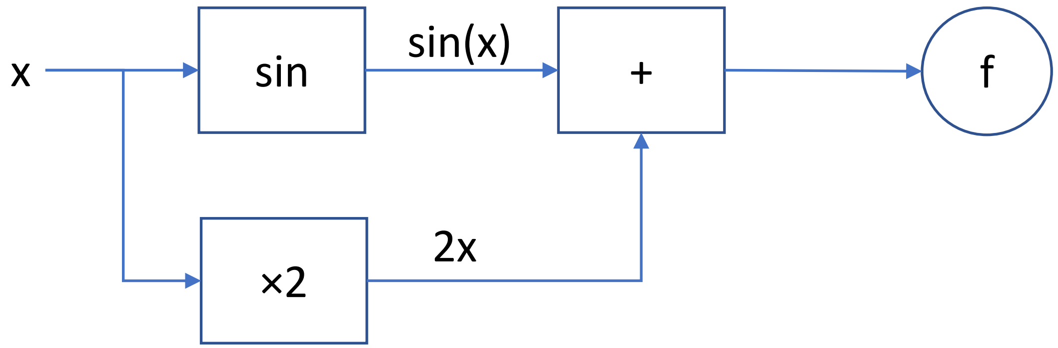
What is a computation graph?

- A computation graph (CG) is a way of modeling a function into intermediate steps.
  - Can be broken down into the composite steps that make up the function

$$f(x) = \sin(x) + 2x$$



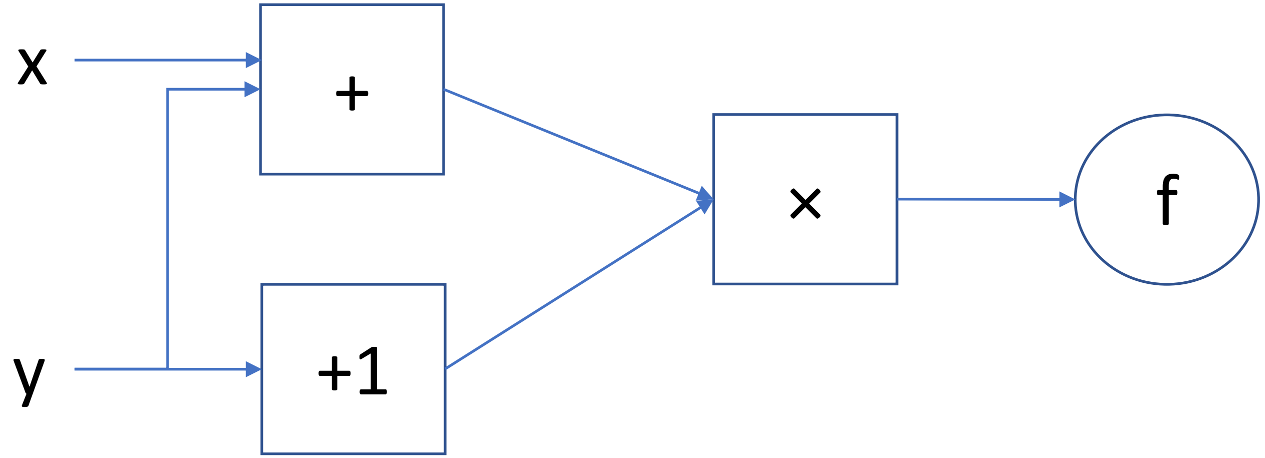
$$= \sin(x) + 2x$$

A neural network is nothing more then a fancy equation.

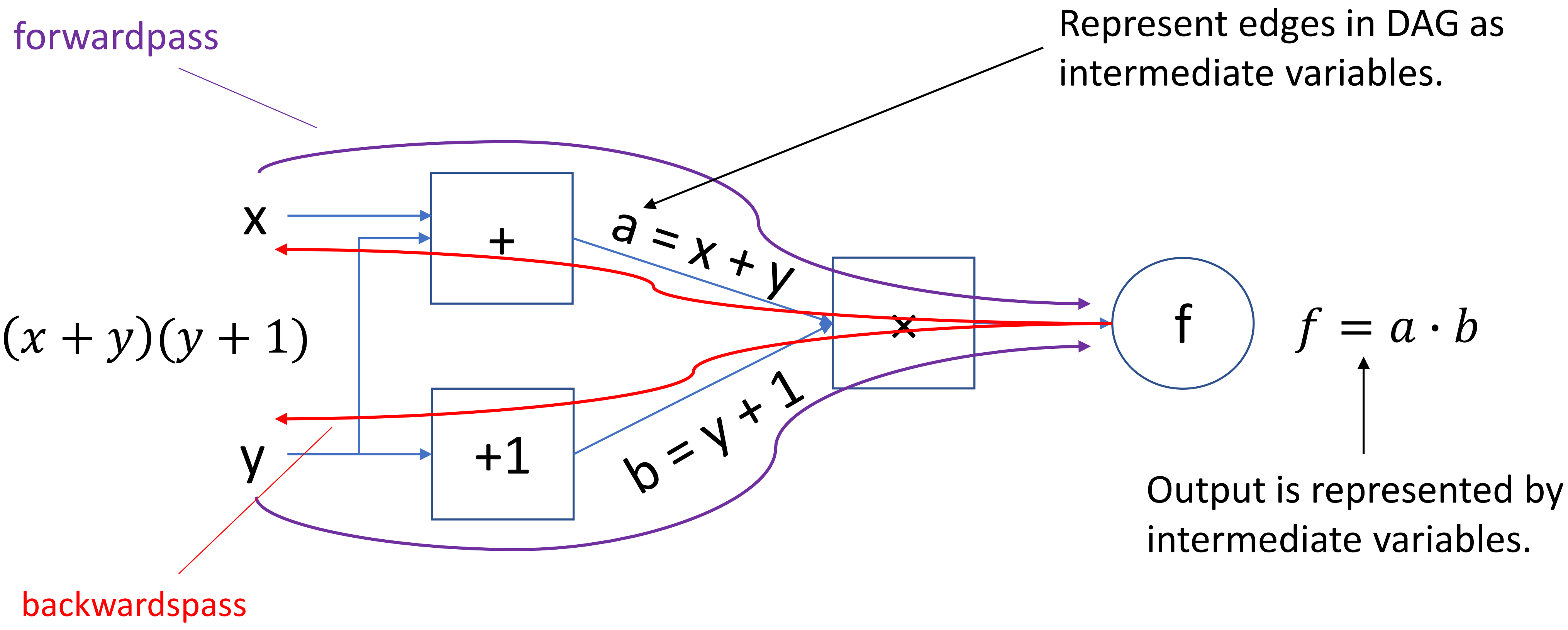
Why do this?

