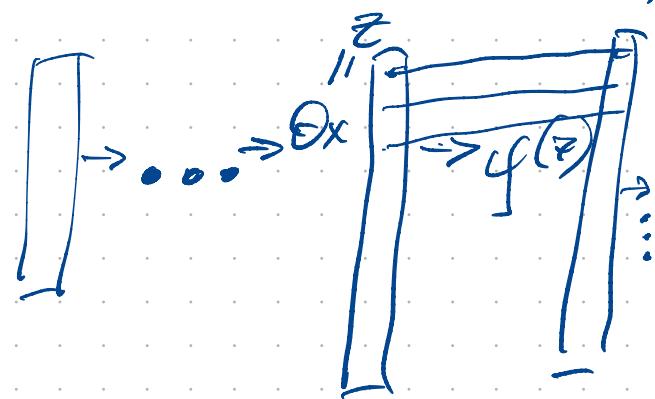


Численные методы

tanh(z)

G(z)

ReLU(z)



$$\text{ReLU} = \max(0, z) = \begin{cases} 0, & z < 0 \\ z, & z \geq 0 \end{cases}$$

$$G(z) = \frac{1}{1 + e^{-z}}$$

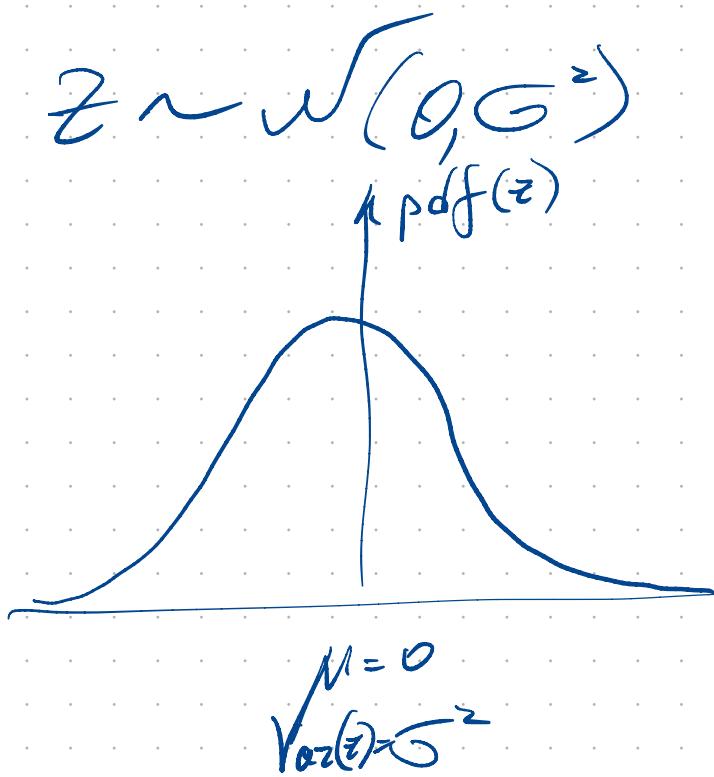
$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} = 2G(2z) - 1$$

