Lecture October 7, 2024

FYS-STK3155/4155 October 7

Automatic differentiation

$$f(x) = \exp x^2 = \exp a$$

$$a = x^2 \qquad b = \exp a$$

$$\frac{df}{dx} = 2x \exp x^2 = 2x \exp(a)$$

(exp)

$$x \to a \to b \to f$$

$$\frac{df}{dx} = \frac{df}{db}\frac{db}{da}\frac{da}{dx} \qquad b = f$$

$$\boxed{\underbrace{\frac{df}{db}}_{=\,1}\ \underbrace{\frac{db}{da}}_{=\,\exp(a)}}\ \underbrace{\frac{da}{dx}}_{=\,2x} = 2x\exp(a)$$

$$= 2x\exp(x^2)$$

Reverse mode

$$\frac{df}{db}\ \boxed{\frac{db}{da}\ \frac{da}{dx}}$$

Forward mode

Example

$$f(x) = \sqrt{\exp(x^2) + x^2}$$ 5 FLOPS (4 FLOPS)

$a = x^2 \qquad b = \exp x^2 = \exp a$

$c = a + b \qquad d = \sqrt{c} = f(x)$

we want $\dfrac{df}{dx} = \dfrac{x(1+\exp x^2)}{\sqrt{x^2+\exp(x^2)}}$

10 FLOPS

$$= \frac{x(1+b)}{d}$$

3 FLOPS

$$\frac{da}{dx} = 2x \qquad \frac{db}{dx} = \frac{db}{da}\frac{da}{dx}$$

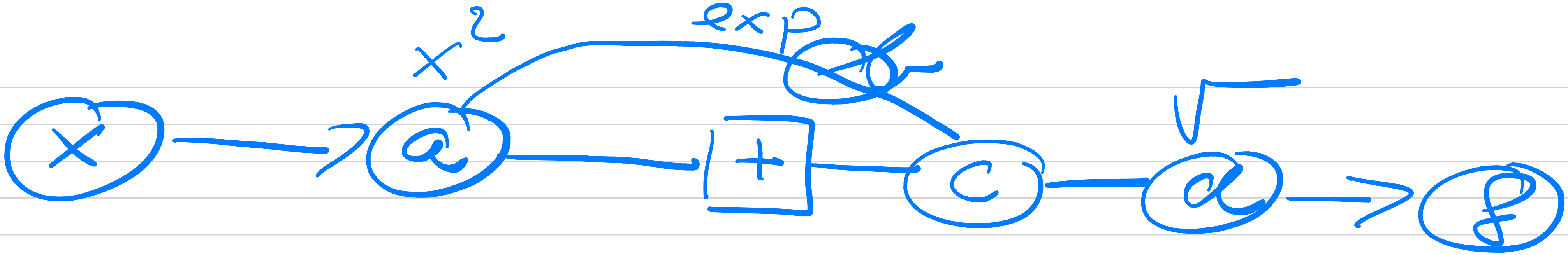
$$= 2x \exp x^2$$

$$= 2x \exp a$$

$$\frac{dc}{dx} = \left[\underbrace{\frac{dc}{da}}_{=1} \underbrace{\frac{da}{dx}}_{2x} + \underbrace{\frac{dc}{db}}_{=1} \underbrace{\frac{db}{da}\frac{da}{dx}}_{2x \exp x^2} \right]$$

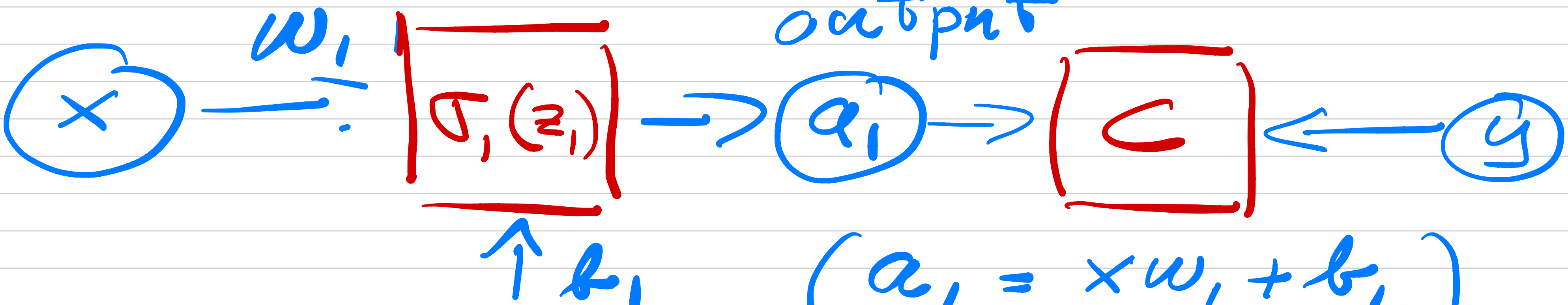
$$\frac{dd}{dc} = \frac{1}{2\sqrt{c}}$$

$$\frac{dd}{dx} = \frac{dd}{dc}\frac{dc}{dx} = \frac{df}{dx}$$



our own code for a neural network.

