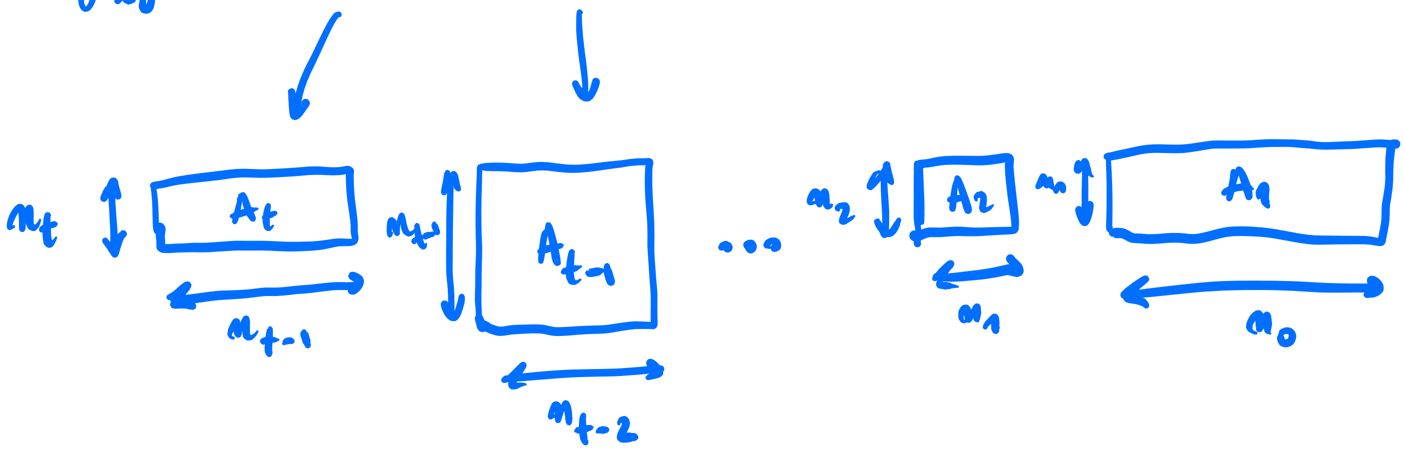
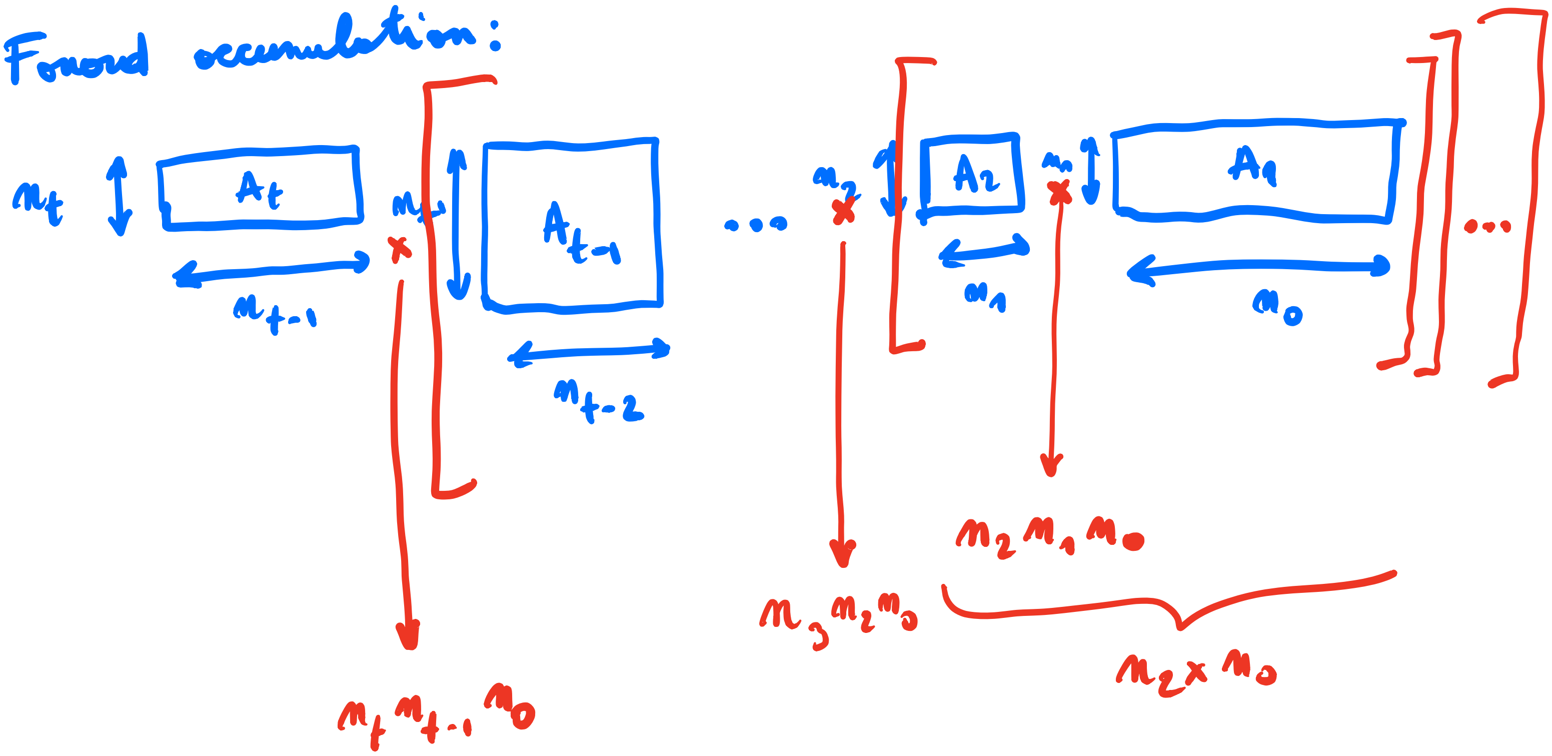
Lecture 3

