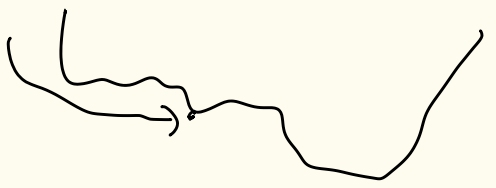
CNN $T = \{(x_1, y_1), \dots\}$ $\underline{W}x + \underline{b} > 0$

$L(f_\theta(x), y^*)$

$\underset{\theta}{\operatorname{argmin}} \sum_i L(f_\theta(x_i), y_i)$



gradient descent

SGD (stochastic g.d.) učenie vždy na dvoch dátach naraz

$$\theta_{t+1} := \theta_t - \lambda \frac{dL}{d\theta}$$

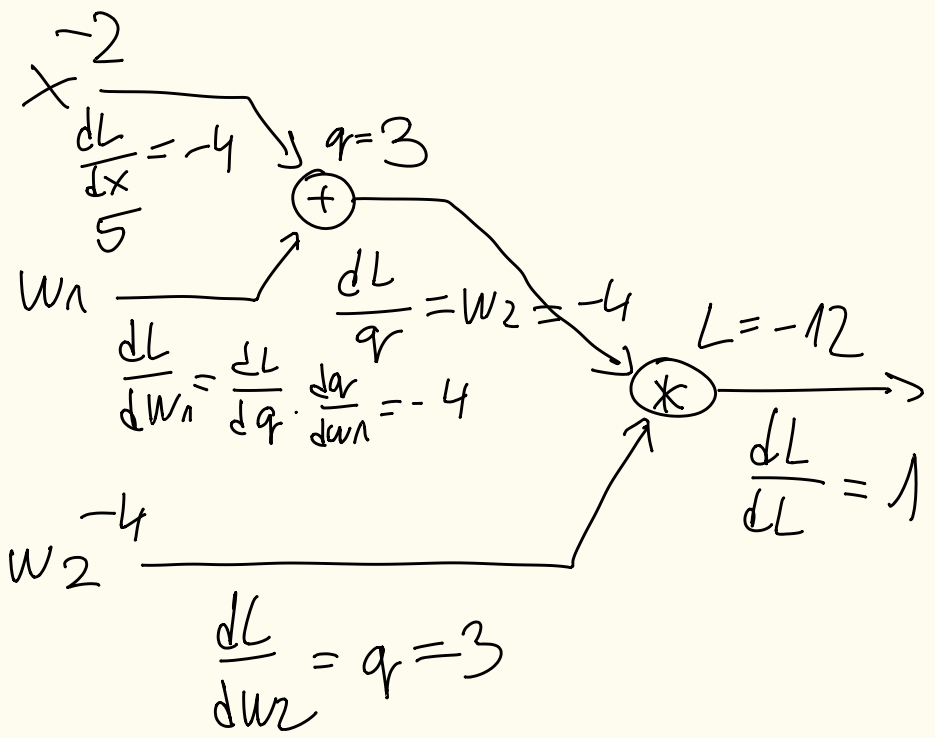
learning rate

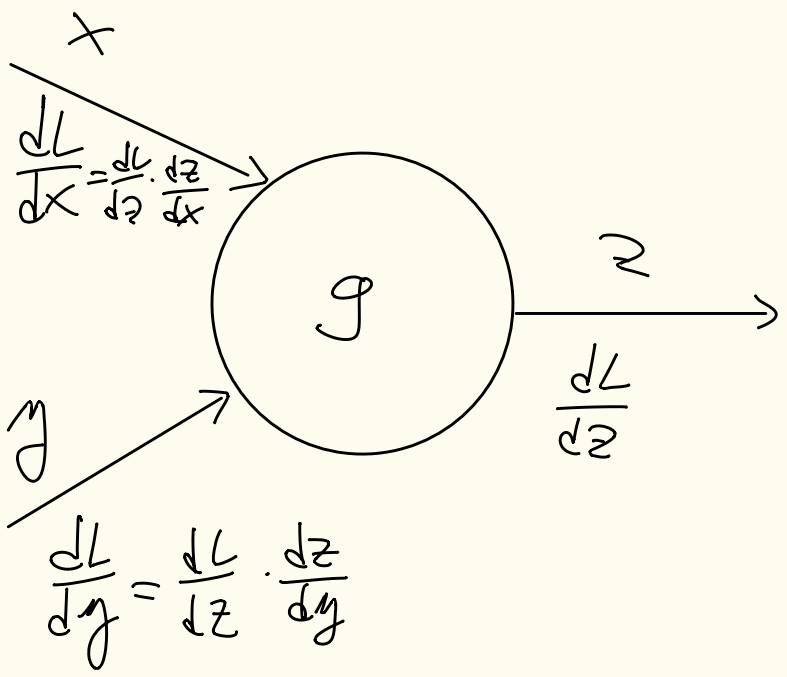
ex. w_2

$f_\theta(x) = (x + w_1) \cdot w_2$

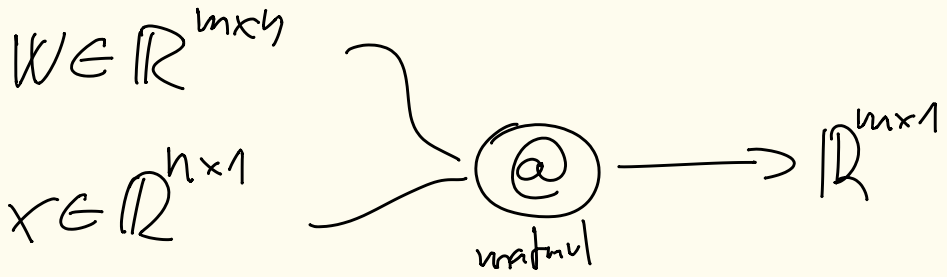
$\dots \{w_1, w_2\}$

$\frac{\partial f}{\partial x} = w_2 \quad \frac{\partial f}{\partial w_1} = w_2 \quad \frac{\partial f}{\partial w_2} = x + w_1$