- It's a way of modeling the forward flow of a function.
- You can use it to model the gradients as they flow through your graph
  - Either forward or backwards.

$$f(x, y) = (x + y)(y + 1)$$



Make use of intermediate variables



$$f(x, y) = (x + y)(y + 1)$$

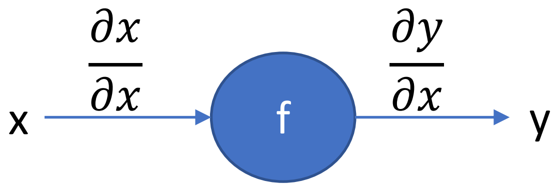
$$f = a \cdot b$$

Can use intermediate variable to help calculate partial derivatives: $\frac{\partial f}{\partial a}, \frac{\partial f}{\partial b} \rightarrow \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$

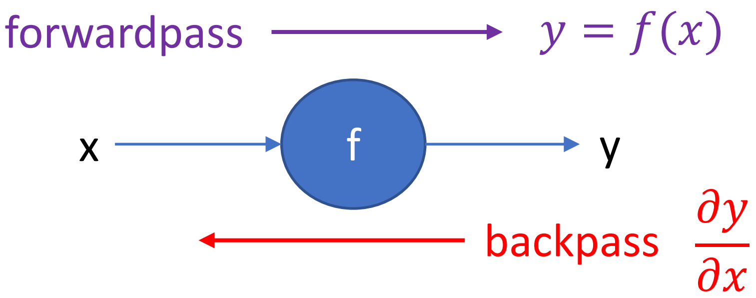
Automatic Differentiation (autodiff)

- Forward-mode Autodiff
  - Gradients are passed forward as you go through the network.

input -> output

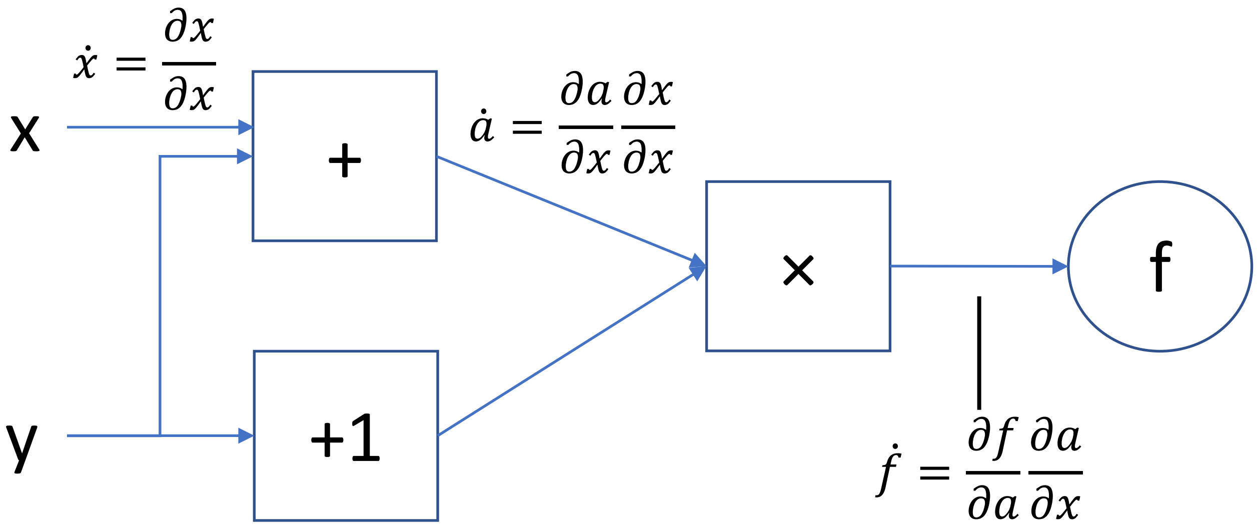


- Reverse-mode Autodiff
  - Gradients start from the output and flow back to the input