Complexity (Slide 17)

$$\frac{\partial x_t}{\partial x_0} = \frac{\partial x_t}{\partial x_{t-1}} \cdot \frac{\partial x_{t-1}}{\partial x_{t-2}} \cdots \frac{\partial x_2}{\partial x_1} \cdot \frac{\partial x_1}{\partial x_0}$$

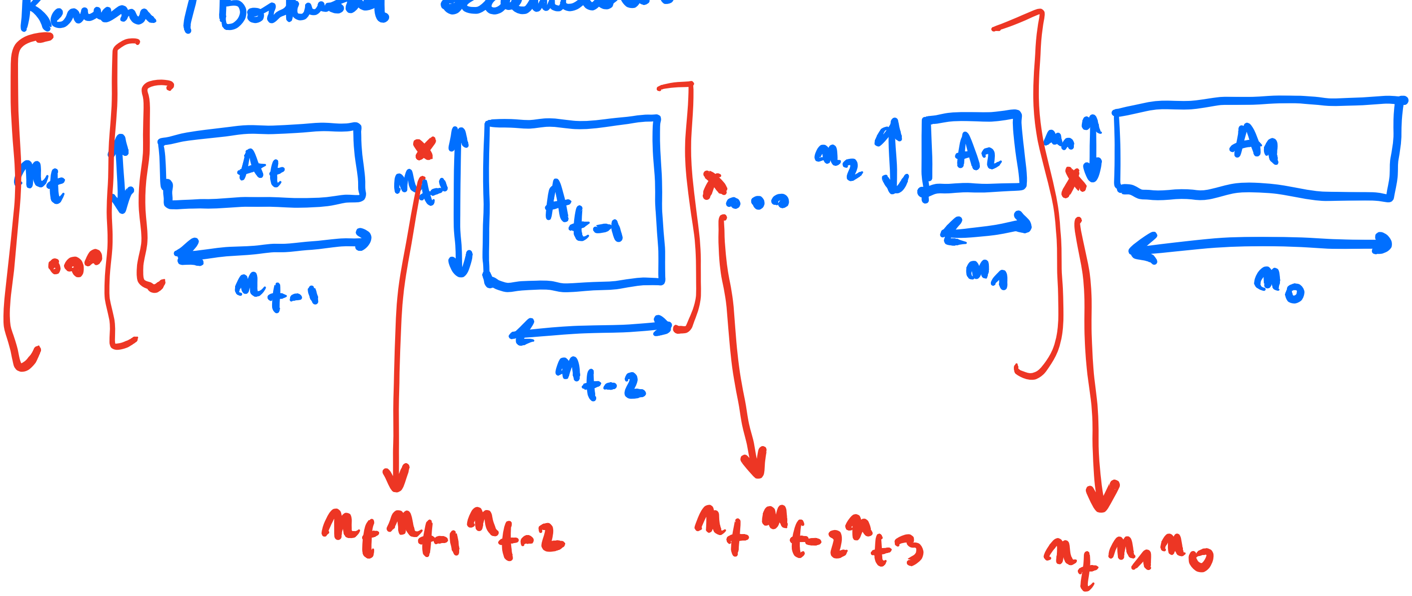


Forward accumulation:



$$\Rightarrow \; n_0 \sum_{k=1}^{t-1} n_k n_{k+1} = \left(n_0 \sum_{k=1}^{t-2} n_k n_{k+1} \right) + n_0 n_{t-1} n_t$$