$$h_i = g(x_i)$$

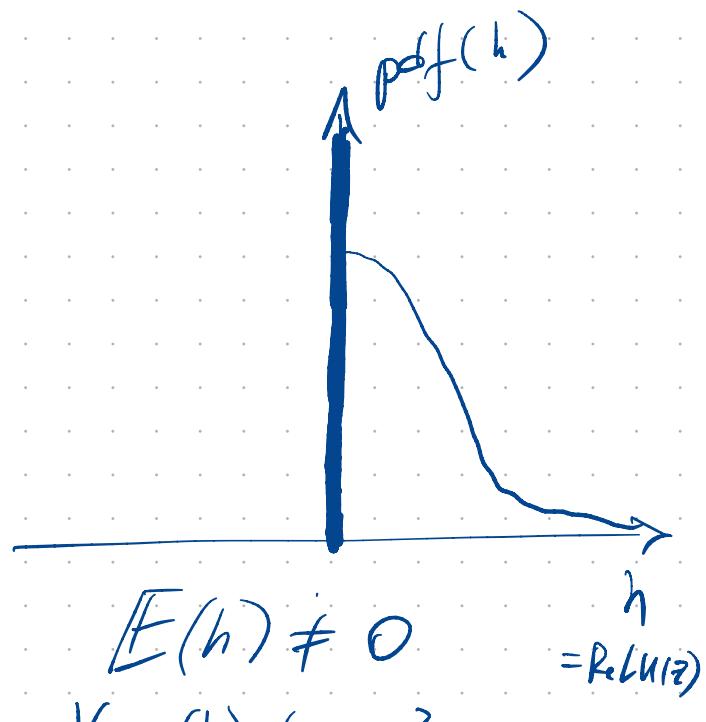
$$\frac{\partial g(z)}{\partial z}$$

$$G(1-G(z))$$

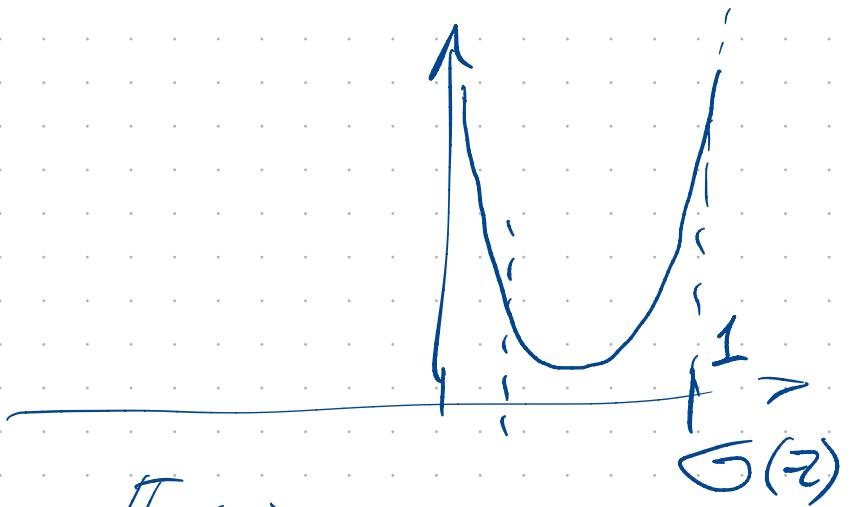
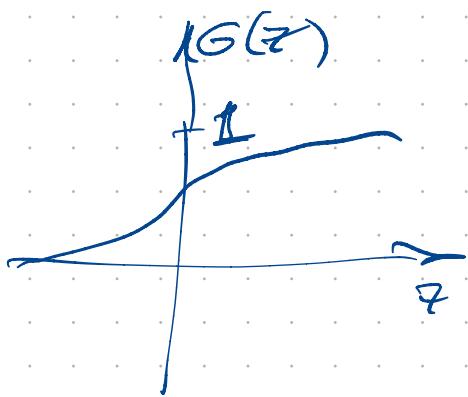
$$4G(z)(1-G(z))$$