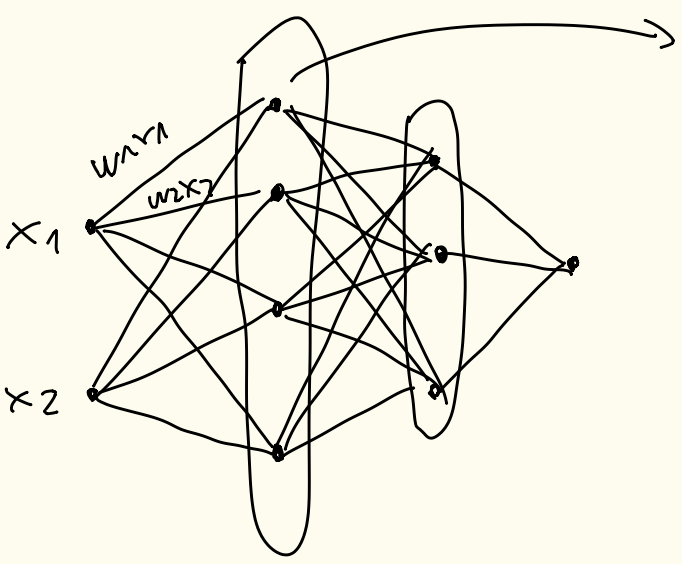



```
class MultiplyGate:
    def forward(self, x, y):
        self.y = y.copy()
        self.x = x.copy()
        return x * y
    def backward(self, dL/dz):
        dL/dx = dL/dz * self.y
        dL/dy = dL/dz * self.x
        return (dL/dx, dL/dy)
```



$x \rightarrow wx + b$

$\min (wx+b)^2$ / $wx + b = 0$



$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cdot \\ \vdots \\ \cdot \end{bmatrix}$$

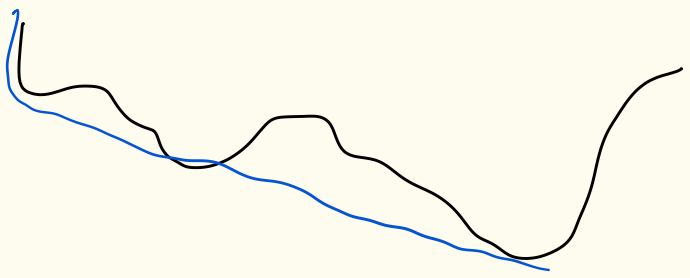
FC

full connected / linear : $wx + b$

sigmoid

FC, Sigmoid, FC ....

optimizer si muže pomahat setrvacnost a presmyknout se přes lokální minima



$$\Theta : \Theta - \lambda \frac{\partial L}{\partial \Theta}$$