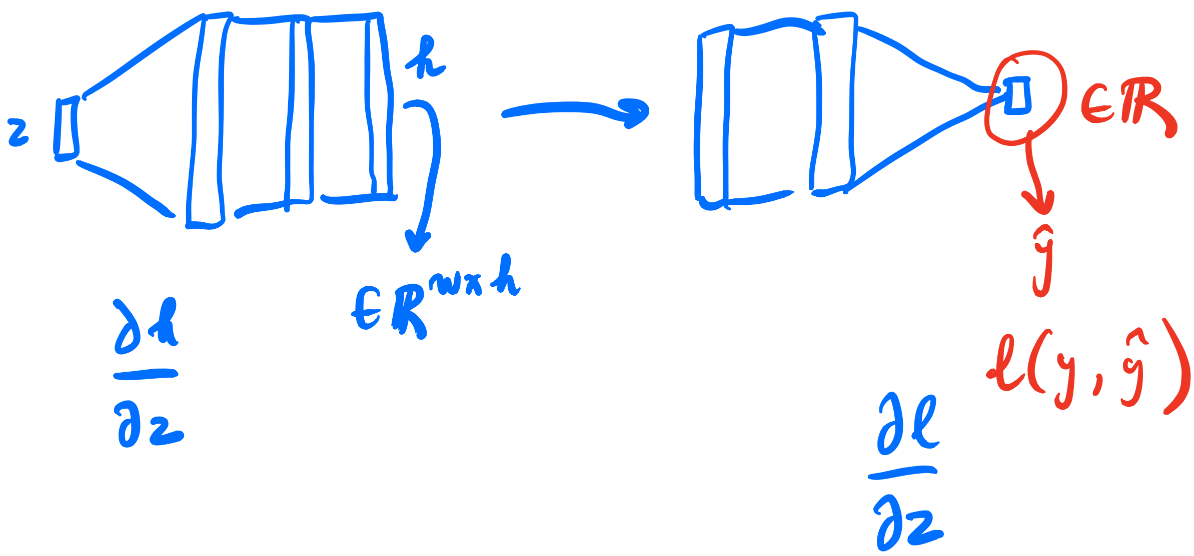
Reverse / Backward accumulation



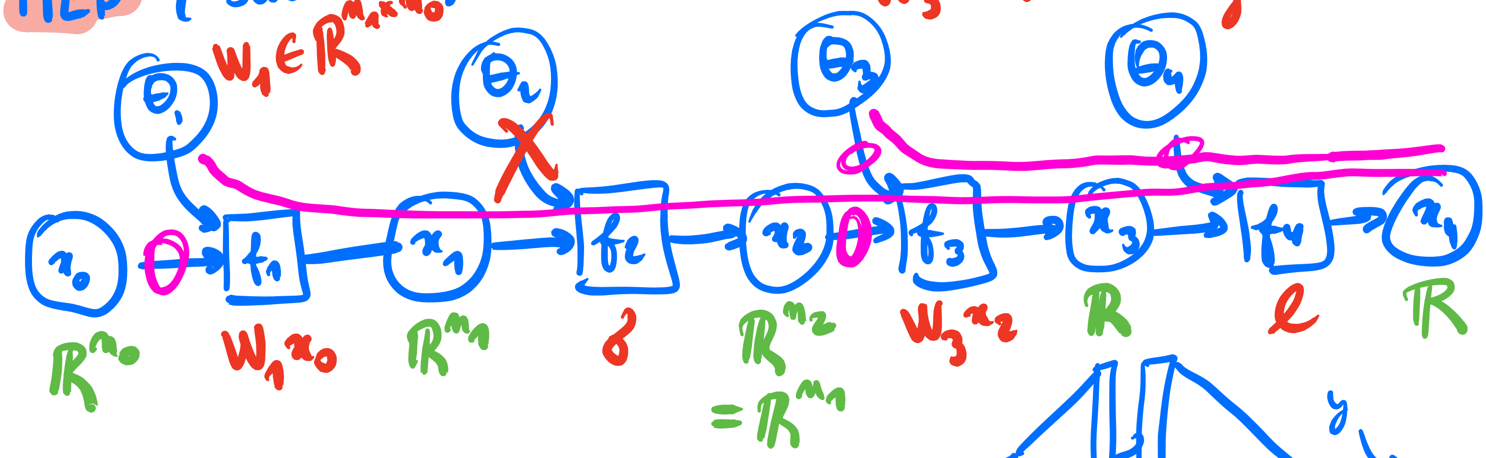
$n_t \times n_{t-2}$

$$\Rightarrow n_t \sum_{k=0}^{t-2} n_k n_{k+1} = \left(n_t \sum_{k=1}^{t-2} n_k n_{k+1} \right) + n_t n_0 n_1$$