$$\begin{aligned} M &= 0 \\ \text{Var}(z) &= G^2 \end{aligned}$$

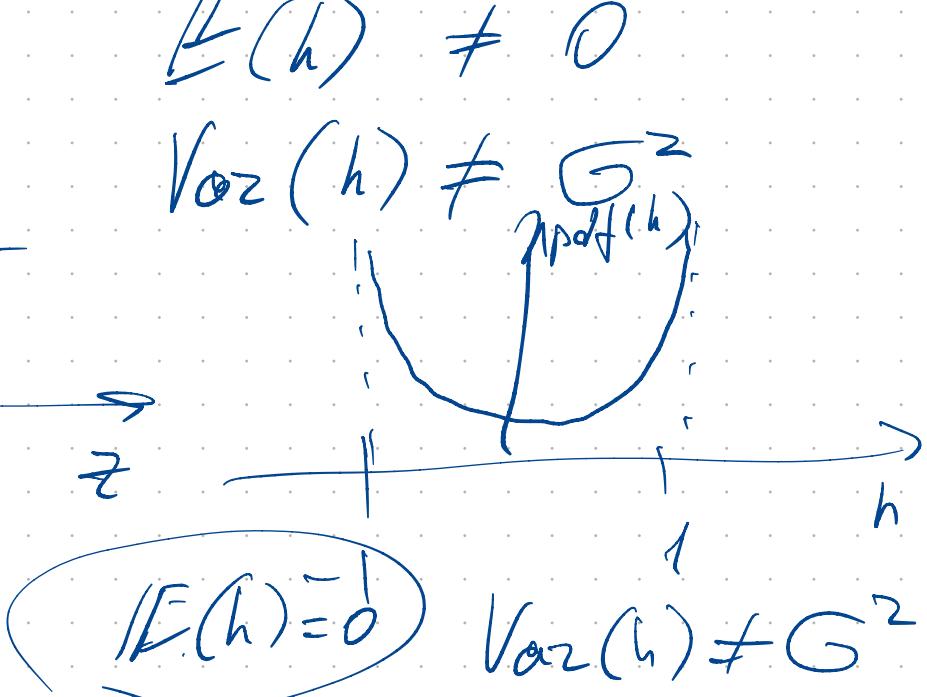
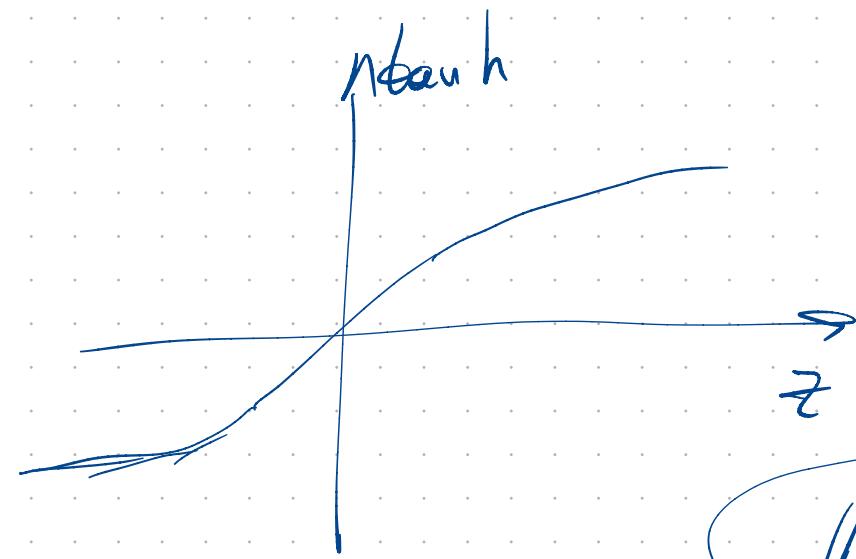