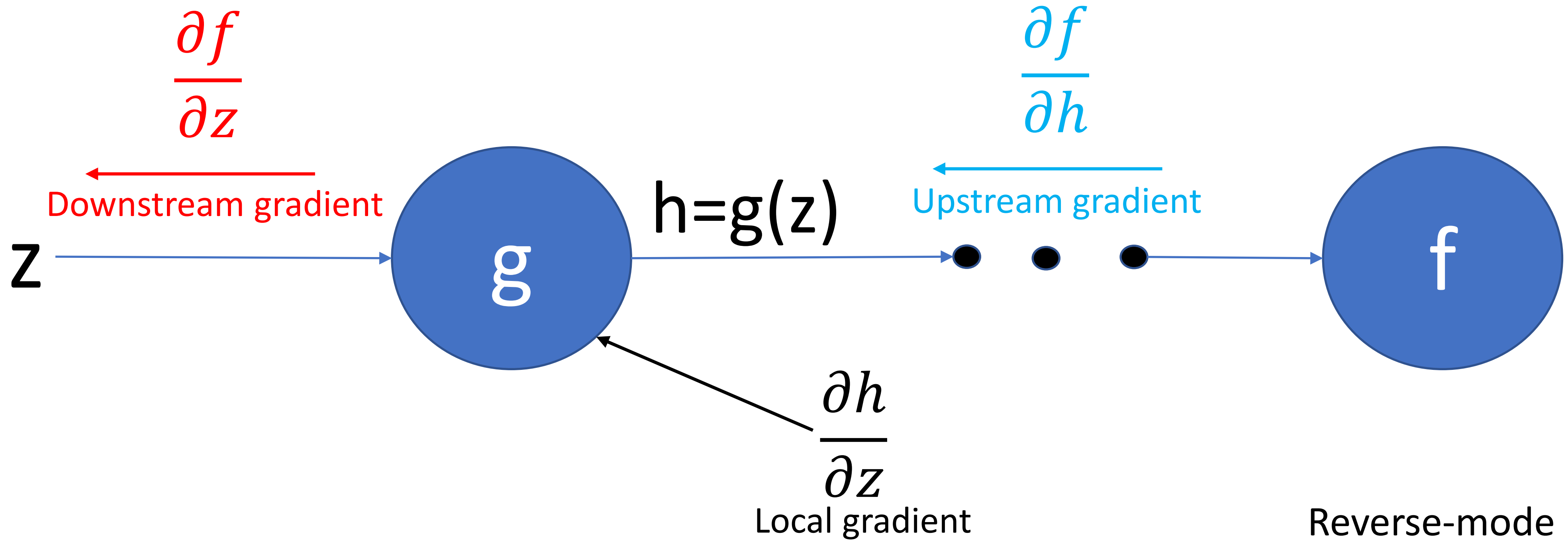
Forward-mode autodiff

Gradients are calculated in a single pass as you go forward in the network



Forward-mode is generally good when the number of inputs (n) is much less then outputs (m): n << m

