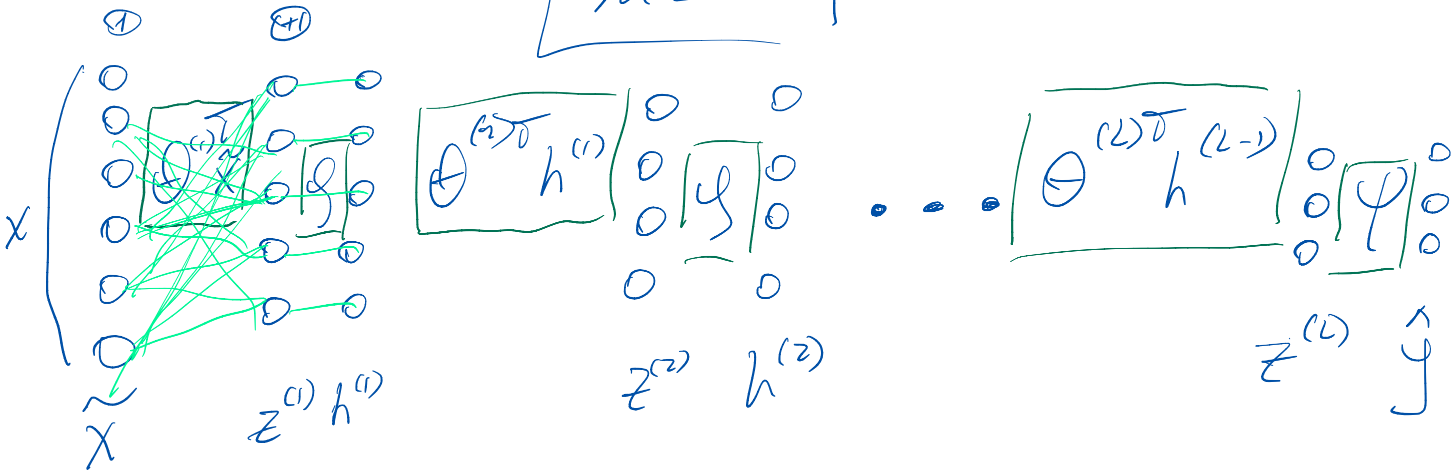
MLP



$$z^{(1)} = \Theta^{(1)T} \tilde{x} \qquad h^{(1)} = \varphi(z^{(1)})$$

$$z^{(2)} = \Theta^{(2)T} h^{(1)} \qquad h^{(2)} = \varphi(z^{(2)})$$

$$\vdots$$

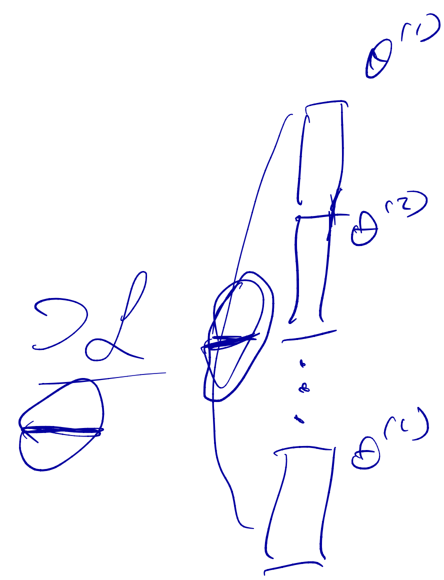
$$z^{(L)} = \Theta^{T(L)} h^{(L-1)} \qquad \hat{y} = \psi(h^{(L-1)})$$

$$\mathcal{T} : \{x_i, y_i\} \qquad x_i \in \mathbb{X}, \quad y_i \in \mathbb{Y}$$

$$\mathcal{L}(\Theta, \{x_i, y_i\}) = \begin{matrix} MSE(\Theta, \mathcal{T}) \\ BCE(\Theta, \mathcal{T}) \\ CE(\Theta, \mathcal{T}) \end{matrix}$$

$$\mathcal{L} = \mathcal{L}(\Theta^{(1)}, \Theta^{(2)}, \Theta^{(3)}, \ldots, \{x_i, y_i\})$$

$$\frac{\partial \mathcal{L}}{\partial \Theta^{(1)}}, \quad \frac{\partial \mathcal{L}}{\partial \Theta^{(2)}}\ ?$$



Градиентная оптимизация

1. θ_0, η, T, B, $C(\varepsilon)$
2. for t in $1 \ldots T$:

   a) $b(\mathcal{T}) \subset \mathcal{T}_B$

   b) $\theta_t = \theta_{t-1} - \eta \nabla_\theta \mathcal{L}(b(\mathcal{T}), \theta_{t-1}))$