$$\begin{aligned} E(h) &\neq 0 \\ \text{Var}(h) &\neq G^2 \end{aligned}$$



$$E(h) \neq 0$$

$$\text{Var}(h) \neq G^2$$

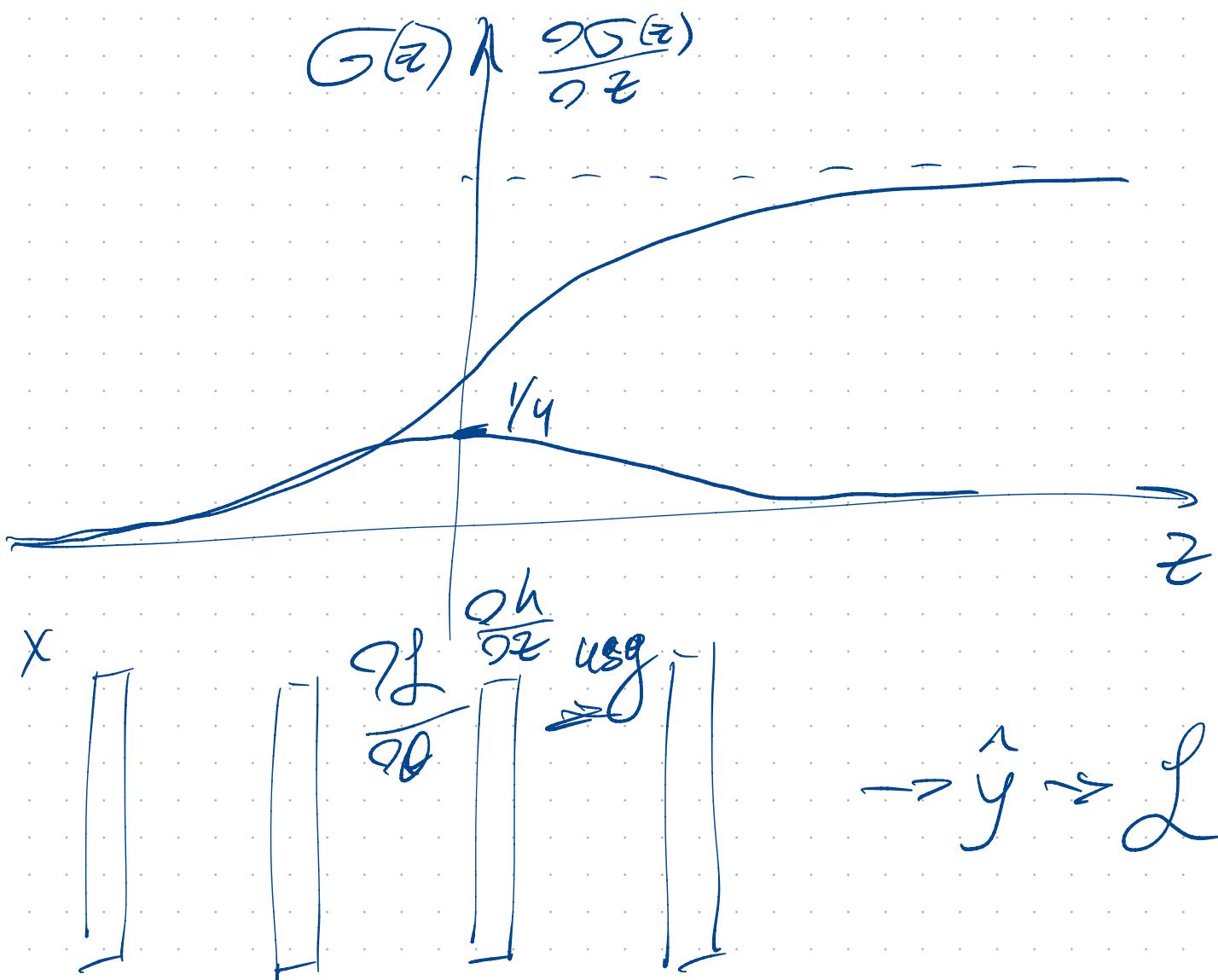


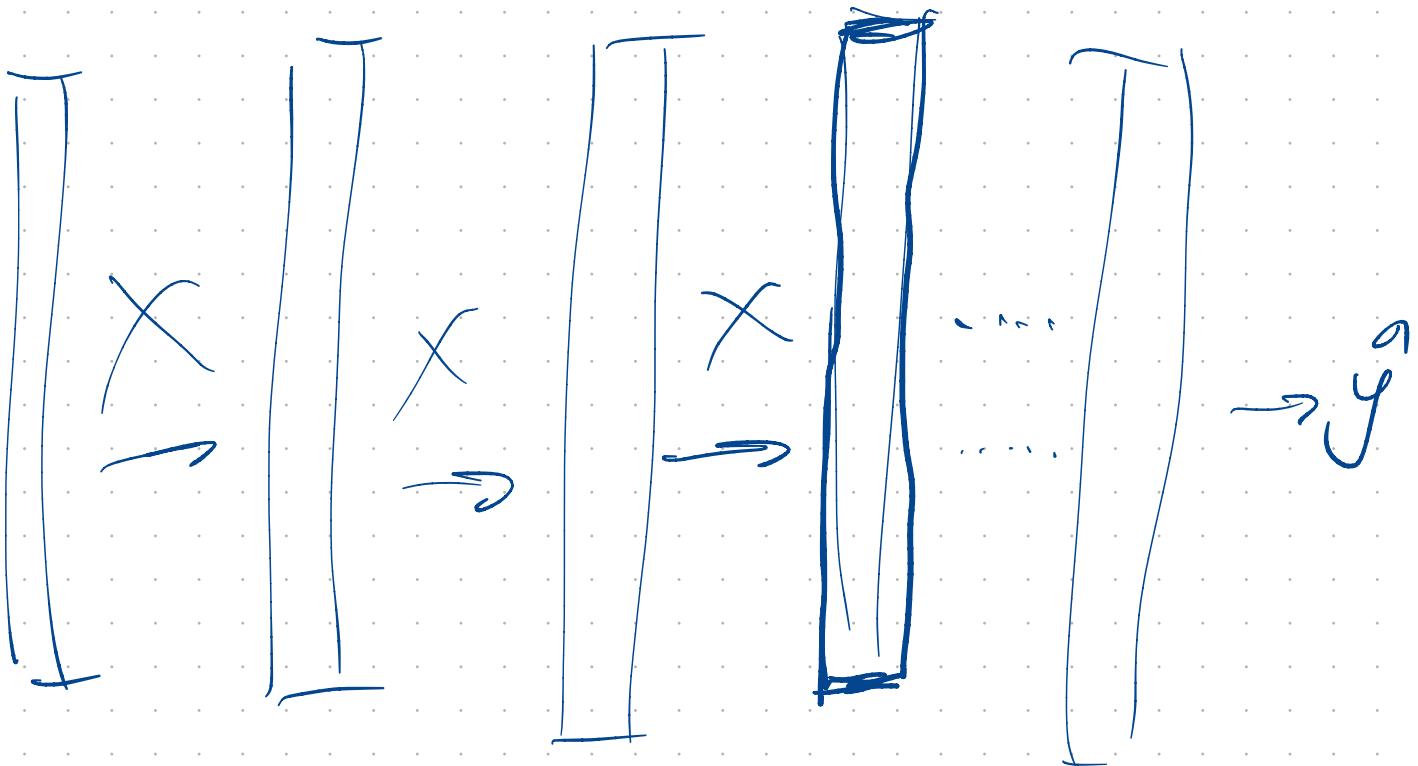