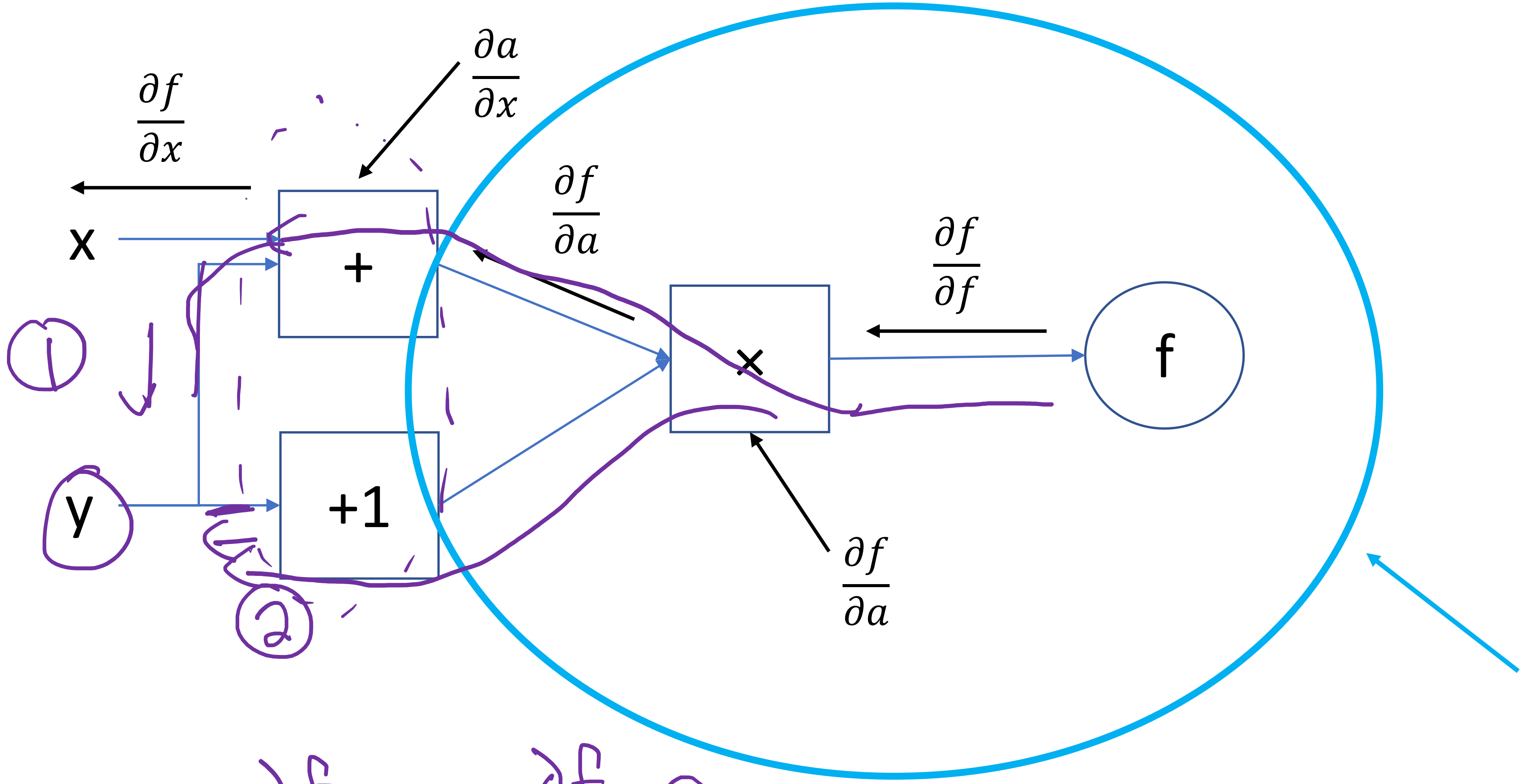
Reverse-mode autodiff



$$DG = UG \times LG \rightarrow \frac{\partial f}{\partial z} = \frac{\partial f}{\partial h} \cdot \frac{\partial h}{\partial z}$$

Reverse-mode is good when the inputs are much greater than the outputs: n >> m

Reverse-mode Example



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

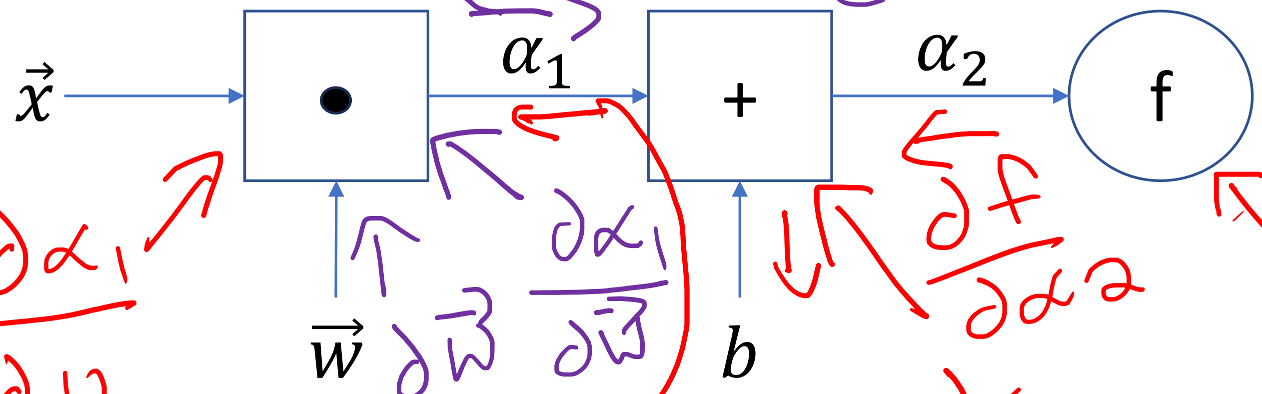
$$\left.\frac{\partial f}{\partial y}\right)_{TOT} = \frac{\partial f}{\partial y}\, 1 + \frac{\partial f}{\partial y}\, 2$$

Example: CG for a simple linear network

Function: $f = \vec{x} \cdot \vec{w} + b$

$$\frac{\partial \alpha_1}{\partial \vec{w}} = \frac{\partial \vec{w}}{\partial \vec{w}} \frac{\partial \alpha_1}{\partial \vec{w}}$$

$$\frac{\partial \alpha_2}{\partial \alpha_1} \qquad \frac{\partial \alpha_2}{\partial \vec{w}} = \frac{\partial \alpha_1}{\partial \vec{w}} \frac{\partial \alpha_2}{\partial \alpha_1} = \frac{\partial f}{\partial \vec{w}}$$



$\alpha_1 + b$

$= \vec{x} \cdot \vec{w} + b$

$$\frac{\partial f}{\partial \alpha_2} \qquad \frac{\partial f}{\partial f} \qquad \frac{\partial f}{\partial \alpha_2} \qquad \frac{\partial f}{\partial \alpha_2} = \frac{\cancel{\partial f}}{\partial f} \frac{\cancel{\partial f}}{\partial \alpha_2}$$

$$\frac{\partial \alpha_1}{\partial w} \qquad \frac{\partial \vec{w}}{\partial w} \qquad \frac{\partial \alpha_1}{\partial \vec{w}}$$

$$\frac{\partial \alpha_2}{\partial b} \Rightarrow \frac{\partial f}{\partial b} = \frac{\partial f}{\partial \alpha_2} \frac{\partial \alpha_2}{\partial b}$$

