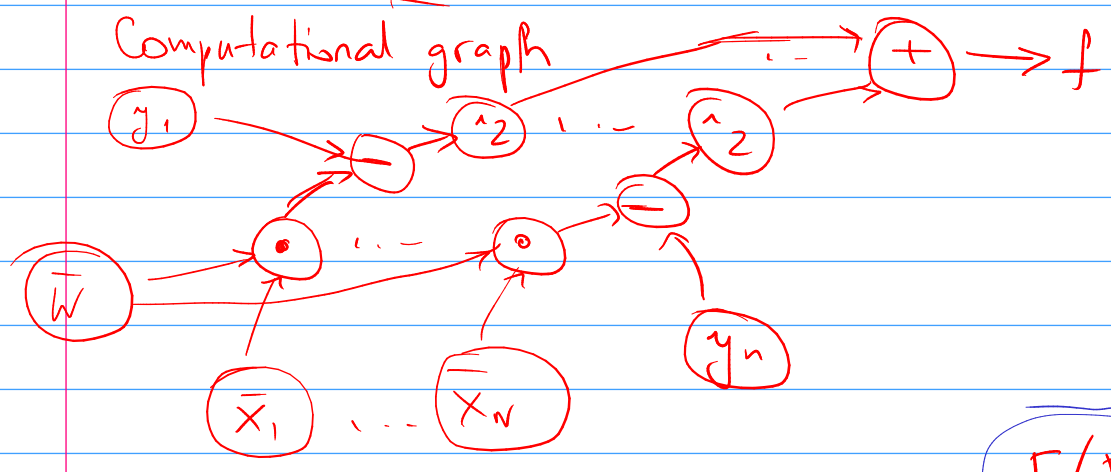


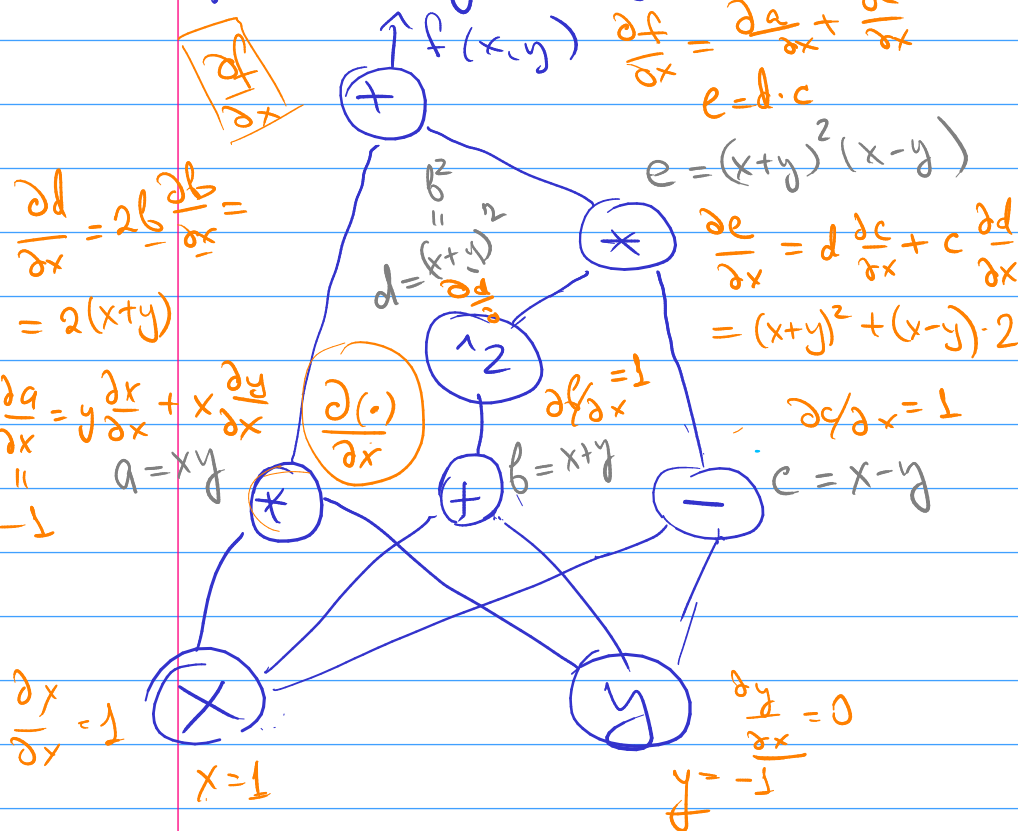
$$L = \frac{1}{N} \sum_{n=1}^N (y_n - \bar{w}^T \bar{x}_n)^2 \rightarrow \min$$

