$\mathcal{L} = \mathcal{L}(\theta, \{x_i, y_i\})$ $\quad \nabla_\theta \mathcal{L}$

$\theta^{(0)}$

1) $\nabla_\theta \mathcal{L}(\theta^{(t)}, B_t)$ $\quad B \sim \mathcal{J}_{(N_b)}$

2) $\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_\theta \mathcal{L}$

ema
exponential
moving
average

eazly stop

$ema(|\mathcal{L}|) \le \varepsilon$

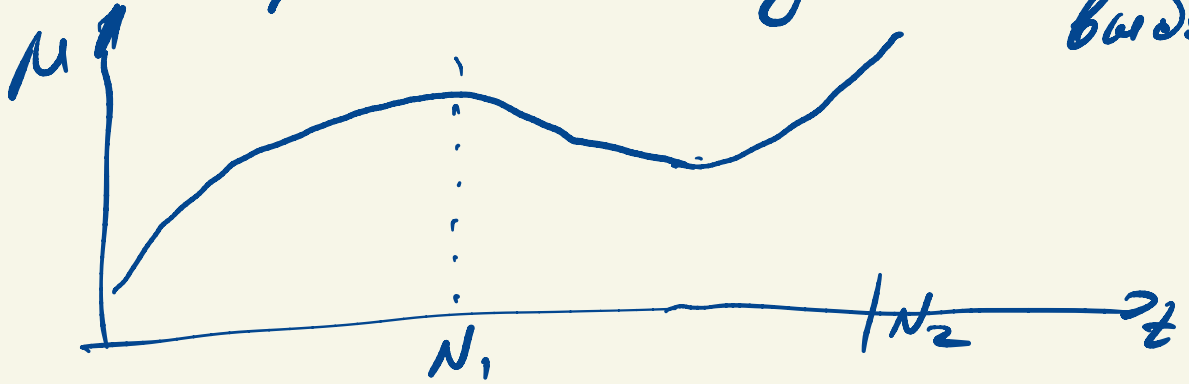
$ema(|\nabla \mathcal{L}|) \le \varepsilon$

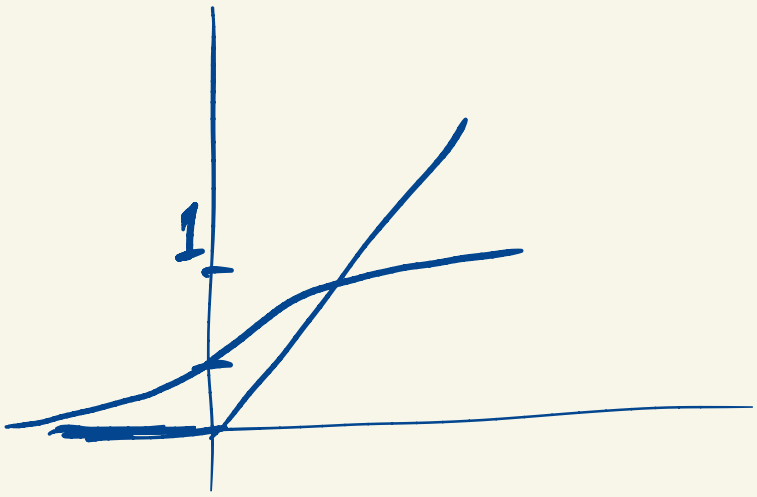
$ema(|M|) \ge (1-\varepsilon)$ (?)

