X

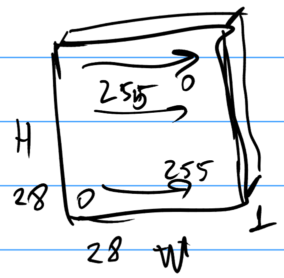
Y

$\updownarrow$ transform(X)

MODEL, θ

t(X)

$\hat{Y}$

$X =$ 

$\downarrow$ tranf $\frac{\cdot}{255}$

1 0
0 1

$\mathcal{L}_\theta(Y, \hat{Y}) \to \min_\theta$

$\frac{1}{10}$

[0, 1]

$X \cdot W + b = \hat{Y} = 10 \begin{bmatrix} \vdots \end{bmatrix} \in \mathbb{R}^{10} \to [0,1]^{10}$

$H \times W$

$X \to \underbrace{[\ \to \mid \to \mid \to \mid \ \mid \to\ ]}_{28^2}$

$H \times W$

$10 \quad \begin{bmatrix} W \end{bmatrix} \begin{bmatrix} X \end{bmatrix}$

$\theta^{n+1} := \theta^{n} - \varepsilon \nabla_\theta \mathcal{L}$

i: $\left(-\sum_{c=1}^{10} Y_c \log \mathrm{sm}(\hat{Y}_c) \right)$

$x, y \in \mathbb{R}$

def $f(x,y) := \left(2xy - \sin(x+y)\right)^{e^{x+\cos y \cdot \log\frac{1}{z}}} \in \mathbb{R}$

y^2

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

$$\frac{df}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \approx \frac{\Delta f}{\Delta x}$$

$$\frac{\partial f}{\partial x} = 2y - \cos(x+y) \cdot 1$$

$$\frac{\partial f}{\partial x}\Big|_{(x_0,y_0)} \approx \frac{f(x_0+h, y_0) - f(x_0, y_0)}{h} \qquad h = 10^{-6}$$

$$\frac{\partial f}{\partial y}\Big|_{(x_0,y_0)} \approx \frac{f(x_0, y_0+h) - f(x_0, y_0)}{h} \qquad (x_0, y_0)$$

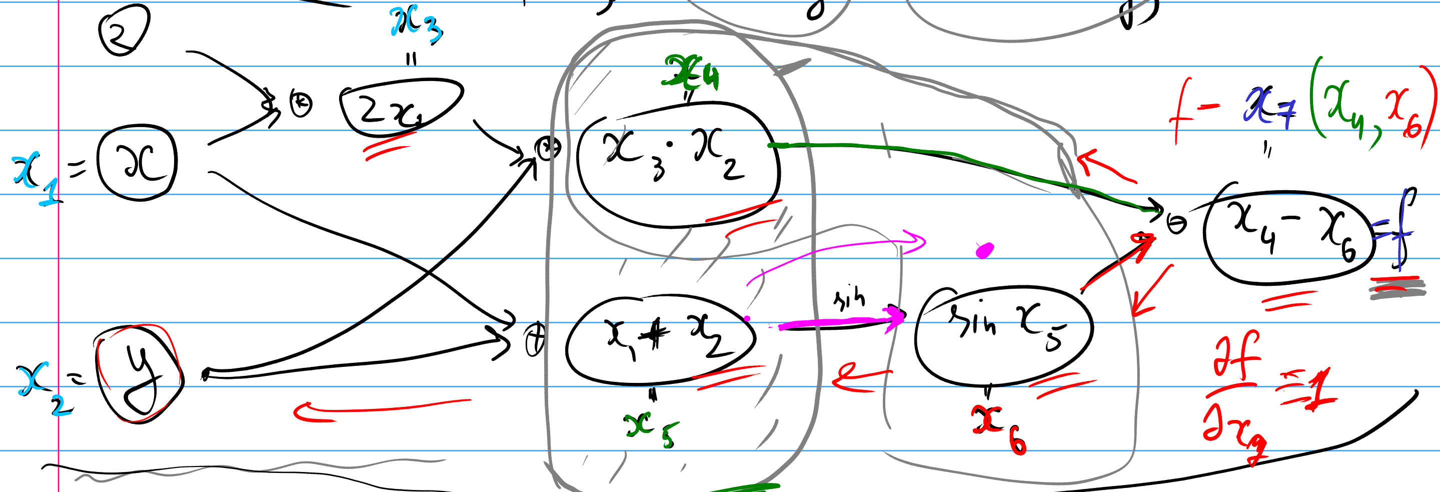
$z - w = f(z, w)$ $z = 2 \cdot xy$

$\frac{\partial f}{\partial z} = 1$, $\frac{\partial f}{\partial w} = -1$ $w = \sin(x+y)$

