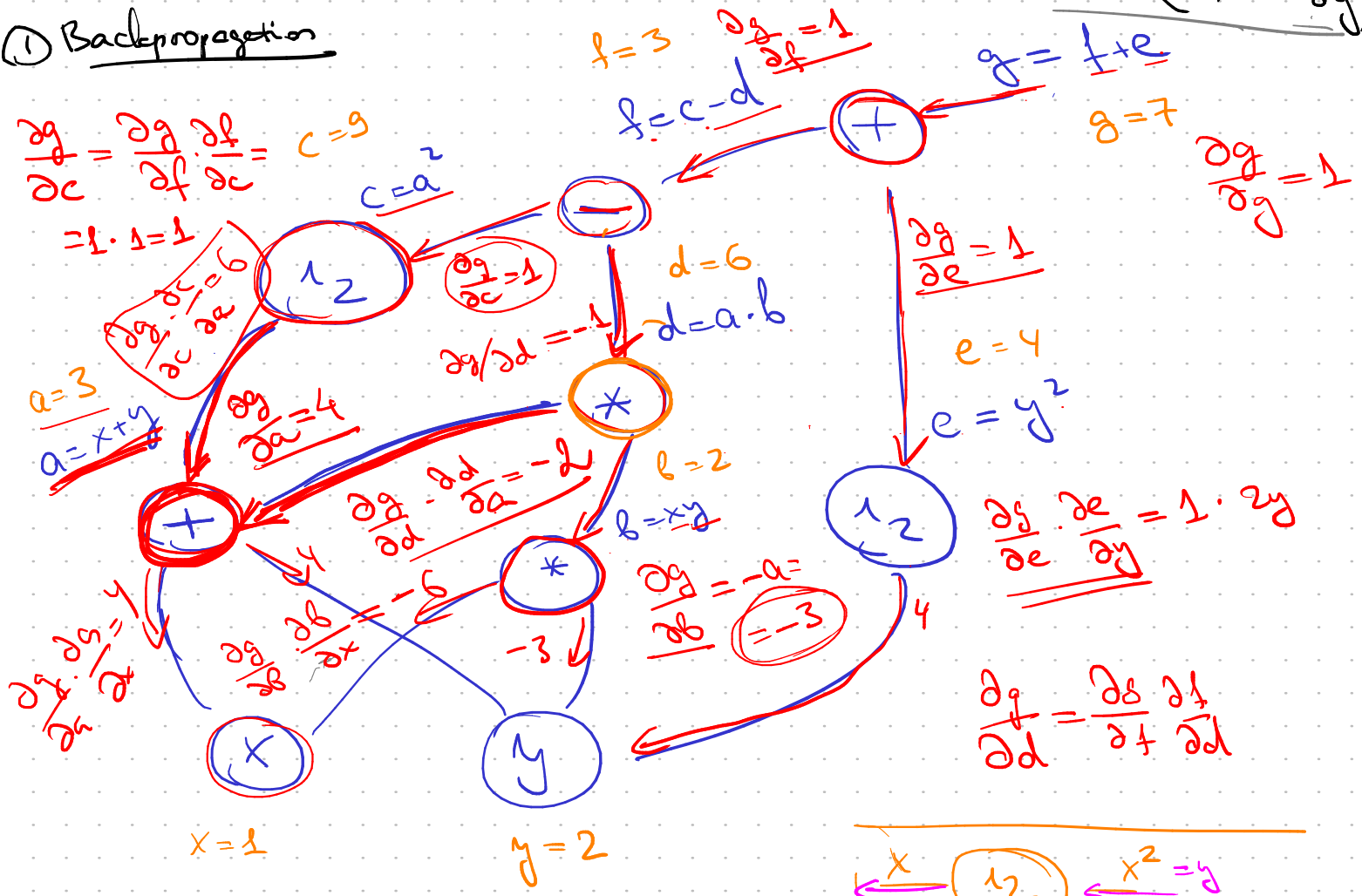
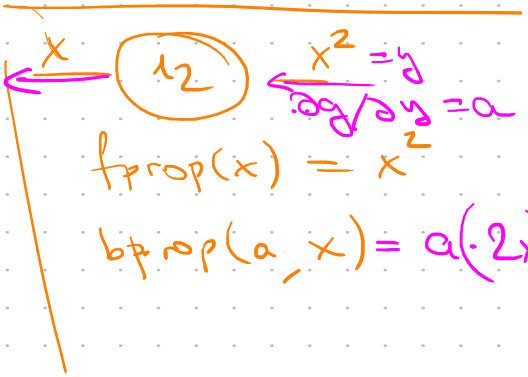
$$g(x,y) = (x+y)^2 - xy(x+y) + y^2 \qquad \nabla g = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right)$$

