- Neural Network code

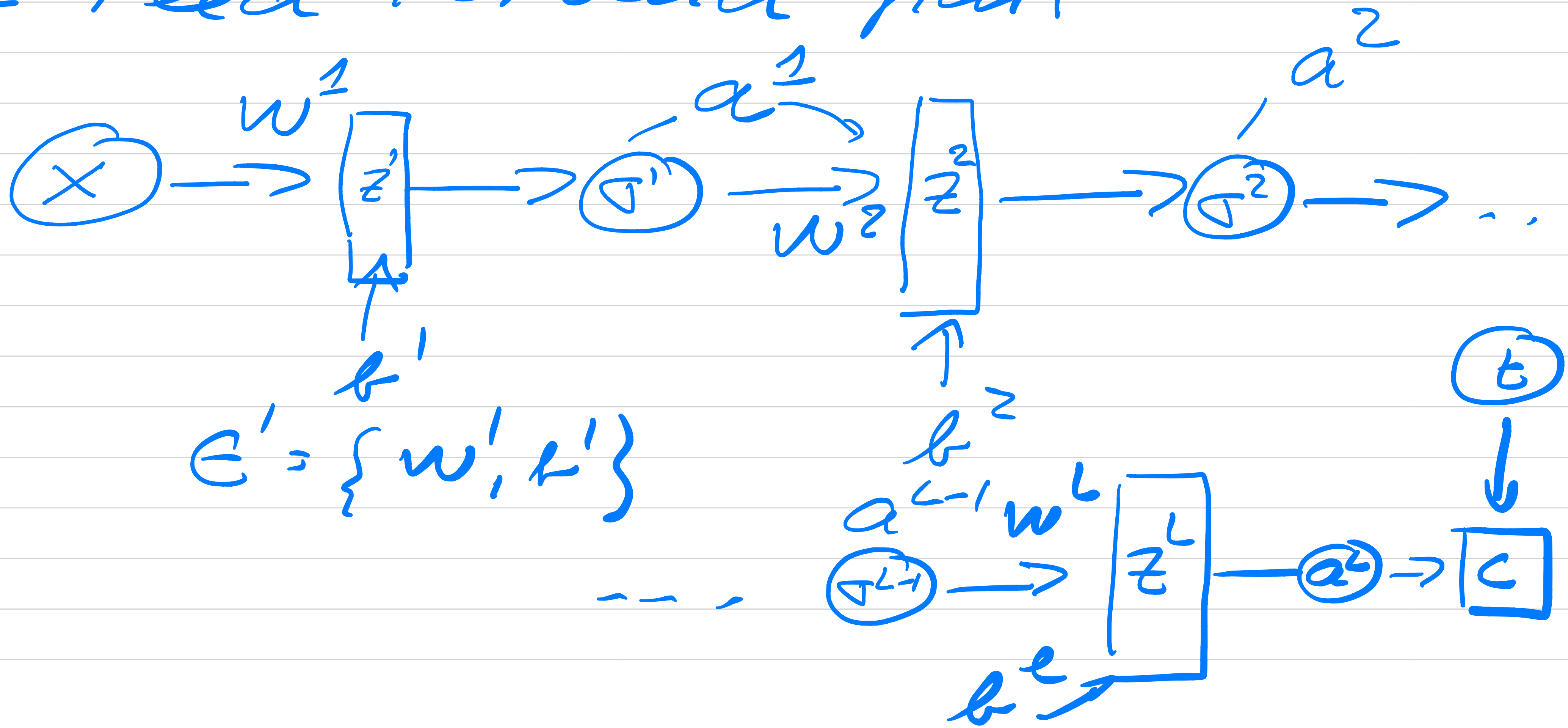
Define architecture (model)
  - # layers
  - # nodes
  - activation functions

Define cost function
  includes hyperparameter λ

Define input x and output
  target _t_

Gradient + Back propagation
  - choose gradient method
    and learning

- initialize parameters $\Theta = \{W, b\}$
- set up λ and learning rate η
- Feed Forward part