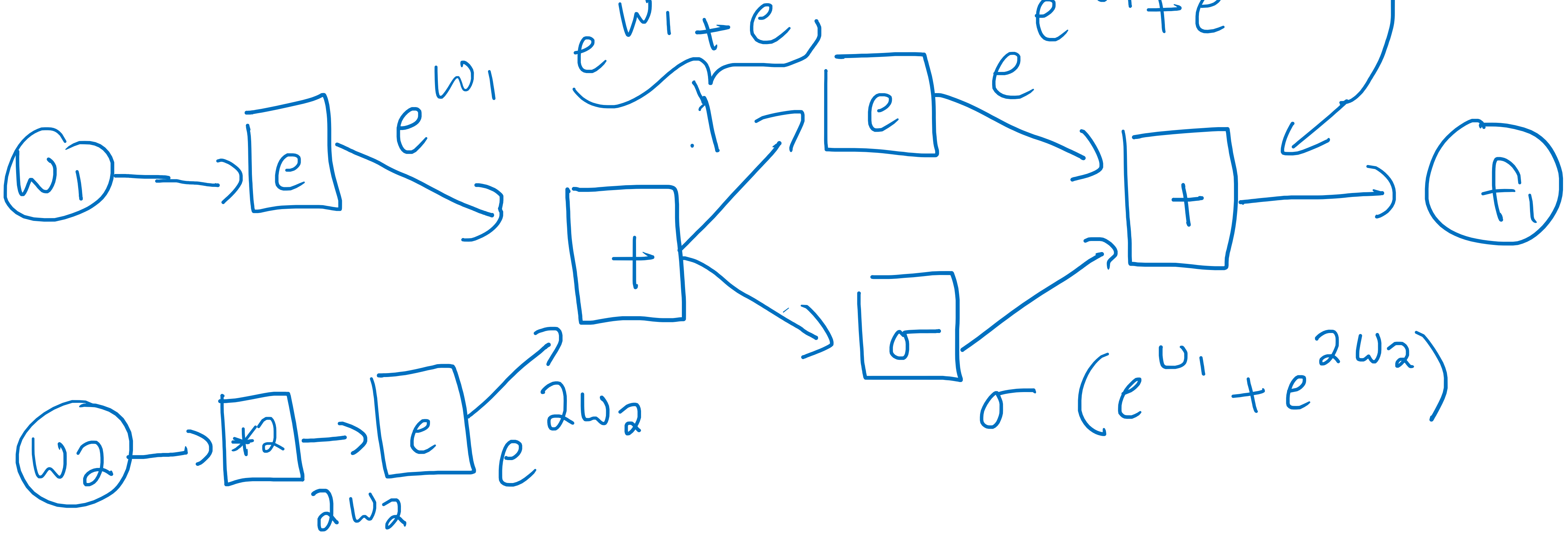
$$\frac{\partial f}{\partial \vec{w}} = \frac{\partial f}{\partial \alpha_1} \frac{\partial \alpha_1}{\partial \vec{w}} \qquad \frac{\partial f}{\partial \alpha_1} = \frac{\partial f}{\partial \alpha_2} \frac{\partial \alpha_2}{\partial \alpha_1}$$

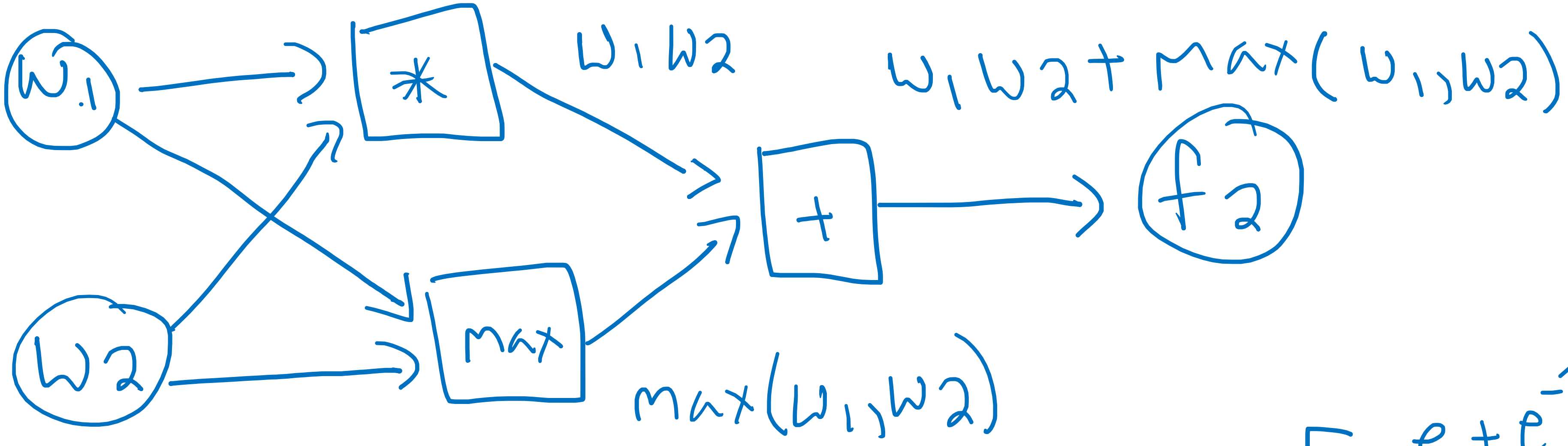
$$f_1(w_1, w_2) = e^{e^{w_1} + e^{2w_2}} + \sigma\left(e^{w_1} + e^{2w_2}\right)$$

$$f_2(w_1, w_2) = w_1 w_2 + \max(w_1, w_2)$$

$$e^{e^{w_1} + e^{2w_2}} + \sigma\left(e^{w_1} + e^{2w_2}\right)$$

(a) Draw out the CG





$$\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow \vec{f} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} e^{e+e^{-2}} + \sigma(e+e^{-2}) \\ (1)(-1) + \max(1,-1) \end{bmatrix}$$

$$= \begin{bmatrix} 18.296 \\ 0 \end{bmatrix}$$

(b)