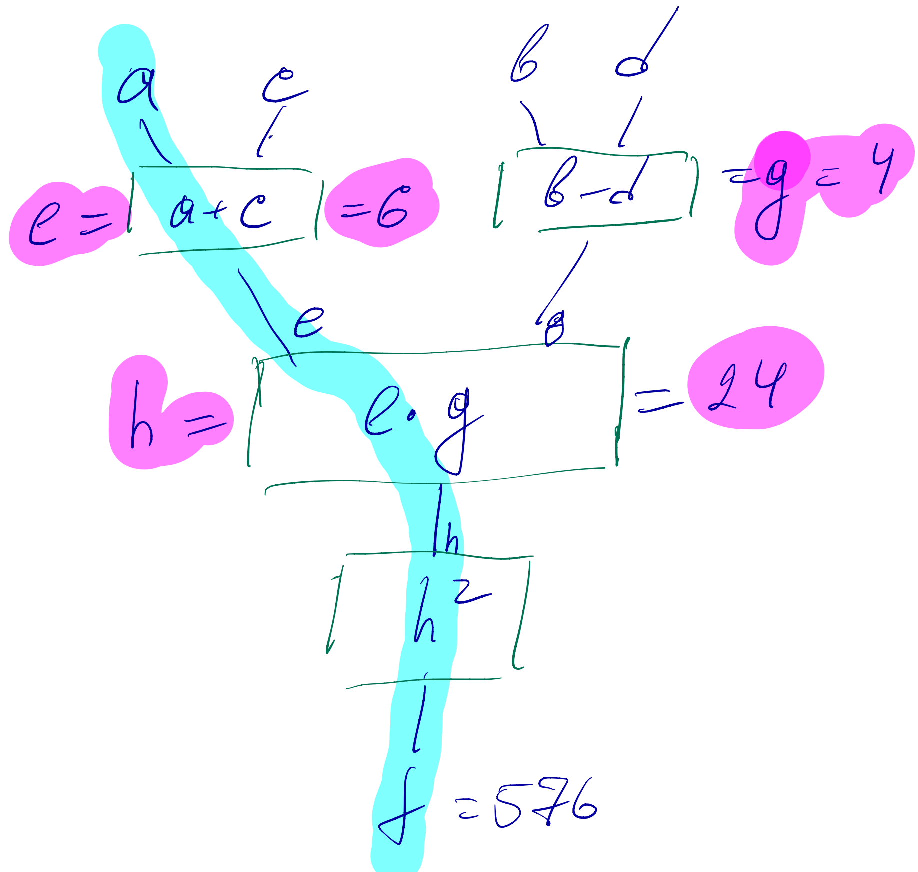
   c) if $C(\varepsilon)$: stop

$$\frac{\partial F(f(x), g(x))}{\partial x} = \frac{\partial F}{\partial f}\frac{\partial f}{\partial x} + \frac{\partial F}{\partial g}\frac{\partial g}{\partial x}$$

$$\frac{\partial f(g(x))}{\partial x} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial x}$$

$$f = ((a+c)\cdot(b-d))^2 = 24^2 = 576$$

$$\quad 1 \quad 5 \qquad 7 \quad 3$$



$$\frac{\partial f}{\partial a} = \frac{\partial f}{\partial h}\,\frac{\partial h}{\partial e}\,\frac{\partial e}{\partial a}$$

$$\frac{\partial f}{\partial a} = 2h\cdot g\cdot 1 =$$

$$= 48\cdot 4\cdot 1 = 192$$

$$f(h) = h^2 \qquad \frac{\partial f}{\partial h} = 2h$$

$$h(e) = eg \qquad \frac{\partial h}{\partial e} = g$$

$$e(a) = a+c \qquad \frac{\partial e}{\partial a} = 1$$

$$\hat{y} = F_{NN}(\theta, x) \qquad \frac{\partial \mathcal{L}}{\partial \theta} \qquad \frac{\partial \mathcal{L}}{\partial \hat{y}} \qquad \frac{\partial \mathcal{L}}{x}$$

$$\mathcal{L}(\hat{y}, x, \theta) \qquad \frac{\partial \hat{y}}{\partial \theta} \qquad \frac{\partial \hat{y}}{\partial x}$$