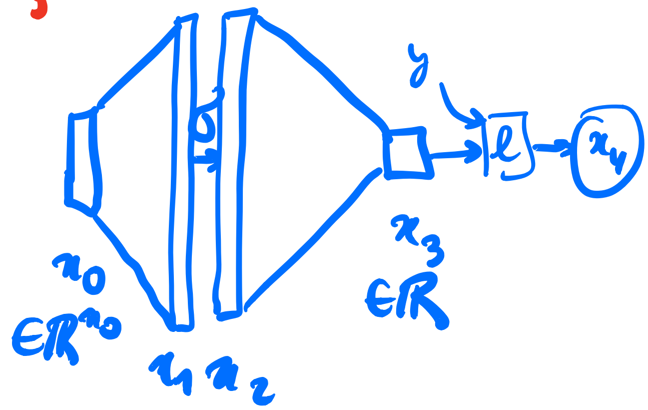
G



MLP (Slide 19)



$$x_0 = x \in \mathbb{R}^{n_0}$$
$$x_1 = f_1(x_0, W_1) = W_1 x_0$$
$$x_2 = f_2(x_1, \emptyset) = \sigma(x_1)$$
$$x_3 = f_3(x_2, W_3) = W_3 x_2 = \hat{y}$$
$$x_4 = f_4(x_3, y) = \frac{1}{2}(x_3 - y)^2$$