① Backpropagation





fprop $\dots \to \frac{\partial \dots}{\partial x}$

bprop $\frac{\partial g}{\partial \dots} \to \dots$

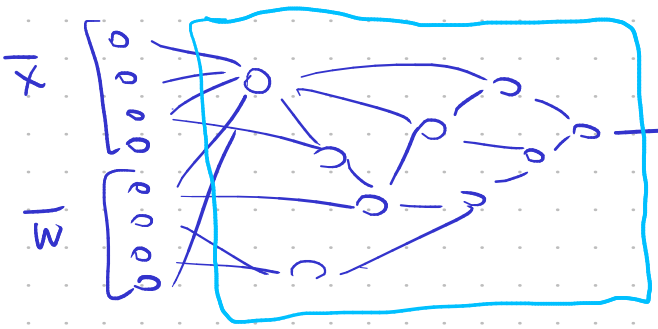
$$\frac{\partial g}{\partial a} = \frac{\partial g}{\partial c}\cdot\frac{\partial c}{\partial a} + \frac{\partial g}{\partial d}\frac{\partial d}{\partial a} = 1\cdot 2a + (-1)\cdot b =$$

$$= 6 - 2 = 4$$

$$\frac{\partial g}{\partial x} = \frac{\partial g}{\partial a}\cdot\frac{\partial a}{\partial x} + \frac{\partial g}{\partial b}\cdot\frac{\partial b}{\partial x} = 4 - 6 = -2$$

$$\frac{\partial g}{\partial y} = \frac{\partial g}{\partial a}\cdot\frac{\partial a}{\partial y} + \frac{\partial g}{\partial b}\cdot\frac{\partial b}{\partial y} + \frac{\partial g}{\partial e}\cdot\frac{\partial e}{\partial y} = 4 - 3 + 4 = 5$$

NN:



GD:

$$\bar{w}^{(k+1)} = \bar{w}^{(k)} - \eta \cdot \nabla_{\bar{w}} L(\bar{w}) \Big|_{\bar{w}^{(k)}}$$

backpropagation

η — learning rate

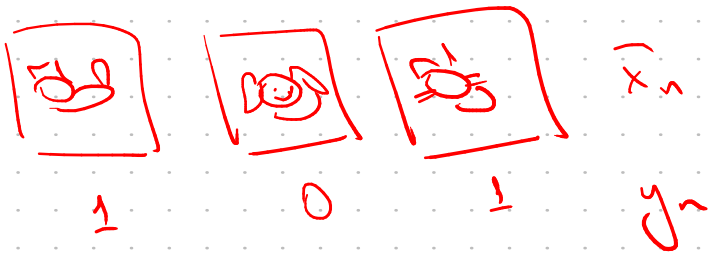
$\bar{w}^{(k)}$

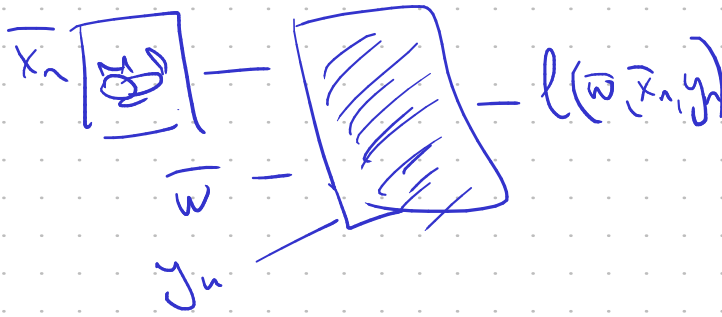
PyTorch: fprop(), bprop()

$\nabla_{\bar{w}} L(\bar{w})$ at $\bar{w}^{(k)}$

② Stochastic gradient descent

$$D = \{(\bar{x}_n, y_n)\}_{n=1}^{N}$$





$$L(\bar{w}) = \frac{1}{N} \sum_{n=1}^{N} \ell(\bar{w}, \bar{x}_n, y_n)$$

$10^6, 10^9$

$$\nabla_{\bar{w}} L(\bar{w}) = \frac{1}{N} \sum_{n=1}^{N} \nabla_{\bar{w}} \ell(\bar{w}, \bar{x}_n, y_n)$$

$\bar{w} \to \min$

Stochastic optimization

$$L(\bar{w}) = \mathbb{E}_{(\bar{x}_n, y_n) \sim \text{Unif}(D)} \left[\ell(\bar{w}, \bar{x}_n, y_n) \right] \approx$$