Computational graph



$$F(D, \bar{w}) \xrightarrow{\bar{w}} \min$$

$$f(x, y) = xy + (x+y)^2(x-y)$$



$$\sum_{d \in D} f(d, \bar{w}) \xrightarrow{\bar{w}} \min$$

$$h(\bar{w}^T \bar{x}) = a$$

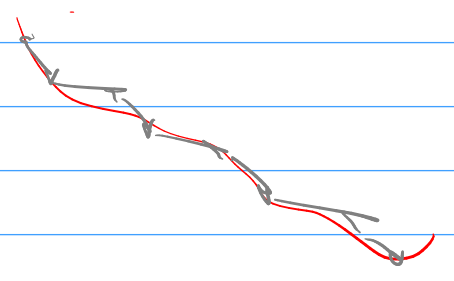
$$\bar{w}^T \bar{x} = h^{-1}(a)$$

$$L(\bar{w}) \xrightarrow{\bar{w}} \min$$

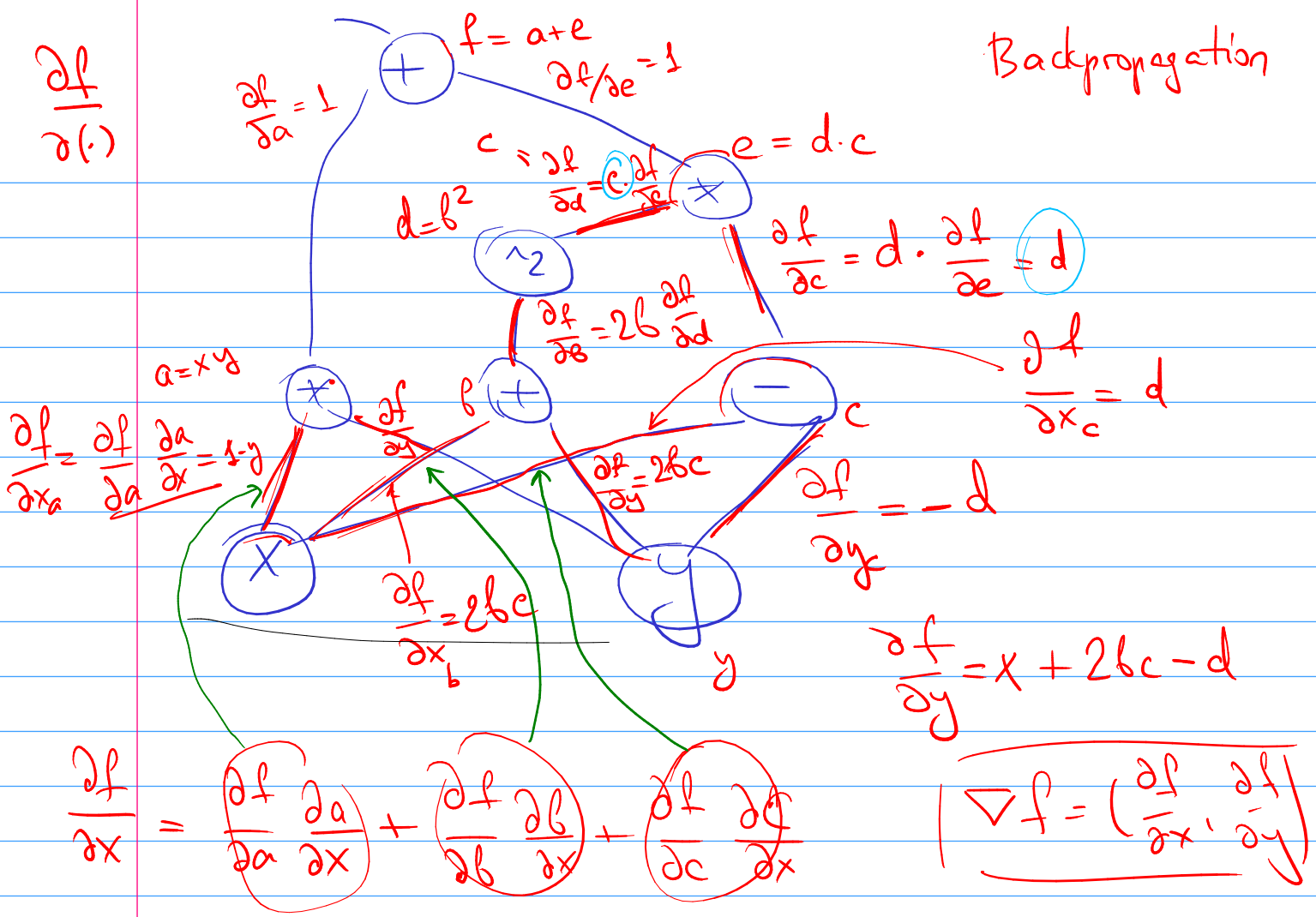
$$\nabla_{\bar{w}} L$$

