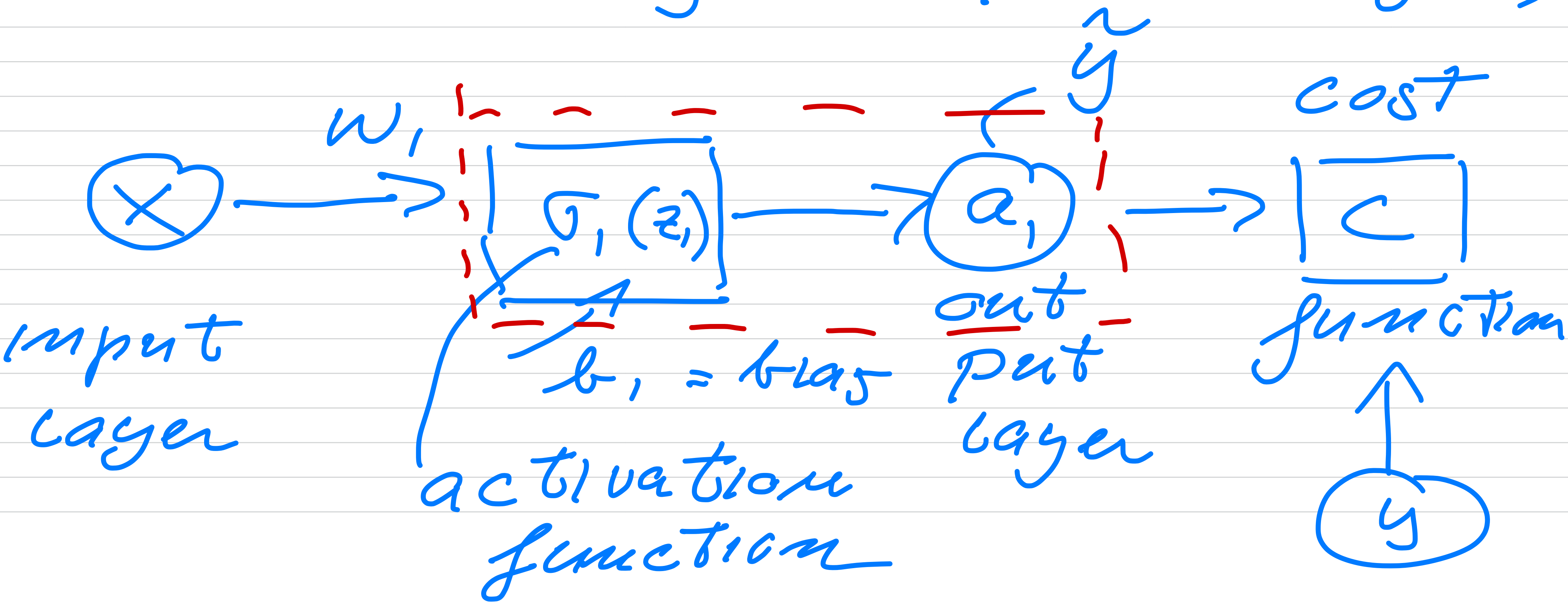
FYS5429/9429 January 29

Basics of FFNN (NN)