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

- Back propagation part

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

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

only w

calculate $\frac{\partial C}{\partial a^L}$

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

For $l = L-1, L-2, \dots l, \dots 1$

$$\boxed{\begin{aligned} \delta_j^l &= \sum_k \delta_k^{l+1} w_{kj}^{l+1} \left(\sigma^l(z_j^l)\right)' \\ w_{jk}^l &\leftarrow w_{jk}^l - \eta\, \delta_j^l a_k^{l-1} \\ b_j^l &\leftarrow b_j^l - \eta\, \delta_j^l \end{aligned}}$$

$$\frac{\partial C}{\partial \theta^{L}} = \frac{\partial C}{\partial a^{L}} \frac{\partial a^{L}}{\partial \theta^{L}}$$

$$\frac{\partial C}{\partial \theta^{L-1}} = \frac{\partial C}{\partial a^{L}} \frac{\partial a^{L}}{\partial \theta^{L-1}}$$

$$\frac{\partial C}{\partial \theta^{L-2}} = \frac{\partial C}{\partial a^{L}} \underbrace{\frac{\partial a^{L}}{\partial a^{L-1}}}_{\text{derivative wrt input}} \underbrace{\frac{\partial a^{L-1}}{\partial \theta^{L-2}}}_{\text{Derivative wrt the parameter of layer } L-2}$$

$$\frac{\partial C}{\partial \epsilon^{l}} = \frac{\partial C}{\partial a^{L}} \frac{\partial a^{L}}{\partial a^{L-1}} \cdots \frac{\partial a^{l+2}}{\partial a^{l+1}} \frac{\partial a^{l+1}}{\partial \epsilon^{l}}$$

Link with autodiff

$$f(x) = \sqrt{x^2 + \exp(x^2)} \qquad \text{5 FLOPs}$$

$$\frac{df(x)}{dx} = \frac{x(1+\exp(x^2))}{\sqrt{x^2+\exp(x^2)}} \qquad \text{10 FLOPs}$$

$$\begin{tikzcd}
x \arrow[r] & \boxed{\cdot}^{\,(x)^2} \arrow[r] & a \arrow[r] \arrow[rrrr, bend left] & \boxed{\exp} \arrow[r] & b \arrow[r] & \boxed{+} \arrow[r] & c \arrow[r] & {}
\end{tikzcd}$$

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

$$\to \boxed{\sqrt{}} \to d \to f(x)$$

$$c = a + b$$

$$d = \sqrt{c} = f(x)$$

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

$$\frac{dc}{dx} = \left[\overset{=1}{\frac{dc}{da}} \underset{2x}{\frac{da}{dx}} + \overset{=1}{\frac{dc}{db}} \frac{db}{dx} \right]$$

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

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

$$= \frac{x(1+b)}{d} \leftarrow 3\text{ Flops}$$

$$f(x) = \underbrace{d = \sqrt{a + \exp(x)}}_{5}$$

we compute $\frac{df}{dx}$ going

Backward, reverse mode

$$\frac{df}{dd} = 1 \qquad d = f$$

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

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

$$c = a+b$$

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

$$= \frac{1}{2\sqrt{c}}\,[1 + \exp(a)] = \frac{1}{2d}(1+e)$$

$$\frac{df}{dx} = \frac{df}{da}\underbrace{\frac{da}{dx}}_{2x} = \frac{x(1+b)}{d}$$

Autodiff:

$x_1, x_2, \dots x_d$ inputs variables
(here $x = x_1$, $d = 1$)

to f, $x_{d+1}, \dots x_{D-1}$
intermediate variables

x_D is the output variable

$x_1 = x$, $d = 1$ $x_2 = a$ $x_3 = b$ $x_4 = c$
$x_D = d = f$

for $i = d+1, D$

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

elementary functions

parent nodes

$g_2 = (x)^2 = a$ parent node

$g_3 = \exp(a) = b$

$g_4 = a + b = c$

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

The reverse mode is

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

$$x_i = pa(x_j)$$

$$\frac{\partial f}{\partial d} = 1$$

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

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

$$\frac{\partial f}{\partial a} = \frac{\partial f}{\partial b}\frac{\partial b}{\partial a} + \frac{\partial f}{\partial c}\frac{\partial c}{\partial a}$$

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

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