$$\bar{w}_{t+1} = \bar{w}_t - \alpha \nabla_{\bar{w}} L(\bar{w}_t)$$

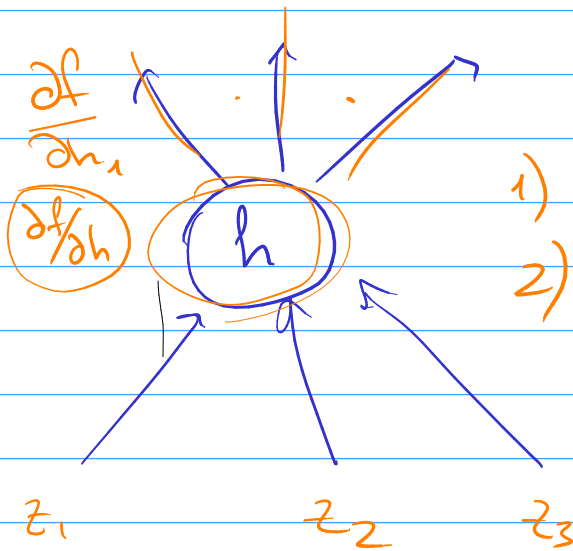
forward propagation, fprop



Backpropagation



$$= y + 2bc + d = y + 2(x+y)(x-y) + (x+y)^2$$



1) $h(z_1, z_2, z_3)$

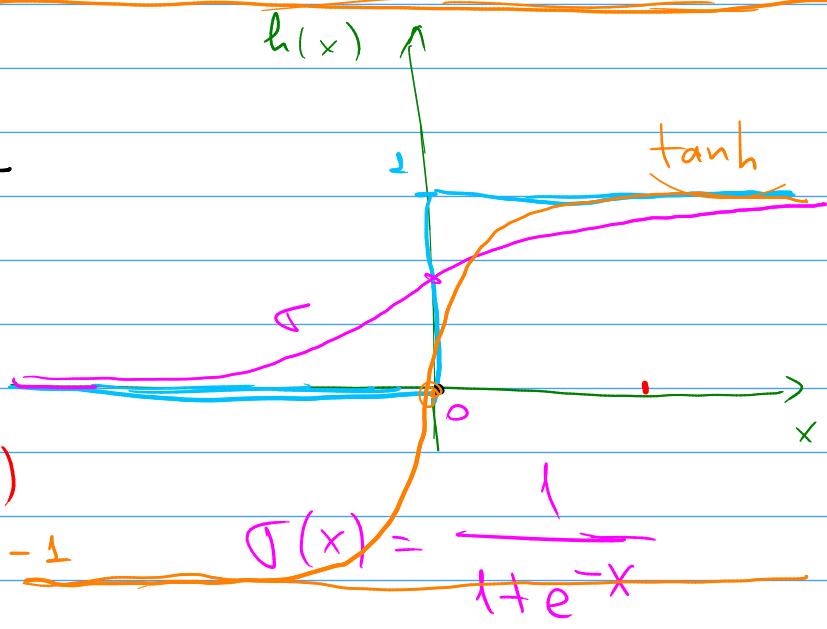
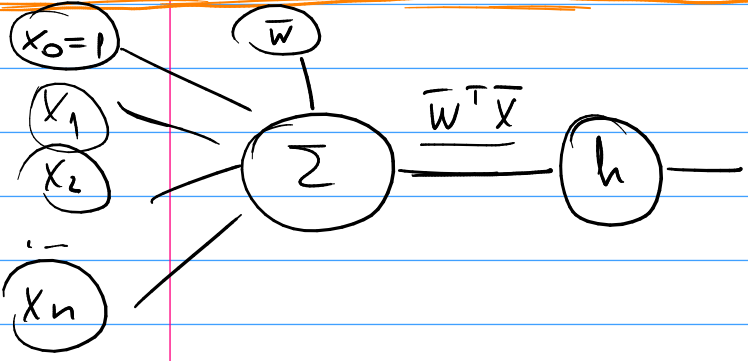
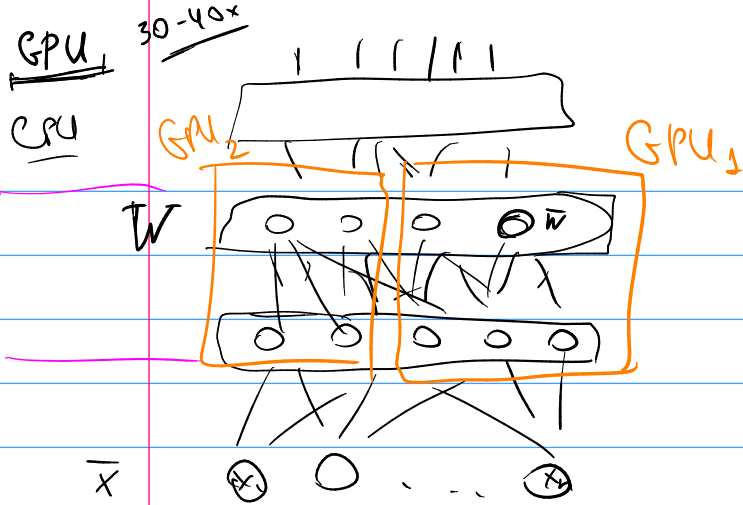
fprop