$\mathcal{L}$ – ф-я потерь

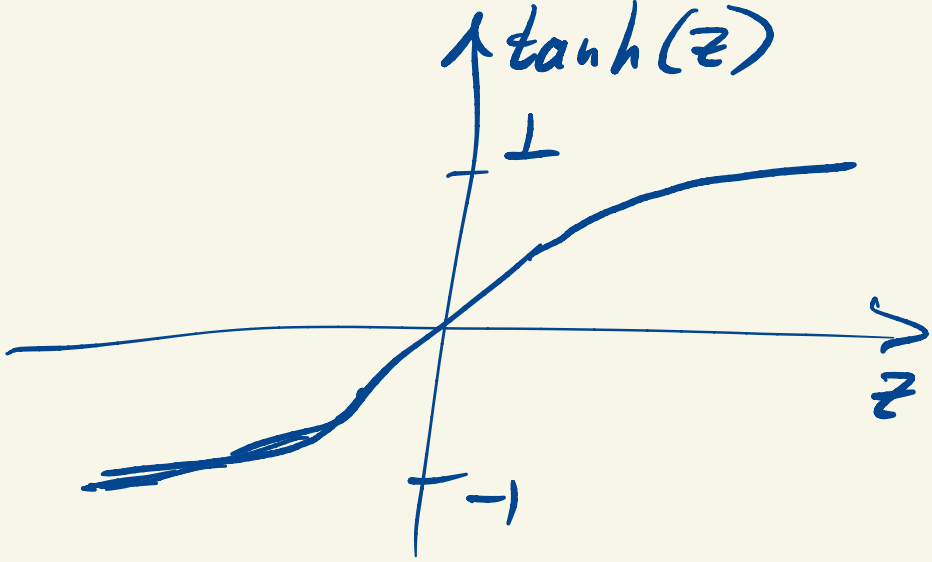
$\nabla \mathcal{L}$ – градиент ф-ии потерь

M – метр кач-ва модели на валид. выборке

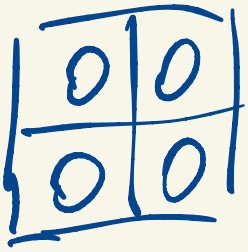


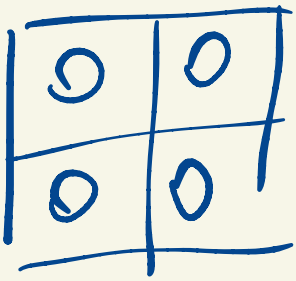

1

$\theta^{(t=0)}$



X 





$\hat{y} = (\theta_L \dots (\theta_2(\theta_1(\theta_0 x))))$

$\hat{y} = 0$ $\quad \nabla_{\theta_i} \mathcal{L}$ $\quad 2(\hat{y} - y) \cdot \frac{\partial \hat{y}}{\partial \theta_i}$

$$\theta^{(e)} = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix}$$

$\hat{y} = \theta^L x$ $\quad \hat{y} >> 0$ $\quad \|\hat{y} - y\| >> 0$

$\mathcal{L} >> 0$

$\nabla_\theta \mathcal{L} >> 0$

$x - z_1 \to h_1 - z_2 \to h_2 - \dots - \hat{y}$ $\mathcal{L}$

θ_0 θ_1

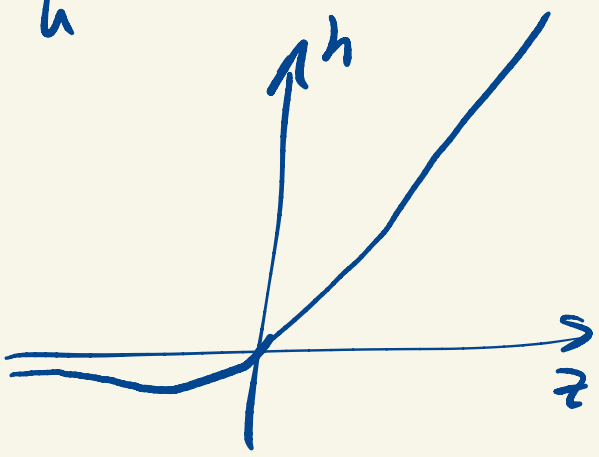
backprop

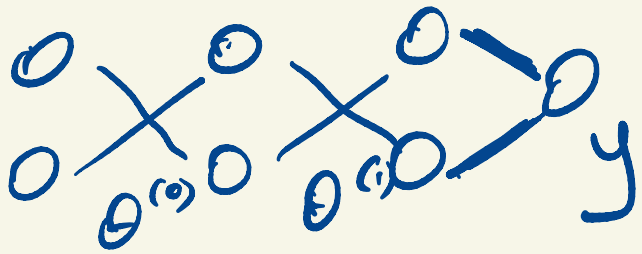
$$\frac{\partial \mathcal{L}}{\theta_i} = h_L h_{L-1} h_{L-2} \dots h_{i+1}$$

$h < L$ $\quad h^{(L-i)}$

$\sigma(z) = h$

$\tanh(z) = h$





$\mathcal{L}$

$\theta \sim N$

$Var\ \hat{y} = 0$