We want $\dfrac{\partial x_4}{\partial \theta_3}$ and $\dfrac{\partial x_4}{\partial \theta_1}$

$$\frac{\partial x_4}{\partial \theta_3} = \frac{\partial x_4}{\partial x_3} \cdot \frac{\partial x_3}{\partial \theta_3}$$

$$\theta_3 := W_3 \in \mathbb{R}^{1 \times n_2}$$
$$|\theta_3| = n_2$$

$$= \overset{1\times 1}{\Big[(x_3 - y) \Big]} \overset{1 \times |\theta_3|}{\Big[\quad j \quad \Big]}$$

$$x_3 = W_2 x_2 = \boxed{W_2} \begin{bmatrix} x_2 \end{bmatrix} = \sum_j W_3^{(j)} x_2^{(j)}$$

$$\frac{\partial x_3}{\partial W_3^{(j)}} = x_2^{(j)}$$

$$= \left[x_2^{(1)} \dots x_2^{(n_2)} \right]$$

$$\frac{\partial x_4}{\partial \theta_1} = \frac{\partial x_4}{\partial x_3} \frac{\partial x_3}{\partial x_2} \frac{\partial x_2}{\partial x_1} \frac{\partial x_1}{\partial \theta_1}$$

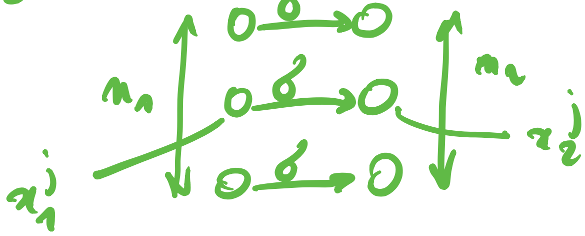
$\theta_1 := W_1 \in \mathbb{R}^{m_1 \times m_0}$

$|\theta_1| = m_1 m_0$

$$= \underset{1\times 1}{\left[(x_3 - y)\right]} \underset{1\times m_2}{\left[\ \cdot\ \right]} \underset{m_2 \times m_1}{\left[\ \mathbb{1}\ \right]} \underset{m_1 \times |\theta_1|}{\left[\ \boxed{0}\ \boxed{0}\ \cdots\ \boxed{0}\ \boxed{0}\ \right]}\ i$$