$4, 16$

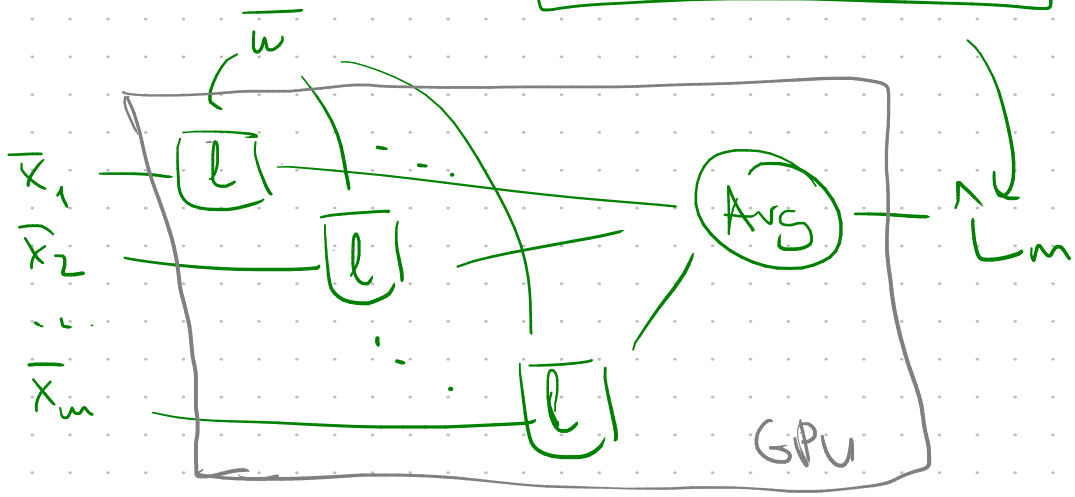
$$\approx \frac{1}{m} \sum_{i=1}^{m} \ell(\bar{w}, \bar{x}_i, y_i)$$

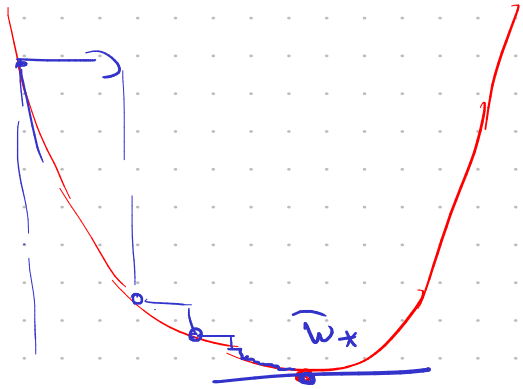
mini-batch

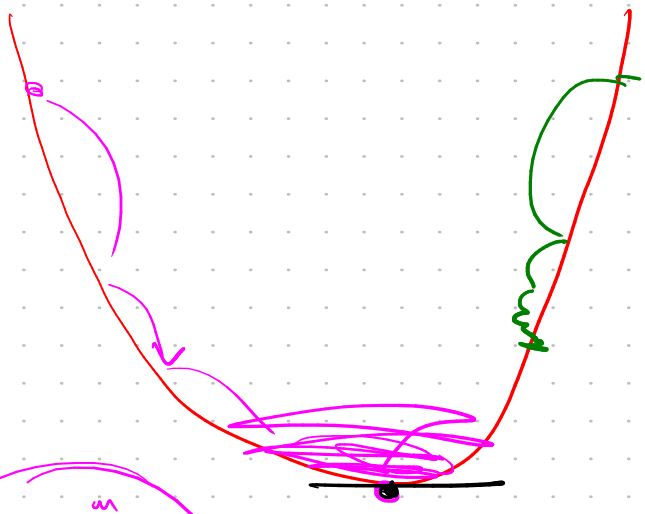
$$\mathbb{E}_{p(x)}[f(x)] \approx \frac{1}{R}\sum f(x_r), \quad \text{где } x_r \sim p(x)$$

stochastic gradient descent: $(\bar{x}_i, y_i) \sim \text{Unif}(D)$

$$\bar{w}^{(k+1)} = \bar{w}^{(k)} - \eta \cdot \frac{1}{m}\sum_{i=1}^{m} \nabla_{\bar{w}} l(\bar{w}, \bar{x}_i, y_i)$$







$\frac{1}{m}\sum_{i}^{m}$

GD: $\bar{w}^{(k+1)} := \bar{w}^{(k)} - \eta \cdot \nabla_{\bar{w}} L\big|_{\bar{w}^{(k)}}$, $\eta = \text{const}$

SGD: $\bar{w}^{(k+1)} := \bar{w}^{(k)} - \eta_k \hat{\bar{g}}^{(k)}$

$\eta_k \xrightarrow{k \to \infty} 0$

"$\frac{1}{m}\sum^{m} \nabla l(\bar{w}, \bar{x}_i, y_i)$"