2) $\frac{\partial f}{\partial z_i} = \frac{\partial f}{\partial h} \cdot \frac{\partial h}{\partial z_i}$

bprop

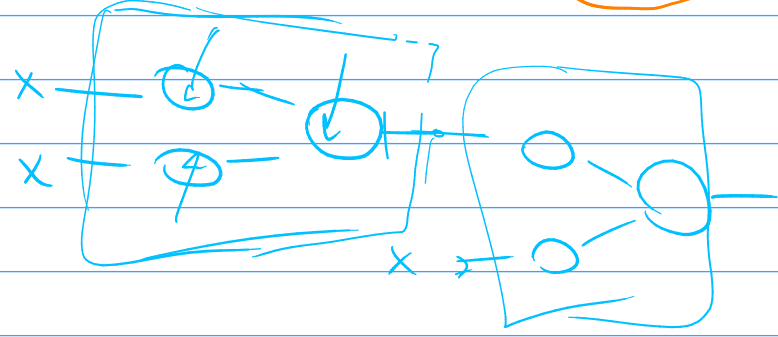
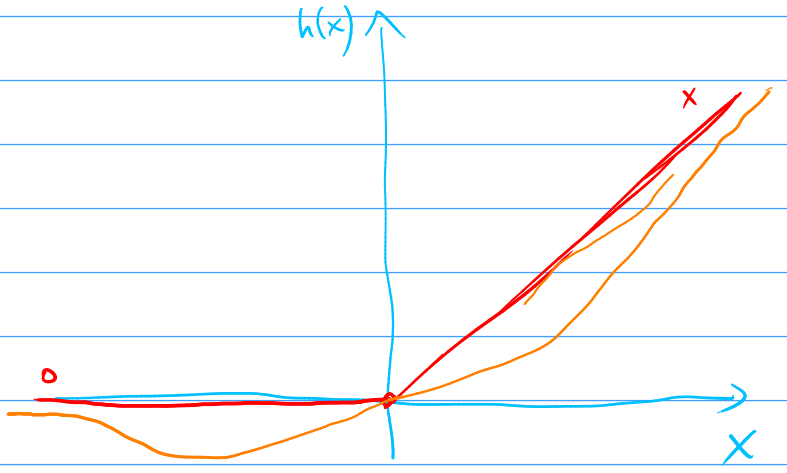
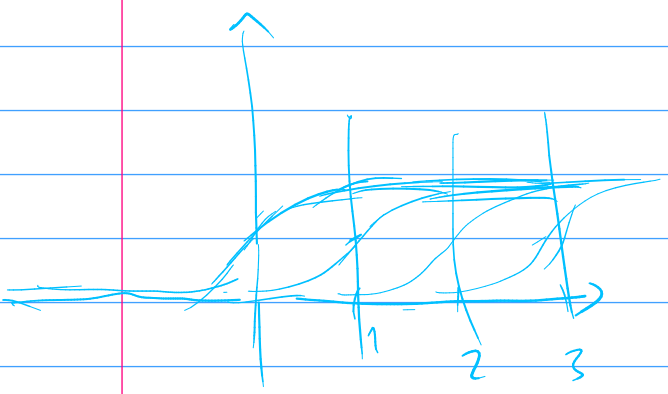
pyTorch

Tensorflow



$ReLU(x) = \max(0, x)$

rectified linear unit



$Swish(x) =$
 $= x \cdot \sigma(\beta \cdot x)$