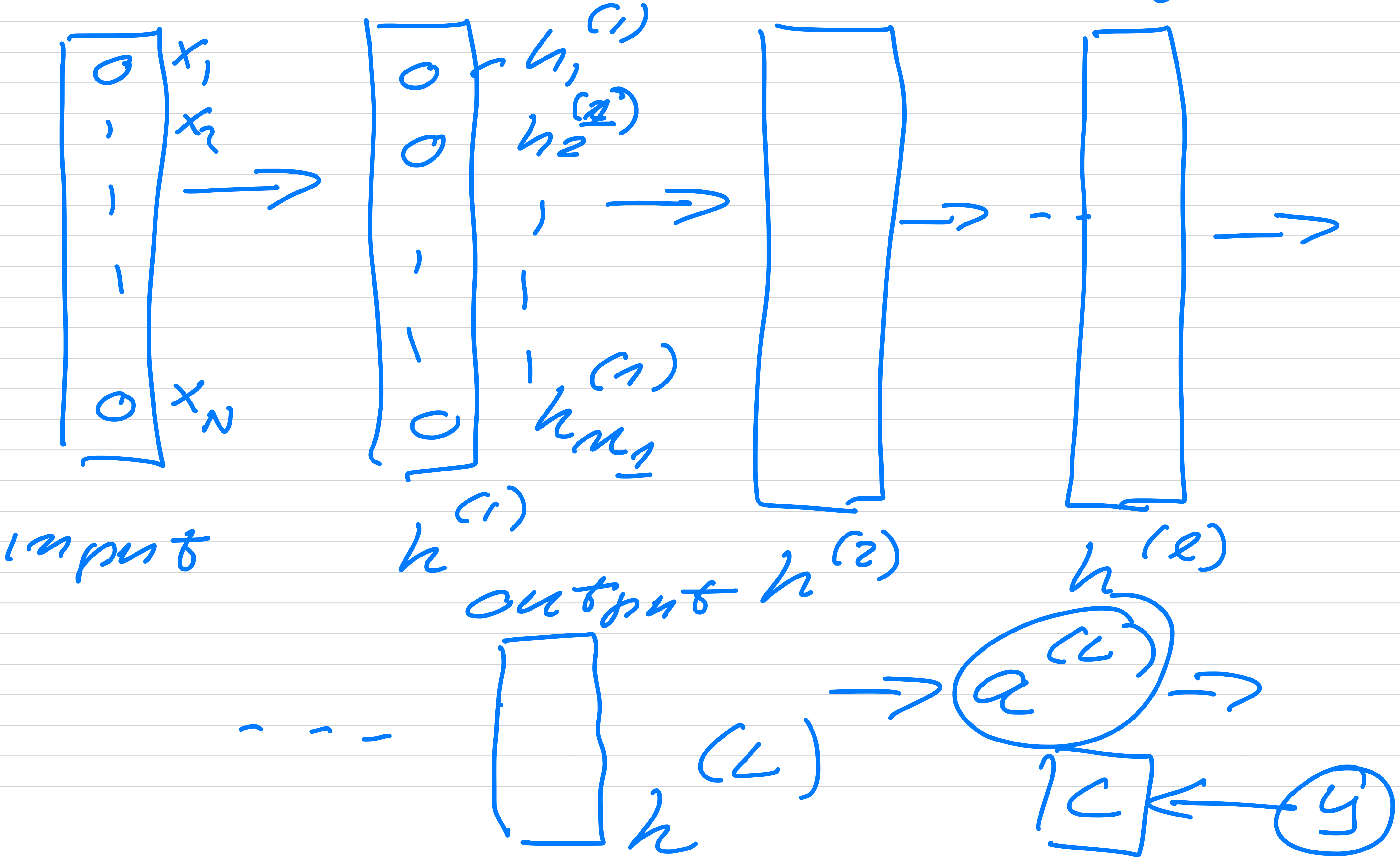
$$\frac{\partial C}{\partial b_1} = \delta_1 \frac{\partial z_1}{\partial b_1} = \delta_1$$