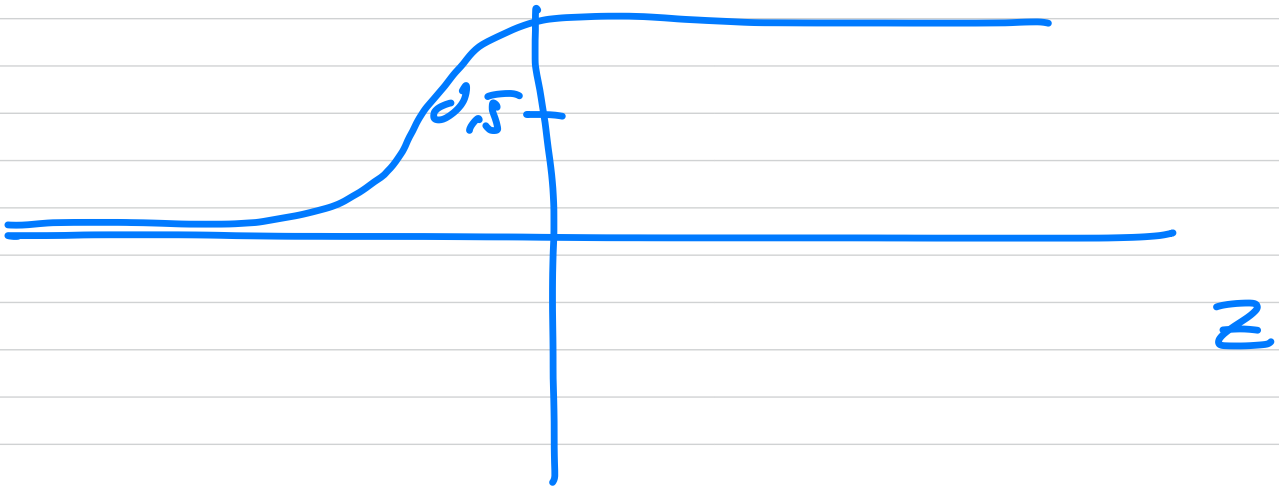
1) No hidden layer
scalar x in
— " — y output (target)



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

$$\sigma_1(z) = \frac{1}{e^{-z}+1} = a_1$$

$$z = w_1 x + b_1$$



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

$$a_1$$

$$\left(\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2 \right)$$

$$C = \frac{1}{2}(\sigma_1(z_1) - y)^2$$

$$w_1 x + b_1$$

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

$$\frac{\partial C}{\partial w_1} = 0 \quad 1 \quad \frac{\partial C}{\partial b_1} = 0$$

$$\frac{\partial C}{\partial w_1} = \frac{\partial C}{\partial a_1} \underbrace{\frac{\partial a_1}{\partial z_1}}_{\sigma_1'} \frac{\partial z_1}{\partial w_1}$$

$$= (a_1 - y)\,\sigma_1'\,x$$

$$z_1 = w_1 x + b_1$$

$$\frac{\partial C}{\partial b_1} = \underbrace{\frac{\partial C}{\partial a_1}\frac{\partial a_1}{\partial z_1}}_{\delta_1}\underbrace{\frac{\partial z_1}{\partial b_1}}_{=1}$$

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

$$\frac{\partial C}{\partial w_1} = \underbrace{\frac{\partial C}{\partial a_1}\frac{\partial a_1}{\partial z_1}}_{\delta_1} \times \frac{\partial z_1}{\partial w_1}$$

$$= \delta_1 x$$

Training of gradients

$$w_1 \leftarrow w_1 - \eta \underbrace{\frac{\partial C}{\partial w_1}}_{\delta_1, x}$$