$$E(h) = 0$$

$$\text{Var}(h) = G^2$$

	$\text{ReLU}(z)$	$G(z)$	$\tanh(z)$
cochante $E(z)$	-	-	+
cochante $G^2(z)$	-	-	-

$G(z)$: Var(z) erzeugend
or end + end

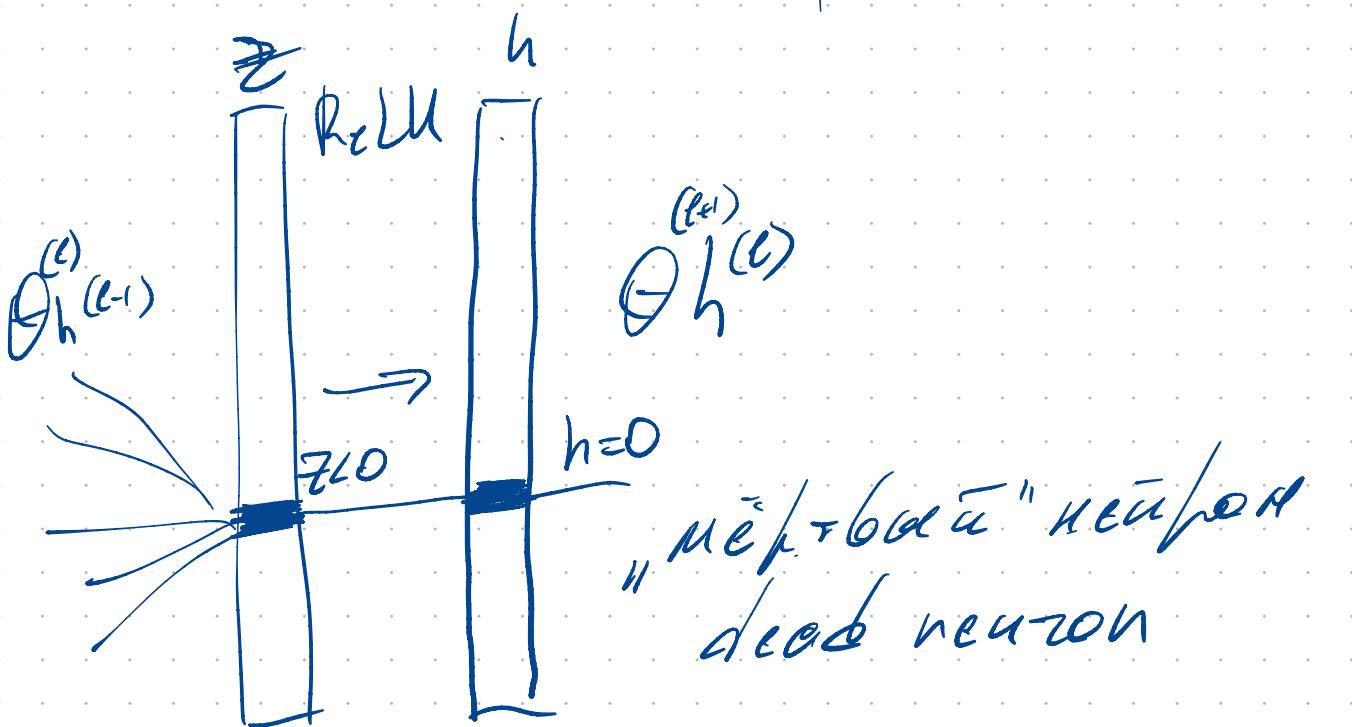
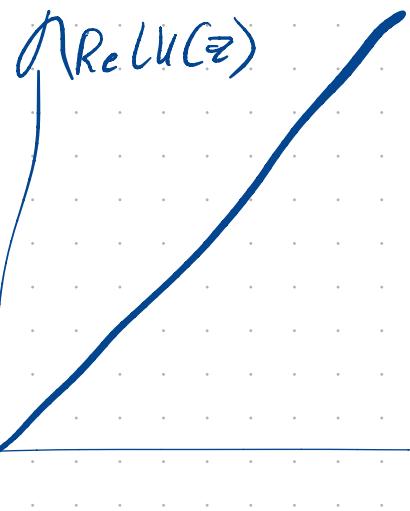




$\sigma(z)$: Проблема исчезающих градиентов (Vanishing gradients)

Никогда не используйте $\sigma(z)$ как f-ю ошибки для танк(z) на 很深 слоях
избыточных нейросетей

ReLU(z)