Auto diff

$$f(x,y) = 2x \cdot y - \sin(x+y)$$

(2)



$$f(x,y) = f(x_1, x_2) = f(x_4, x_5) = f(x_4(x_2, x_3), x_5(x_1, x_2))$$

$$= f(x_4, x_6) = f(x_7) = x_7$$

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}$$

$$\underbrace{\frac{\partial f}{\partial x} = \frac{\partial f}{\partial x_1} = \frac{\partial f(x_4, x_5)}{\partial x_1}}= \underbrace{\frac{\partial f}{\partial x_4}}_{=1}\underbrace{\frac{\partial x_4}{\partial x_1}}_{=0} + \frac{\partial f}{\partial x_5}\underbrace{\frac{\partial x_5}{\partial x_1}}_{=1}$$

$$x_4 = x_3 x_2$$

$$\frac{\partial x_4}{\partial x_1} = 0 \qquad \frac{\partial x_5}{\partial x_1} = \frac{\partial(x_1 + x_2)}{\partial x_1} = 1$$

$$1 = \underbrace{\frac{\partial f}{\partial x_4}} = \frac{\partial f(x_7)}{\partial x_4} = \underbrace{\frac{\partial f}{\partial x_7}}_{=1} \cdot \underbrace{\frac{\partial x_7}{\partial x_4}}_{=1} + \underbrace{\frac{\partial f}{\partial x_6} \cdot \cancel{\frac{\partial x_6}{\partial x_4}}^{0}}_{0}$$

$$\frac{\partial x_7}{\partial x_4} = \frac{\partial(x_4 - x_6)}{\partial x_4} = 1 \qquad f = x_7$$

$$\frac{\partial f}{\partial x_7} = 1$$

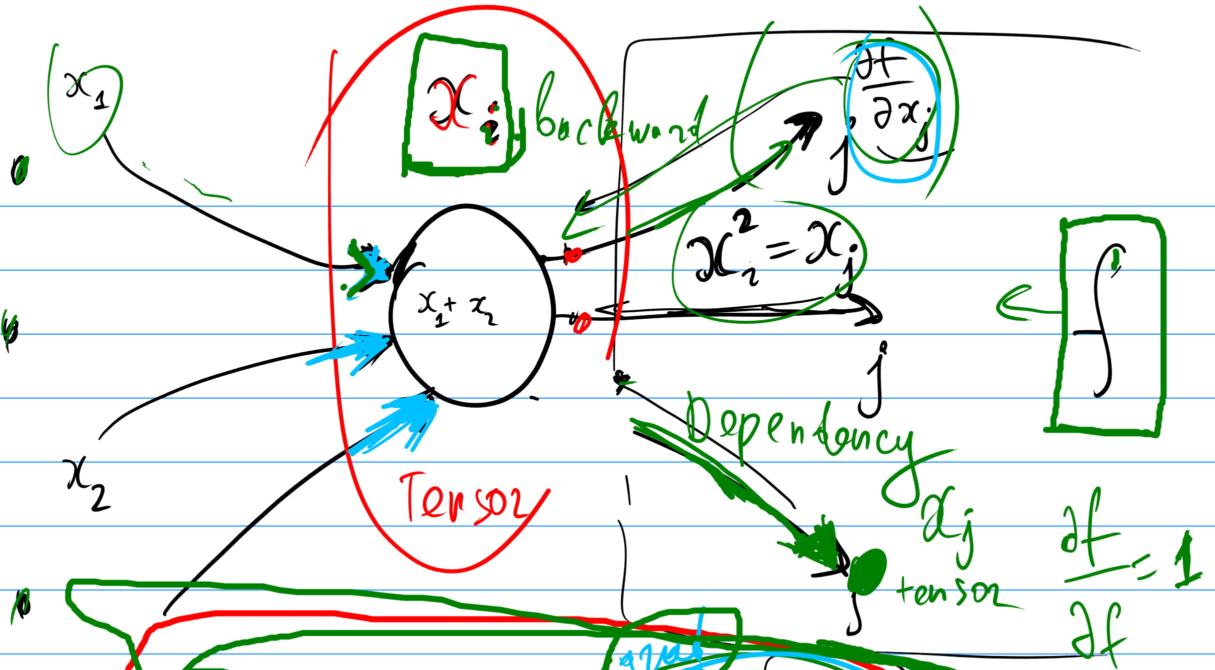
$$\frac{\partial f}{\partial x_5} =$$

$$\boxed{f = z - w}$$

$$\boxed{z = 2xy} \qquad w = \sin(x+y)$$

$$\frac{\partial f}{\partial x} = \underbrace{\frac{\partial f}{\partial z}}_{1} \boxed{\frac{\partial z}{\partial x}} + \underbrace{\frac{\partial f}{\partial w}}_{-1} \underbrace{\frac{\partial w}{\partial x}}$$

$$\frac{\partial z}{\partial x} = 2y \qquad \frac{\partial w}{\partial x} = \cos(x+y)$$



grad

$$\frac{\partial f}{\partial x_i} = \sum_j \frac{\partial f}{\partial x_j} \frac{\partial x_j}{\partial x_i}$$

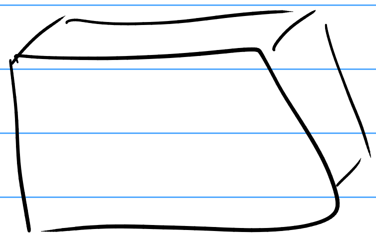
grad_fn $2x_i$

$$\text{grad\_fn}(X) = \frac{\partial x_j}{\partial x_i} X, \quad X = \frac{\partial f}{\partial x_j}$$

$$AX \Rightarrow \frac{\partial (AX)}{\partial X}$$

$$\text{reduse\_sum}(x) \to x_1 + x_2 + \dots + x_N$$

x



$$\frac{\partial (x_1 + \dots + x_n)}{\partial \vec{x}} = \begin{pmatrix} \frac{\partial (x_1 + \dots + x_n)}{\partial x_1} \\ \vdots \\ \frac{\partial (x_1 + \dots + x_n)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$