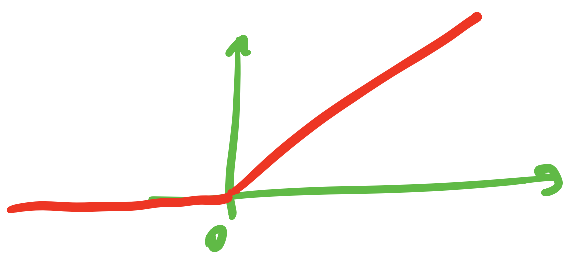
x_0

$$\frac{\partial x_3}{\partial x_2^{(j)}} = W_3^{(j)}$$

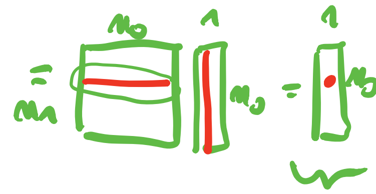


$\sigma := \mathrm{ReLU}$

$x_1 > 0$



$x_1 = W_1 \; x_0$



$$x_1^{(i)} = \sum_j W_1^{i,j} x_0^{(j)}$$

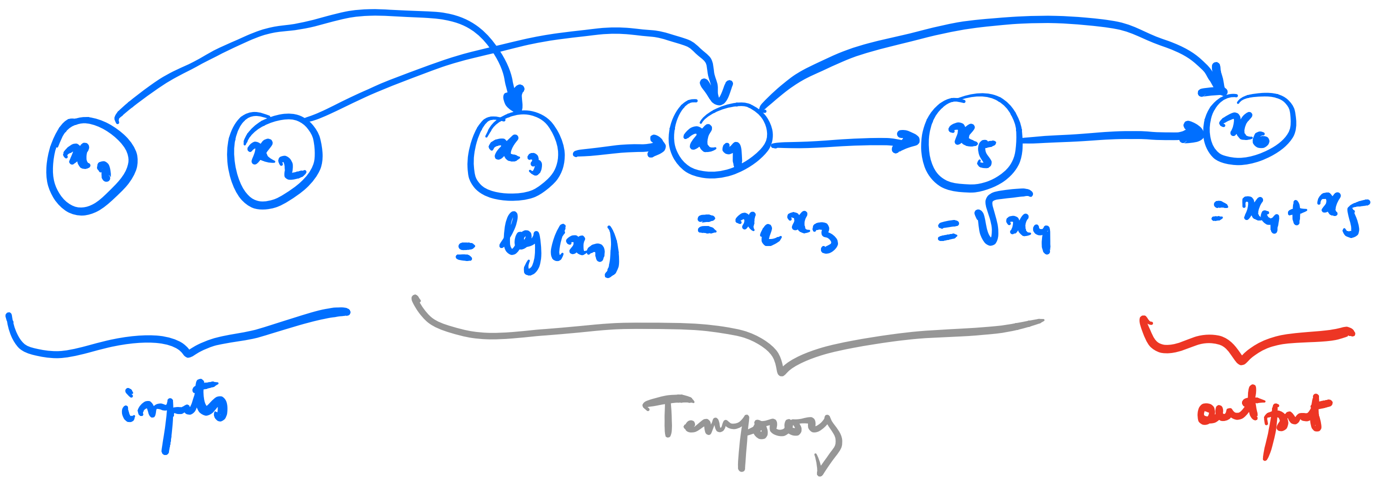
$$\frac{\partial x_1^i}{\partial W_1^{ij}}$$

$$\begin{bmatrix} x_0 & 0 & 0 \\ 0 & x_0 & 0 \\ 0 & 0 & x_0 \end{bmatrix}$$

$[\ x_0^T \otimes \mathbb{1}$

Forward mode AD (Slide 25)

$$f(x_1, x_2) = x_2 \log(x_1) + \sqrt{x_2 \log(x_1)}$$



$\frac{\partial x_6}{\partial x_1}$:

$$\boxed{\frac{\partial x_k}{\partial x_1} = \sum_{l \in parents(k)} \left[\frac{\partial x_k}{\partial x_l}\right] \frac{\partial x_l}{\partial x_1}}$$

$$\frac{\partial x_1}{\partial x_1} = 1$$

$$\frac{\partial x_2}{\partial x_1} = 0$$

$$\frac{\partial x_3}{\partial x_1} = \left[\frac{\partial x_3}{\partial x_1}\right] \frac{\partial x_1}{\partial x_1} = \frac{1}{x_1} \cdot 1$$

$$\frac{\partial x_4}{\partial x_1} = \left[\frac{\partial x_4}{\partial x_3}\right] \frac{\partial x_3}{\partial x_1} + \left[\frac{\partial x_4}{\partial x_2}\right] \frac{\partial x_2}{\partial x_1}$$

$$= x_2 \frac{\partial x_3}{} + 0$$

$$\partial x_1$$

$$\frac{\partial x_5}{\partial x_1} = \left[\frac{\partial x_5}{\partial x_4}\right] \frac{\partial x_4}{\partial_1}$$