$$\frac{\partial \vec{f}}{\partial \vec{w}} = f \downarrow \begin{array}{c} \quad w \longrightarrow \\ \begin{bmatrix} \dfrac{\partial f_1}{\partial w_1} & \dfrac{\partial f_1}{\partial w_2} \\ \dfrac{\partial f_2}{\partial w_1} & \dfrac{\partial f_2}{\partial w_2} \end{bmatrix} \end{array} \qquad \Delta w = 0.01$$

$$\frac{\partial f_1}{\partial w_1} = \frac{f_1(w_1 + \Delta w, w_2) - f_1(w_1, w_2)}{\Delta w} = \frac{18.778 - 18{,}296}{0.01} = 48{,}192$$

$$f_1(1 + 0.01, -1) = e^{e^{1.01} + e^{-2}} + \sigma\left(e^{1.01} + e^{-2}\right) = 18{,}778$$

$$\frac{\partial f_1}{\partial w_2} = \frac{f_1(w_1, w_2 + \Delta w) - f_1(w_1, w_2)}{\Delta w} = \frac{18.344 - 18{,}296}{0.01} = 4.764$$

$$f_1(1, -1 + 0.01) = f_1(1, -0.99) = 18{,}344$$

$$\frac{\partial f_2}{\partial w_1} = \frac{f_2(w_1+\Delta w, w_2) - \cancel{f_2(w_1, w_2)}^{\,0}}{\Delta w} = 0$$

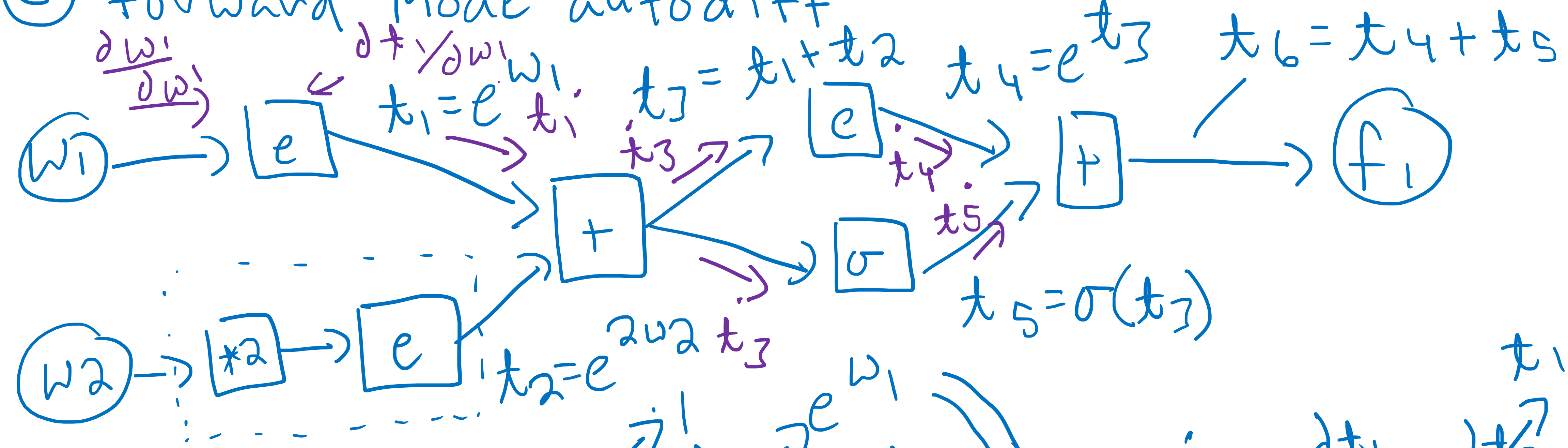
$$f_2(1.01, 1) = (1.01)(-1) + \max(1.01, -1) = 0$$

$$\frac{\partial f_2}{\partial w_2} = \frac{f_2(w_1, w_2+\Delta w) - \cancel{f_2(w_1, w_2)}^{\,0}}{\Delta w} = \frac{0.01}{0.01} = 1$$

$$f_2(1, -0.99) = (1)(-0.99) + \max(1, -0.99) = 0.01$$

$$\frac{\partial \vec{f}}{\partial \vec{w}} = \begin{bmatrix} 48.192 & 4.764 \\ 0 & 1 \end{bmatrix}$$

(c) forward-mode autodiff



$t_4 = e^{t_3}$

$t_1 e^{t_3}$

$$\dot{w}_1 = \frac{\partial w_1}{\partial w_1} = 1$$

$$\dot{t}_1 = \overset{1}{\cancel{\frac{\partial w_1}{\partial w_1}}}\ \overset{e^{w_1}}{\cancel{\frac{\partial t_1}{\partial w_1}}} = \frac{\partial t_1}{\partial w_1} = t_1$$

$$\dot{t}_4 = \frac{\partial t_4}{\partial w_1} = \cancel{\frac{\partial t_3}{\partial w_1}}\ \cancel{\frac{\partial t_4}{\partial t_3}}$$

$$= t_1 t_4$$

$$\dot{w}_2 = \frac{\partial w_2}{\partial w_1} = 0$$

$$\dot{t}_2 = \overset{0}{\cancel{\frac{\partial w_2}{\partial w_1}}}\ \overset{2e^{2w_2}}{\cancel{\frac{\partial t_2}{\partial w_2}}} = 0$$

$$\dot{t}_3 = \frac{\partial t_3}{\partial w_1} = \overset{t_1}{\cancel{\frac{\partial t_1}{\partial w_1}}}\ \overset{1}{\cancel{\frac{\partial t_3}{\partial t_1}}} = t_1$$

$$\dot{t}_5 = \frac{\partial t_5}{\partial w_1} = \frac{\cancel{\partial t_3}^{\,t_1}}{\cancel{\partial w_1}} \frac{\partial t_5}{\partial t_3} = t_1\, \sigma(t_3)(1-\sigma(t_3)) = t_5(1-t_5)\, t_1$$

$$t_5 = \sigma(t_3) \Rightarrow \sigma'(x) = \sigma(x)(1-\sigma(x))$$

$$\dot{t}_6 = \frac{\partial t_6}{\partial w_1} = \frac{\partial t_4}{\partial w_1} + \frac{\partial t_5}{\partial w_1}$$