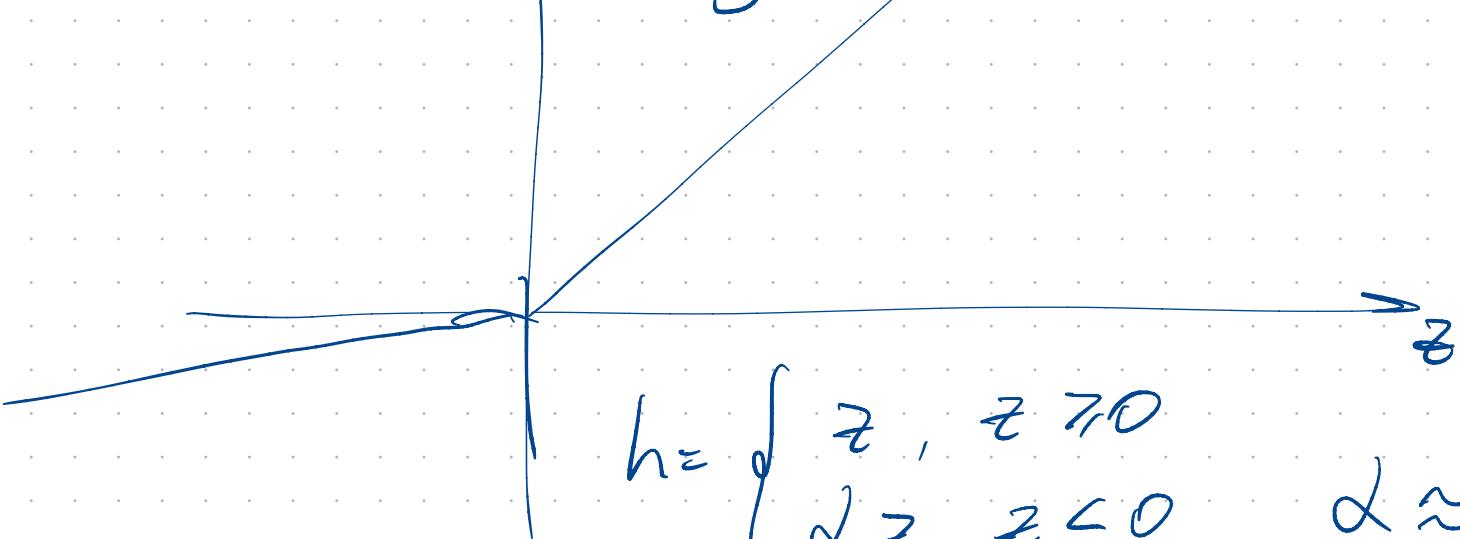
Reverence:

- ① Uebergangsphase
fuer aktiveren
- ② Uebergangsphase BN, Dropout
- ③ Residual connections
- ④ Parametric ReLU
Leaky ReLU

④

Leaky ReLU

Leaky ReLU



$$h = \begin{cases} z, & z \geq 0 \\ dz, & z < 0 \end{cases}$$

$d \approx 0.01$

PRReLU

$$h = \begin{cases} z, & z \geq 0 \\ p z, & z < 0 \end{cases}$$

p-trainable

$$\frac{\partial L}{\partial P}$$

$$\frac{\partial h}{\partial P} = \begin{cases} 0, & z \geq 0 \\ z, & z < 0 \end{cases}$$

Exponential Linear Unit

$$\psi(z) \in \text{ELU}(z) = \begin{cases} z, & z \geq 0 \\ \lambda(e^z - 1), & z < 0 \end{cases}$$

Even $z \sim \mathcal{N}(0, 1)$

$$\theta \sim \mathcal{N}\left(0, \frac{\sqrt{6}}{\sqrt{m+k}}\right) \quad | \quad \text{SD}$$

$$E\psi(z) = E(z)$$

Scaled ELU

$$\psi(z) \in \text{SELU}(z) = \gamma \begin{cases} z, & z \geq 0 \\ \lambda(e^z - 1), & z < 0 \end{cases}$$

SELU exchanges $E(z)$, $\text{Var}(z)$

$$\text{even } \theta \sim \mathcal{N}\left(0; \frac{1}{(m+n)}\right)$$

$$z \sim \mathcal{N}(0, 1)$$

$$\gamma = 1.0507 \quad \lambda = 1.6752$$

$$\text{Swish} = XG(X)$$

Swish - f-с активации, подогреваемое в процессе обучения комбинацией предыдущих f-с.

Swish - порождает более многошар
нейросети любой неравнинной;

- Не доставляет проблем "мертвых" нейронов
- Содержит способность к глубоким
нейросетям ($\sim 40 - 50$ слоёв)
и обучению долгое и
стабильное в ReLU
- Каждико порождает ReLU
при предыдущих предыдущих
batch-size

$$\text{Mish}(z) = z \tanh(\log(1 + e^z))$$

$$\approx x \left(1 - \frac{9}{4 \left(1 + \left(1 + \frac{x}{16} \right)^{16} \right)^2} \right)$$