$\hat{y}_i = \mu_i = I(\theta^T x_i)$

θ, x_i – вектор-столбцы

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y} - y)^2$$

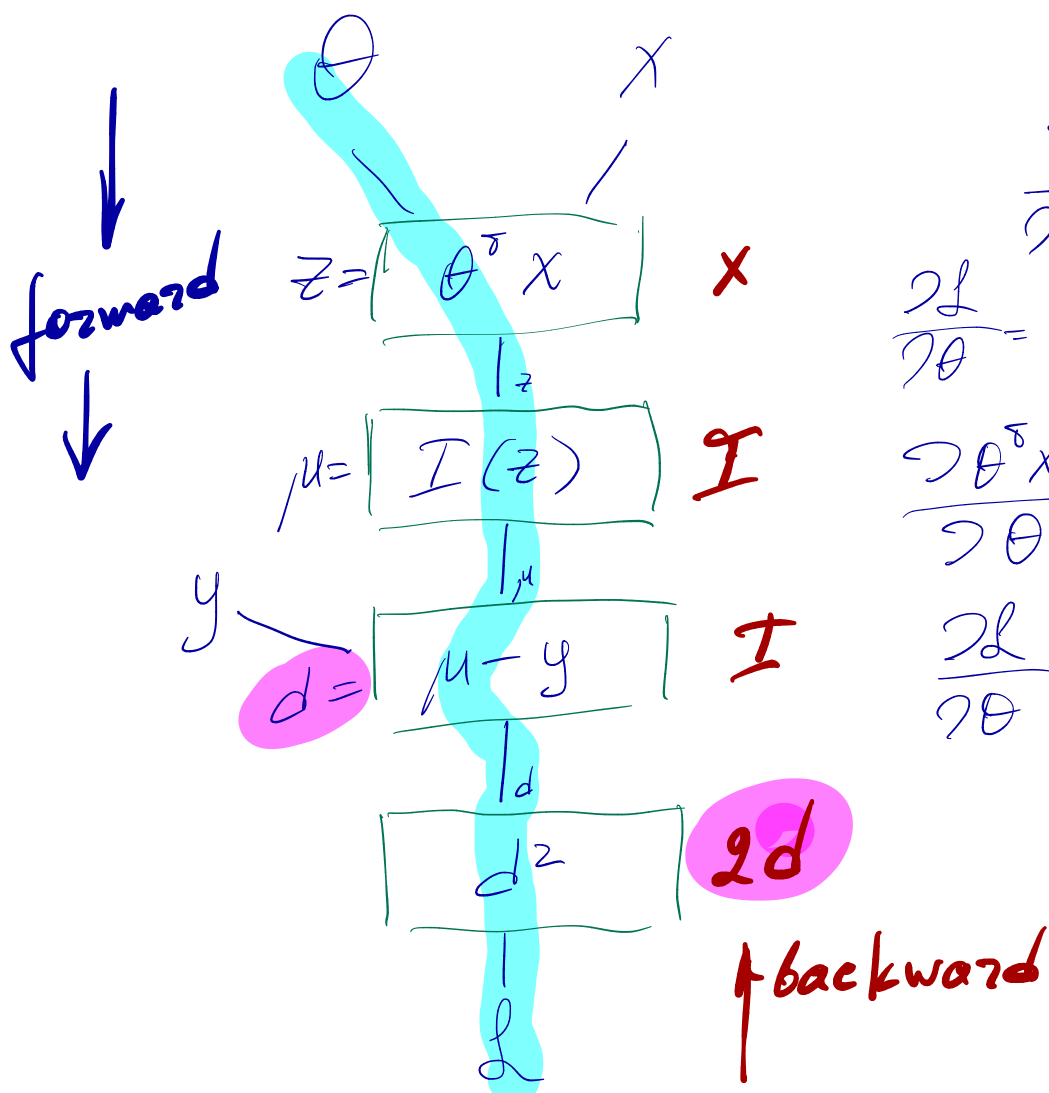
$\frac{\partial \mathcal{L}}{\partial \theta}$ – ?

$x \in \mathbb{R}^D$

$\theta \in \mathbb{R}^D$

$$\frac{\partial \mathcal{L}}{\partial \theta} = \vec{e}_0 \cdot \frac{\partial \mathcal{L}}{\partial \theta_0} + \vec{e}_1 \frac{\partial \mathcal{L}}{\partial \theta_1} + \dots + \vec{e}_D \frac{\partial \mathcal{L}}{\partial \theta_D}$$

$$\frac{\partial \mathcal{L}}{\partial \theta} \in \mathbb{R}^D$$



$\mathcal{L} = (\mu - y)^2$

$\frac{\partial \mathcal{L}}{\partial \theta}$ – ?

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial \mathcal{L}}{\partial d} \cdot \frac{\partial d}{\partial \mu} \cdot \frac{\partial \mu}{\partial z} \cdot \frac{\partial z}{\partial \theta}$$

$$\frac{\partial \theta^T x}{\partial \theta} = x$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = 2d \cdot I \cdot I \cdot x$$