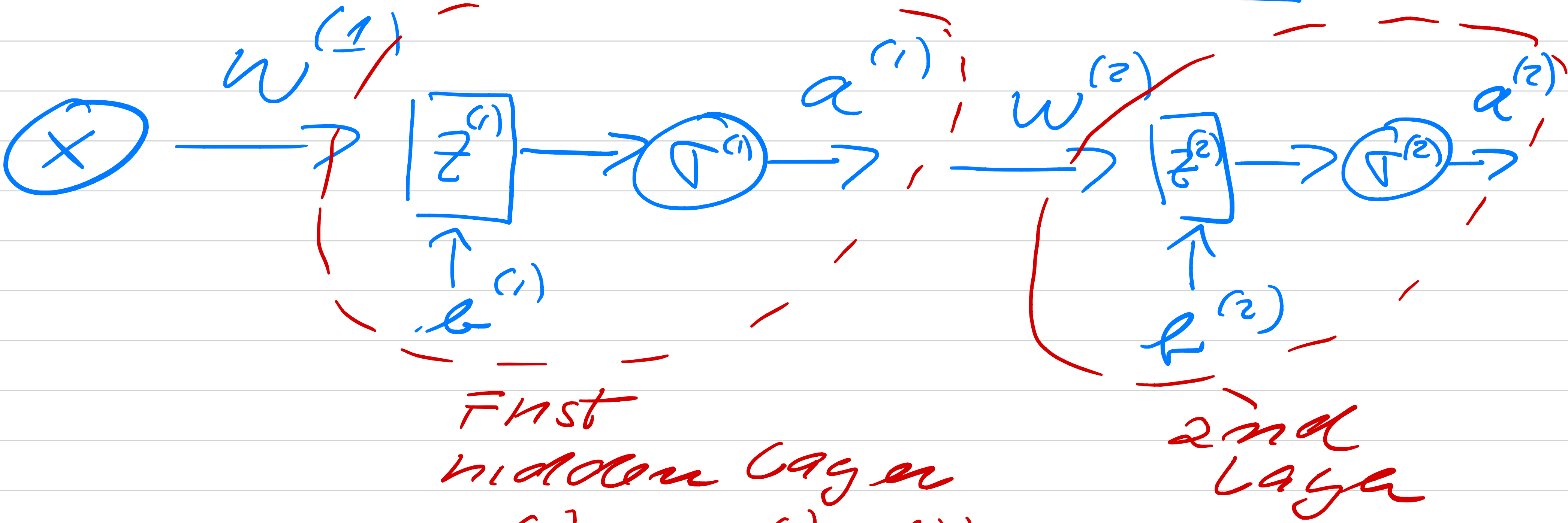
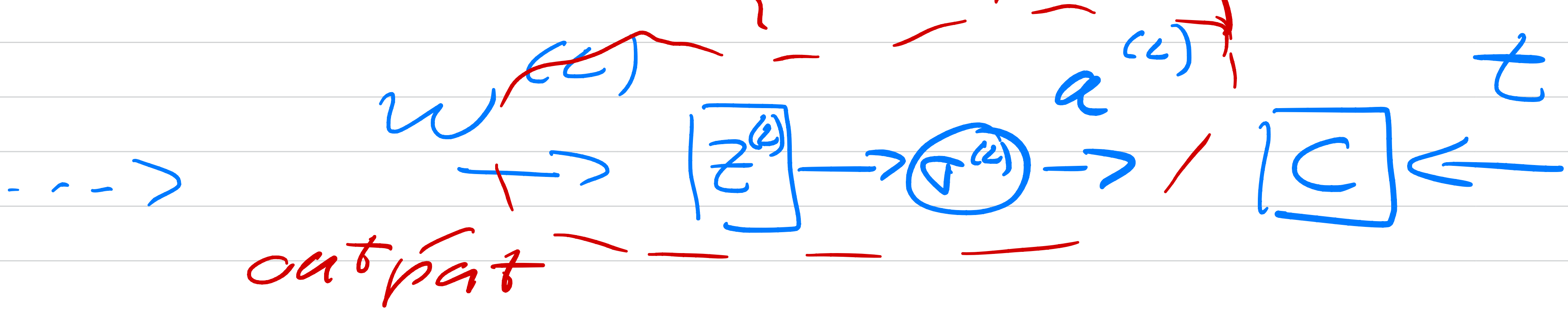
Erasmus+ course on Machine Learning, December 12, 2023