learning rate

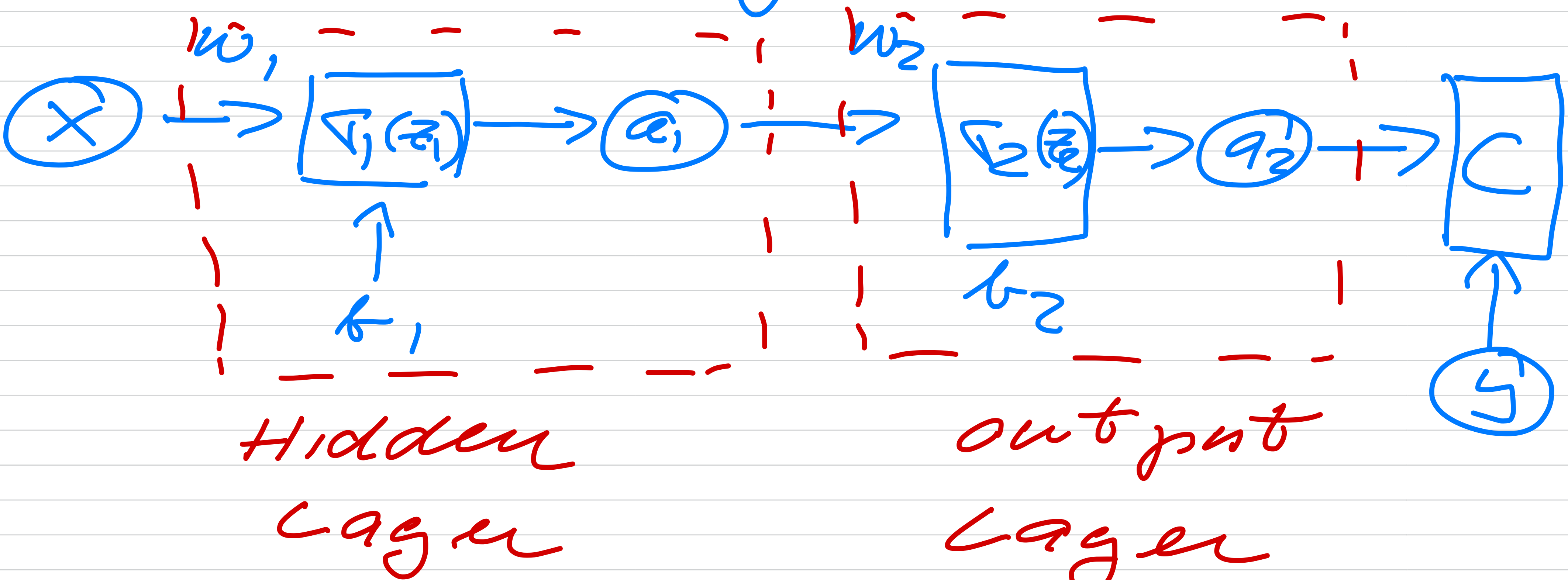
$$b_1 \leftarrow b_1 - \eta \underbrace{\frac{\partial C}{\partial b_1}}_{\delta_1}$$

one iteration

(i) Feed Forward

(ii) Back propagation ($\nabla_6 C$) + gradient descent

1-hidden layer



$$\Theta = \{ w_1, b_1, w_2, b_2 \}$$

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

$$\frac{\partial C}{\partial w_2} = \underbrace{\frac{\partial C}{\partial a_2}\,\frac{\partial a_2}{\partial z_2}}_{\delta_2}\;\underbrace{\frac{\partial z_2}{\partial w_2}}$$

σ_2'

$$z_2 = w_2 \cdot a_1 + b_2$$

$$\frac{\partial z_2}{\partial w_2} = a_1$$

$$\frac{\partial C}{\partial w_2} = \delta_2\, a_1 \qquad \frac{\partial C}{\partial b_2} = \delta_2$$

$$z_2 = w_2 a_1 + b_2$$

$$= w_2 \sigma_1(z_1) + b_2$$

$$= w_2 \sigma_1(w_1 x + b_1) + b_2$$

$$\frac{\partial C}{\partial w_1} = \underbrace{\frac{\partial C}{\partial a_2} \frac{\partial a_2}{\partial z_2}}_{\delta_2} \underbrace{\frac{\partial z_2}{\partial z_1}}_{\underbrace{\frac{\partial z_2}{\partial a_1}}_{w_2} \underbrace{\frac{\partial a_1}{\partial z_1}}_{\sigma_1'}} \underbrace{\frac{\partial z_1}{\partial w_1}}_{x = a_0}$$

$$\frac{\partial C}{\partial b_1} = ?$$

$$\frac{\partial C}{\partial w_1} = \underbrace{\delta_2 \sigma_1' w_2}_{\delta_1} \cdot \overset{= x}{a_0}$$

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

$$z_1 = w_1 x + b_1$$

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

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

$$w_1 \leftarrow w_1 - \eta \cdot \delta_1 \overset{= x}{a_0}$$

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