$$= \frac{1}{2} \frac{1}{\sqrt{x_4}} \cdot \frac{\partial x_4}{\partial x_1}$$

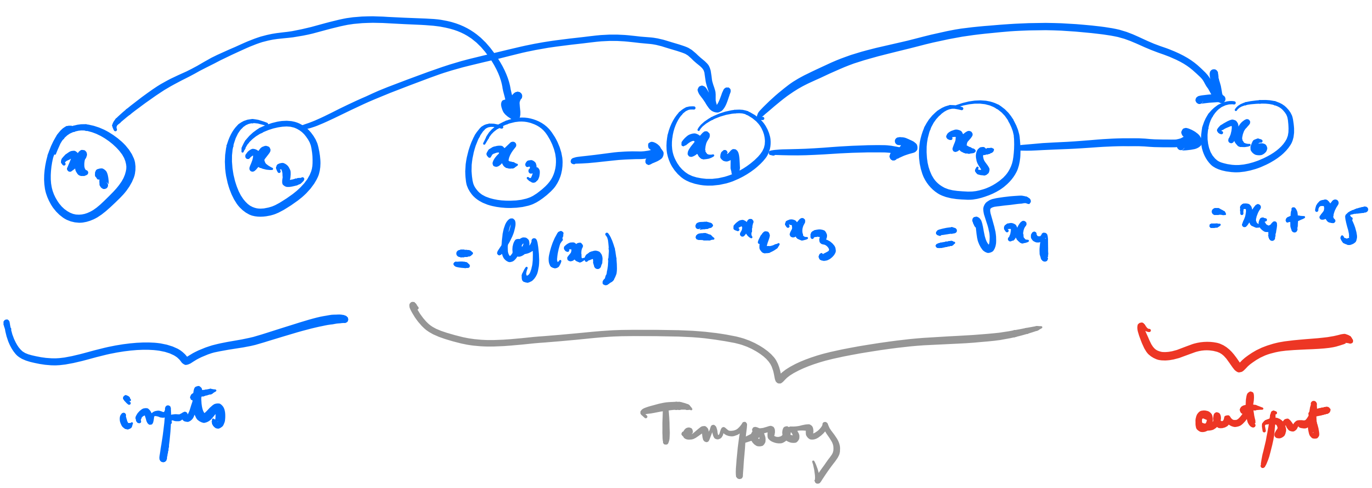
$$\frac{\partial x_6}{\partial x_1} = \left[\frac{\partial x_6}{\partial x_4}\right] \frac{\partial x_4}{\partial x_1} + \left[\frac{\partial x_6}{\partial x_5}\right] \frac{\partial x_5}{\partial x_1}$$

$$= 1 . \frac{\partial x_4}{\partial x_1} + 1 \frac{\partial x_5}{\partial x_1}$$

$$\frac{\partial x_6}{\partial x_2} : \quad (\ldots)$$

Backward mode of AD (Slide 29)

$$f(x_1, x_2) = x_2 \log(x_1) + \sqrt{x_2 \log(x_1)}$$



$$\frac{\partial x_6}{\partial x_6} = 1$$

$$\boxed{\frac{\partial x_t}{\partial x_k} = \sum_{m \in \text{child}(k)} \frac{\partial x_t}{\partial x_m} \left[\frac{\partial x_m}{\partial x_k}\right]}$$

$$\frac{\partial x_6}{\partial x_5} = \frac{\partial x_6}{\partial x_6}\left[\frac{\partial x_6}{\partial x_5}\right]$$

$$= \frac{\partial x_6}{\partial x_6} \cdot 1$$

$$\frac{\partial x_6}{\partial x_4} = \frac{\partial x_6}{\partial x_5} \cdot \left[\frac{\partial x_5}{\partial x_4}\right] + \frac{\partial x_6}{\partial x_6}\left[\frac{\partial x_6}{\partial x_4}\right]$$

$$= \frac{\partial x_6}{\partial x_5} \frac{1}{2} \frac{1}{\sqrt{x_4}} + \frac{\partial x_6}{\partial x_6} 1$$

$$\frac{\partial x_6}{\partial x_3} = \frac{\partial x_6}{\partial x_4}\left[\frac{\partial x_4}{\partial x_3}\right]$$

$$= \frac{\partial x_6}{\partial x_4} \cdot x_2$$

$$\frac{\partial x_6}{\partial x_2} = \frac{\partial x_6}{\partial x_4}\left[\frac{\partial x_4}{\partial x_2}\right]$$

$$= \frac{\partial x_6}{\partial x_4} x_3$$

$$\frac{\partial x_6}{\partial x_1} = \frac{\partial x_6}{\partial x_1}\left[\frac{\partial x_3}{\partial x_1}\right] = \frac{\partial x_6}{\partial x_3} \frac{1}{x_1}$$

We got both at once!