Neural Network Layout



$$a^{(1)} = \sigma^{(1)}(z^{(1)})$$

$$\theta^{(1)} = \{ w^{(1)}, b^{(1)} \}$$



Final output $a^{(L)} = a^{(L)}(\Theta; x)$

$$\Theta = \left\{ \Theta^{(1)}, \Theta^{(2)}, \ldots \Theta^{(L)} \right\}$$

$$= \left\{ w^{(1)}, b^{(1)}, w^{(2)}, b^{(2)}, \ldots w^{(L)}, b^{(L)} \right\}$$

$$C(a^{(L)}, t) = C(a^{(L)}(\Theta, x), t)$$

$$C(a^{(L)}, t) = \| t - a^{(L)}(\Theta; x) \|_2^2 + \lambda \|\Theta\|_2^2$$

Example for regression problem

$$\frac{\partial C}{\partial \theta^{(L)}}$$ which ends being a function of all

$$\Theta = \{ \theta^{(1)}, \theta^{(2)}, \dots \Theta^{(L)} \}$$

Expression at final layer

$$\frac{\partial C}{\partial \theta^{(L)}} = \frac{\partial C}{\partial a^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}} \frac{\partial z^{(L)}}{\partial \Theta^{(L)}}$$

$$z^{(L)} = w^{(L)} a^{(L-1)} + b^{(L)}$$

$$\delta^{(L)} = \underbrace{\left(\sigma^{(L)}(z^L)\right)'}_{\frac{\partial a^{(L)}}{\partial z^{(L)}}} \frac{\partial C}{\partial a^L}$$

$$\ell = L-1,\ L-2,\ \ldots\ \underline{1}$$

$$\delta_j^{(\ell)} = \sum_k \delta_k^{(\ell+1)} w_{kj}^{(\ell+1)} \sigma'\left(z_j^{(\ell)}\right)$$

$$w_{jk}^{(\ell)} \leftarrow w_{jk}^{(\ell)} - \eta\, \delta_j^{(\ell)} a_k^{(\ell-1)}$$

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

$$\frac{\partial C}{\partial B^{(L-1)}} = \frac{\partial C}{\partial B^{(L)}} \frac{\partial B^{(L)}}{\partial B^{(L-1)}}$$

and then continue till first layer.

Automatic differentiation (reverse mode) = Back prop.

Not symbolic differentiation

Not finite difference

$$f'(x) \simeq \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

Example

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

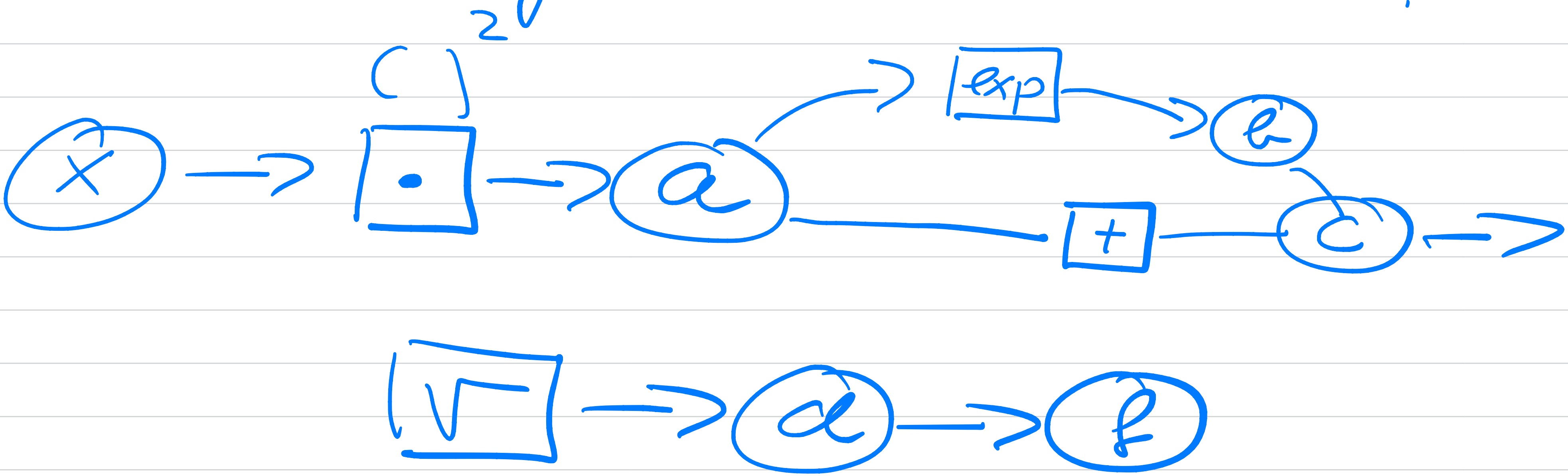
Number of operations

1) $x \cdot x$ 1 FLOP
2) addition $x^2 + \exp(x^2)$ 1 Flop
3) $\exp(x^2)$ 2 FLOP
4) $\sqrt{\text{arg}}$ 1 FLOP

5 FLOPs

$$\frac{df}{dx} = \frac{x\left(1+\exp(x^2)\right)}{\sqrt{x^2+\exp(x^2)}}$$

Brute force 10 FLOPS ?