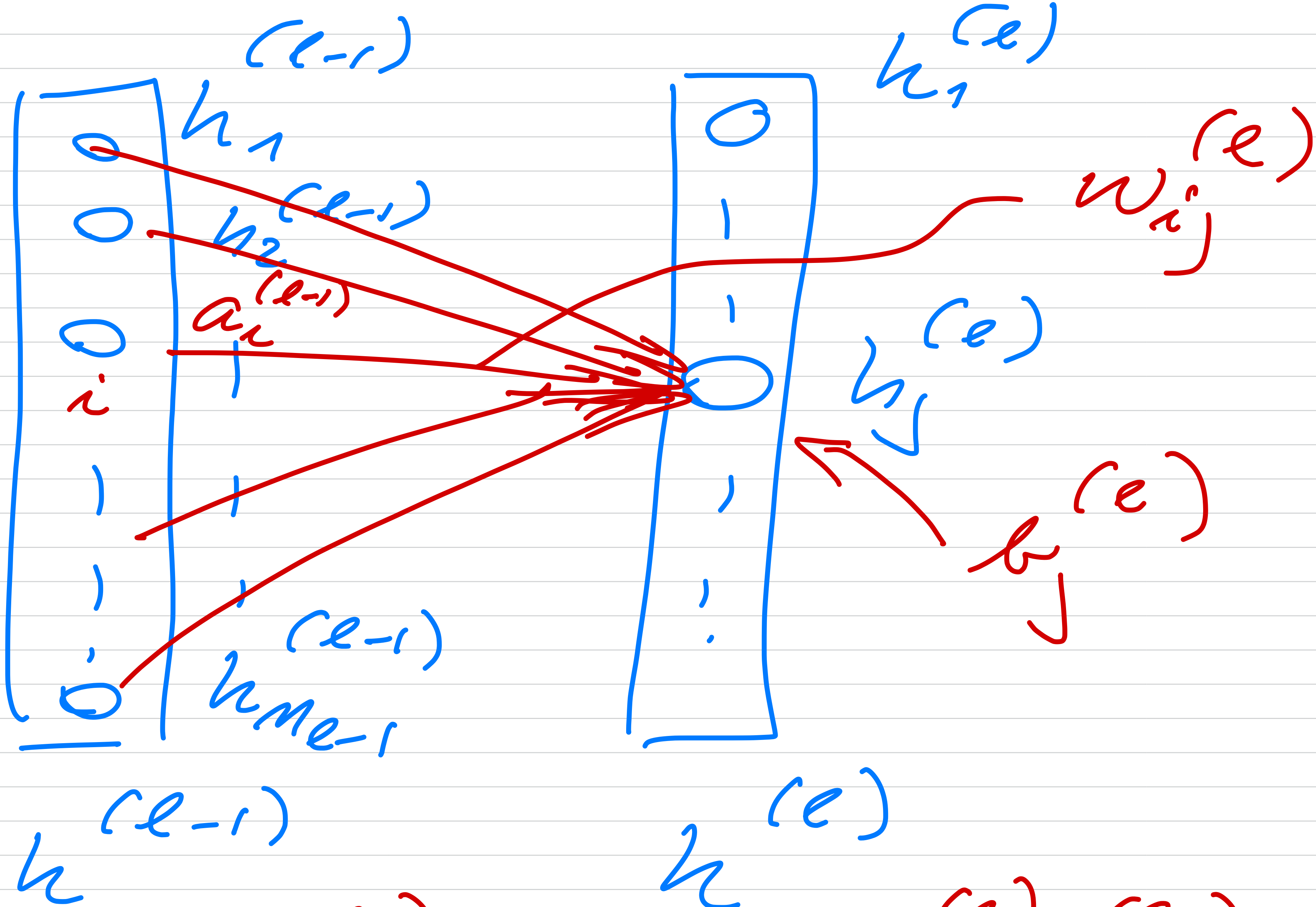
updates hidden layer

$$w_1 \leftarrow w_1 - \eta \delta_1 x$$

$$b_1 \leftarrow b_1 - \eta \delta_1$$

Backpropagation algo





$$z_j^{(l)} = \sum_i w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)}$$

output from node $h_j^{(\ell)}$

$$a_j^{(\ell)} = \sigma\left(z_j^{(\ell)}\right)$$

$$z^{\ell} = \left[w^{(\ell)}\right]^T a^{(\ell-1)} + b^{(\ell)}$$

$$a_k^{(\ell-1)} = \sigma\left(z_k^{(\ell-1)}\right)$$

We start we the output layer $-L-$

$$C(\Theta) = \frac{1}{2} \sum_{i=1}^{n_L} (\hat{y}_i - y_i)^2$$

$$a_i^{(L)}$$

$$\frac{\partial C}{\partial \Theta^{(L)}} = ?$$

some intermediate steps first

$$\frac{\partial z_j^{(\ell)}}{\partial w_{ij}^{(\ell)}} = a_i^{(\ell-1)}$$

$$\left(z_j^{(\ell)} = \sum_i w_{ij}^{(\ell)} a_i^{(\ell-1)} \right)$$

$$\frac{\partial z_j^{(\ell)}}{\partial a_i^{(\ell-1)}} = w_{ij}^{(\ell)}$$

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (\text{example})$$

$$\frac{\partial a_j^{(\ell)}}{\partial z_j^{(\ell)}} = \sigma' = \sigma\left(z_j^{(\ell)}\right)$$

$$\times \left(1 - \sigma\left(z_j^{(\ell)}\right)\right)$$

$$= a_j^{(\ell)}\left(1 - a_j^{(\ell)}\right)$$

$$\ell = L \quad (\text{output})$$

$$\frac{\partial C}{\partial w_{jk}^{(l)}} = \left(a_j^{(l)} - y_j\right) \frac{\partial a_j^{(l)}}{\partial w_{jk}^{(l)}}$$

cost func. specific

$$\frac{\partial a_j^{(l)}}{\partial z_j^{(l)}} \frac{\partial z_j^{(l)}}{\partial w_{jk}^{(l)}}$$

$$= a_j^{(l)}\left(1 - a_j^{(l)}\right) a_k^{(l-1)}$$

activation func specific

$$\delta_j^{(L)} = \sigma' \frac{\partial C}{\partial a_j^{(L)}}$$

$$\underbrace{\qquad\qquad}_{(a_j^{(L)} - y_j)}$$

$$\boxed{\begin{aligned} \frac{\partial C}{\partial w_{jk}^{(L)}} &= \delta_j^{(L)} \cdot a_k^{(L-1)} \\ \frac{\partial C}{\partial b_j^{(L)}} &= \delta_j^{(L)} \end{aligned}}$$

gradients

$$w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \eta \, \delta_j^{(l)} a_k^{(l-1)}$$

$$b_j^{(l)} \leftarrow b_j^{(l)} - \eta \, \delta_j^{(l)}$$

$L \rightarrow e$ and

final expression

next week.