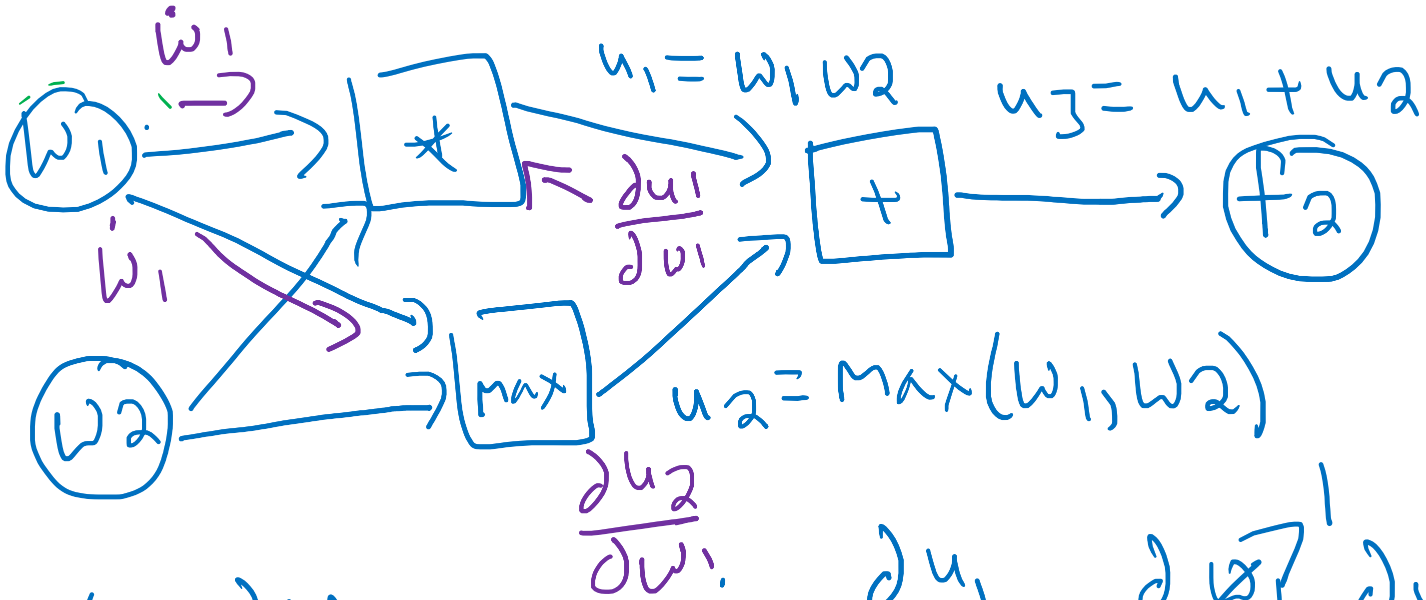
$$= t_1 t_4 + t_5(1-t_5)\, t_1$$

$$= e^{w_1} e^{t_3} + \sigma(t_3)(1-\sigma(t_3))\, e^{w_1}$$

$$= e^{w_1} \cdot e^{e^{w_1}+e^{2w_2}} + \sigma(e^{w_1}+e^{2w_2})(1-\sigma(\,)\,)\, e^{w_1}$$

$$= e^{e+e^{-2}+1} + \sigma(e+e^{-2})(1-\sigma(e+e^{-2})) \cdot e$$

$$= 47.303$$



$$\dot w_1 = \frac{\partial w_1}{\partial w_1} = 1$$

$$\dot u_1 = \frac{\partial u_1}{\partial w_1} = \overset{1}{\cancel{\frac{\partial w_1}{\partial w_1}}}\ \overset{w_2}{\cancel{\frac{\partial u_1}{\partial w_1}}} = w_2$$

$$\dot w_2 = \frac{\partial w_2}{\partial w_1} = 0$$

$$\dot u_2 = \frac{\partial u_2}{\partial w_1} = \overset{1}{\cancel{\frac{\partial w_1}{\partial w}}}\ \frac{\partial u_2}{\partial w_1}$$

$$= \frac{\partial}{\partial w_1} \max(\overset{1}{\cancel{w_1}}, \overset{-1}{\cancel{w_2}})$$

$$= \frac{\partial}{\partial w_1} w_1 = 1$$

The derivative of max is dep. on the input.