1) simple perceptron model



$(a_1 = x w_1 + b_1)$ model

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

$C = C(a_1, y; \theta)$

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

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

Example of $\sigma_1 = \frac{1}{1+e^{-z_1}}$

chain rule

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

$$a_1 = \sigma_1(z_1) \qquad \frac{\partial a_1}{\partial z_1} = \sigma_1'$$

$$z_1 = x w_1 + b_1$$

$$\frac{\partial z_1}{\partial w_1} = x$$

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

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

$$\frac{\partial C}{\partial b_1} = \underbrace{(a_1 - y)}\,\sigma_1' = \delta_1$$

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

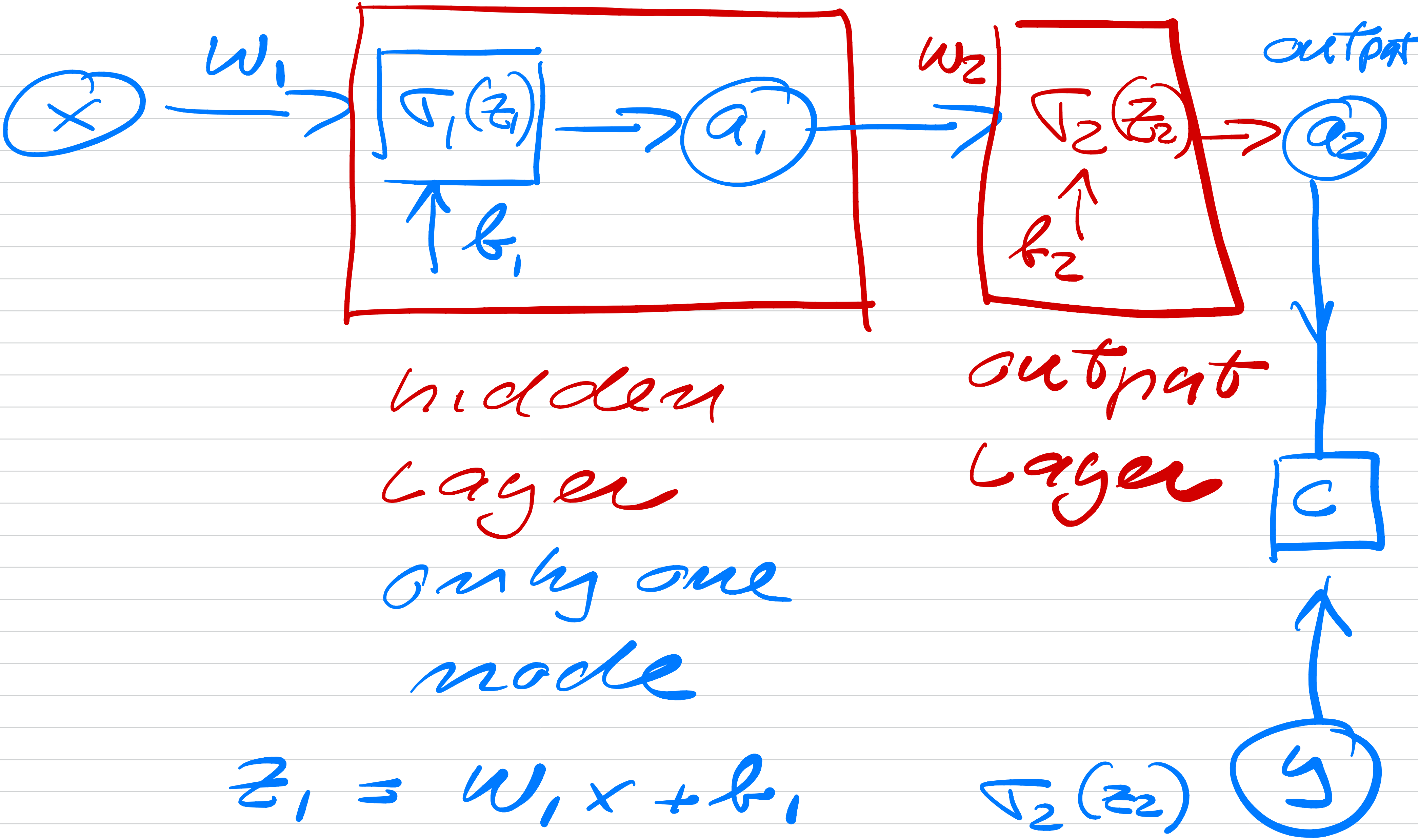
Gradient descent

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

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

one hidden layer but

scalar x and y



$$z_1 = w_1 x + b_1 \qquad \sigma_2(z_2)$$

$$z_2 = a_1 w_2 + b_2 \qquad \sigma_1(z_1)$$

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

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

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

$$a_2 = a_2(w_2, b_2)$$

$$z_2 = w_2 a_1 + b_2$$

$$\frac{\partial C}{\partial w_2} = \underbrace{(a_2 - y)\sigma_2'}_{\delta_2} \cdot a_1$$

$$w_2 \leftarrow w_2 - \eta \frac{\partial C}{\partial w_2}$$

$$\frac{\partial C}{\partial b_2} = \frac{\partial C}{\partial a_2} \frac{\partial a_2}{\partial z_2} \underbrace{\frac{\partial z_2}{\partial b_2}}_{=1} = \delta_2$$

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

$$b_2 \leftarrow b_2 - \eta \frac{\partial C}{\partial b_2}$$

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

$$= \delta_1 x$$

$$z_1 = w_1 x + b_1$$

$$z_2 = a_1 w_2 + b_2$$

$$\frac{\partial C}{\partial b_1} = \underbrace{\frac{\partial C}{\partial a_2}\frac{\partial a_2}{\partial z_2}\frac{\partial z_2}{\partial a_1}\frac{\partial a_1}{\partial z_1}}_{\delta_1}\underbrace{\frac{\partial z_1}{\partial b_1}}_{=0}$$

$$w_1 \Leftarrow w_1 - \eta\frac{\partial C}{\partial w_1} \wedge b_1 \Leftarrow b_1 - \eta\frac{\partial C}{\partial b_1}$$