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

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

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

$$\frac{dc}{dx} = \left[\frac{dc}{da}\frac{da}{dx} + \frac{dc}{db}\frac{db}{dx}\right]$$

$$\frac{dc}{da} = \frac{db}{dc} = 1$$

$$\frac{dc}{dx} = \frac{da}{dx} + \frac{db}{da}\frac{dg}{dx}$$

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

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

we want $\frac{df}{dx}$

walking backward

$$\frac{df}{dd} = 1$$

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

$$\frac{df}{db} = \frac{df}{dc}\,\frac{dc}{db} = \frac{1}{2\sqrt{c}}$$

$$\frac{df}{da} = \frac{df}{db}\,\frac{db}{da} + \frac{df}{dc}\,\frac{dc}{da}$$

$$= \frac{1}{2\sqrt{c}}\left[1 + \exp(a)\right]$$

$$\frac{df}{dx} = \frac{df}{da}\,\frac{da}{dx} = \frac{x(1+\exp(a))}{\sqrt{c}}$$

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

$$d = \sqrt{x^2 + \exp(x^2)}$$

assume we have $x_1, \ldots x_d$ input variables to f, x_{d+1}, $x_{d+2}, \ldots x_{D-1}$ intermediate variables, x_D is the output

$x_1 = x \qquad d = \underline{1}$

$x_2 = a \qquad x_3 = b \qquad x_4 = c = a+b$

$x_D = x_5 = d = f$

also

for $i = d+1, D$

$x_i = g_i(x_{Pa(x_i)}))$

$$g_2 = (\cdot)^2 = a$$

$$g_3 = \exp(\cdot) \quad \swarrow$$

$$g_4 = a + b$$

$$g_5 = \sqrt{c} = d$$

By definition $\dfrac{df}{dx_D} = \underline{1}$

Reverse mode $\dfrac{\partial f}{\partial x_i} =$

$$\sum_{x_j} \frac{\partial f}{\partial x_j} \frac{\partial g_j}{\partial x_i}$$

$$x_i = Pa(x_j)$$

$$\frac{\partial f}{\partial a} = 1 \qquad \frac{\partial f}{\partial c} = \underbrace{\frac{\partial f}{\partial d}}_{=1} \frac{\partial d}{\partial c} = \frac{1}{2\sqrt{c}}$$

$$\frac{\partial f}{\partial b} = \frac{\partial f}{\partial c} \frac{\partial c}{\partial b} = \frac{1}{2\sqrt{c}}$$

$$\frac{\partial f}{\partial a} \rightarrow \frac{\partial f}{\partial x} = \frac{\partial f}{\partial a} \frac{\partial a}{\partial x}$$

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

Ordinary Diff eqs (ODEs)

$$\frac{df}{dx} = g(x)$$

$$f = f(x)$$

$$\mathcal{L}(x, g) = \frac{df}{dx} - g(x) = 0$$

initial conditions

$$f(x_0) = f_0$$

discretize eqs

$x \Rightarrow x_i = x_0 + i\Delta x \quad i = 0, 1, 2, \ldots n$

$$\Delta x = \frac{x_n - x_0}{n}$$

$$f(x) \Rightarrow f(x_i) = f_i$$

$$f_{i\pm 1} = f(x_i \pm \Delta x)$$

$$\frac{df}{dx} \approx \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x}$$

$$= (f_{i+1} - f_i)/\Delta x \quad \Rightarrow$$

$$f_{i+1} = f_i + \Delta x \left.\frac{df}{dx}\right|_{x_i}$$

$$= f_i + \Delta x\, f_i'$$

$$f_i' = g(x_i) = g_i$$

$$f_{i+1} = f_i + \Delta x\, g_i$$ known

with neural networks

$x_i,\ f_i,\ g_i$

The solution is approxima-
ted with a neural network

$$f_t(x) = h_1(x) + h_2(x, N(x;\theta))$$

neural network

θ = parameters of neural network

$$C(x;\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[\mathcal{L}(x, f', f'' \ldots)\right]^2$$

$$\mathcal{L} = \left[\frac{df}{dx} - g(x)\right]$$