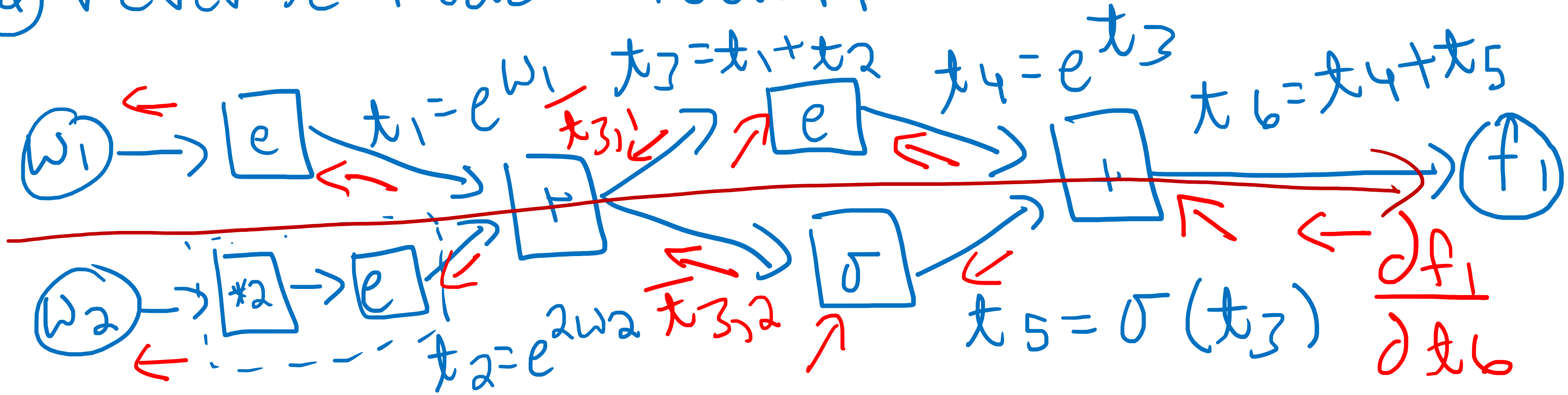
$$\dot u_3 = \frac{\partial u_3}{\partial w_1} = \frac{\partial u_1}{\partial w_1} + \frac{\partial u_2}{\partial w_2}$$

$$= w_2 + 1$$

$$= -1 + 1$$

$$= 0$$

② reverse-mode autodiff



$$\bar{t}_6 = \frac{\partial f_1}{\partial t_6} = 1$$

$$\bar{t}_5 = \frac{\partial f}{\partial t_5} = \cancelto{1}{\frac{\partial f_1}{\partial t_6}}\ \cancelto{1}{\frac{\partial t_6}{\partial t_5}} = 1$$

$$\bar{t}_4 = \frac{\partial f_1}{\partial t_4} = \cancelto{1}{\frac{\partial f_1}{\partial t_6}}\ \cancelto{1}{\frac{\partial t_6}{\partial t_4}} = 1$$

$$\bar{t}_3 = \frac{\partial f_1}{\partial t_3} = \bar{t}_{3,1} + \bar{t}_{3,2}$$

$$\bar{t}_{3,1} = \frac{\partial f_1}{\partial t_{3,1}} = \cancelto{1}{\frac{\partial f_1}{\partial t_4}}\ \cancelto{e^{t_3}}{\frac{\partial t_4}{\partial t_3}} = t_4$$

$$\bar{t}_{3,2} = \frac{\partial f_1}{\partial t_{3,2}} = \frac{\partial f_1}{\partial t_5}\ \frac{\partial t_5}{\partial t_3}$$

$$= t_5(1-t_5)$$

$$\bar{t}_2 = \frac{\partial f_1}{\partial t_2} = \frac{\partial f_1}{\partial t_3} \cancel{\frac{\partial t_3}{\partial t_2}}^{\,1} = \bar{t}_3$$

$$t_3 = t_1 + t_2$$

$$\bar{t}_1 = \frac{\partial f_1}{\partial t_1} = \frac{\partial f_1}{\partial t_3} \cancel{\frac{\partial t_3}{\partial t_1}}^{\,1} = \bar{t}_3$$

$$\bar{w}_2 = \frac{\partial f_1}{\partial w_1} = \frac{\partial f_2}{\partial t_2} \cancel{\frac{\partial t_2}{\partial w_2}}^{\,2e^{2w_2}} = \bar{t}_3 \cdot 2t_2$$

$$t_2 = e^{2w_2} \qquad t_1 = e^{w_1}$$

$$\bar{w}_1 = \frac{\partial f_1}{\partial w_1} = \frac{\partial f_1}{\partial t_1} \cancel{\frac{\partial t_1}{\partial w_1}}_{\,e^{w_1}} = \bar{t}_3 \cdot t_1$$

$$\bar{t}_3 = t_4 + t_5(1-t_5) = e^{t_3} + \sigma(t_3)(1-\sigma(t_3))$$

$$= e^{t_1+t_2} + \sigma(t_1+t_2)(1-\sigma(\,))$$

$$= e^{e^{w_1}+e^{2w_2}} + \sigma(e^{w_1}+e^{2w_2})(1-\sigma(\,))$$

$$= e^{e+e^{-2}} + \sigma(e+e^{-2})(1-\sigma(e+e^{-2}))$$

$$= 17.4$$

$$\bar{w}_2 = \bar{t}_3 \cdot 2t_2 = 17.4 \cdot 2 \cdot e^{-2} = 4.71 = \frac{\partial f_1}{\partial w_2}$$

$$\bar{w}_1 = \bar{t}_3 \cdot t_1 = 17.4 \cdot e = 47.3 = \frac{\partial f_1}{\partial w_1}$$

AI tips

1. Draw out the CG
   - Make it similar to CG in slide 8
   - Break out operations into their own nodes.

2. Keep track of the shapes as you go back in the graph.
   - If no shape change, do an elementwise prod.
   - If there is a shape change, do a dot product,
   - Transpose when necessary.