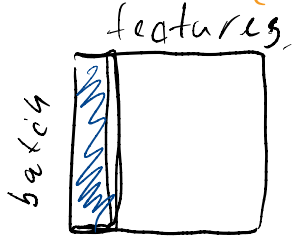
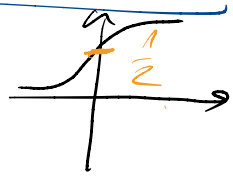
b p f.

h (batch x features) $h = \theta^T x$.

$$\hat{h} = \frac{h - \mu(h)}{std(h) + \varepsilon} \cdot \gamma + \zeta$$



$\sigma(x)$ $\frac{1}{1+e^{-x}}$

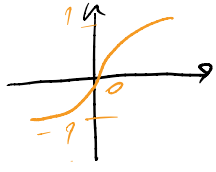


1) запускаем градиент.

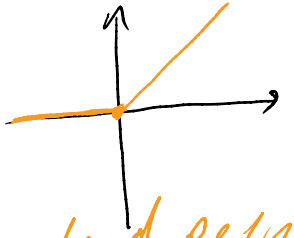
2).

$\tanh = 2\sigma(2x) - 1$

1) — // —



$ReLu(x) = \max(0, x)$

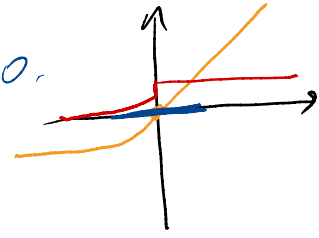


1) Смещение.

2) Неравное. активации

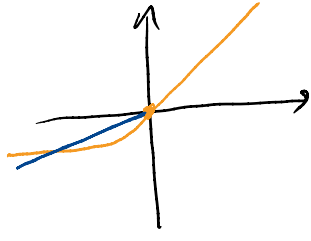
dead ReLu.

$$elu(x) = \begin{cases} x, & x > 0 \\ \alpha \cdot (e^x - 1), & x \le 0 \end{cases}$$



$$L\_relu(x) \quad \text{leaky relu} \begin{cases} x, & x > 0 \\ \alpha \cdot x & x \le 0 \end{cases}$$

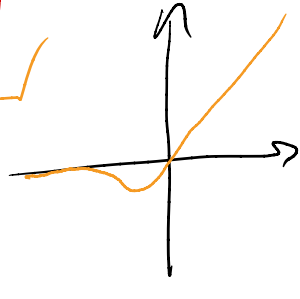
$$\max(\alpha x, x)$$



$$selu(x) = \lambda \begin{cases} x, & x > 0 \\ \alpha \cdot (e^x - 1), & x \le 0 \end{cases}$$

$$Gelu(x) = x \, P(X \le x)$$

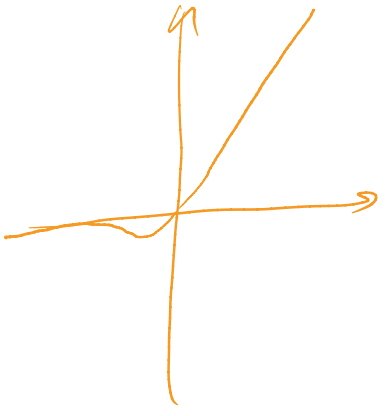
$$X \sim N(0,1)$$



swish

gate.

$$swish(x) = x\,\sigma(x)$$



$$mish(x) = x \cdot \tanh(softplus(x))$$

$$\ln(1 + e^x)$$

Dropout

претренировка

$$h_i = h_i \cdot b / \alpha$$

$$b \sim Ber(\alpha)$$

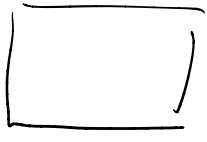
0.1

$$\hat{h}_i = h_i / \alpha$$

$\hat{h}_i = h_i \cdot \alpha_i$ — независимо.

$b \sim N(\ )$

1) много признаков. $640 \times 480 \times 3 =$

$= 921\,600$ $\Rightarrow$  517

921 600

2) признаки не независимые.

3) нет инвариантности по сдвигу.

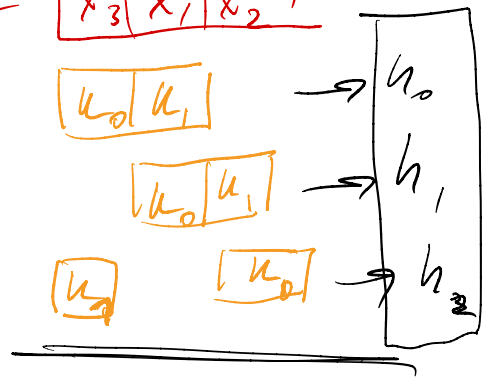
$x \in \mathbb{R}^3$.

$h_1 = \theta x$

$S^R = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

$S^R x = [x_3 | x_1 | x_2]$

S



$\theta(Sx) = S(\theta x)$

$\theta S = S \theta$

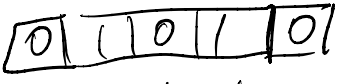
$\theta = \begin{pmatrix} k_0 & k_1 & k_2 \\ k_2 & k_0 & k_1 \\ k_1 & k_2 & k_0 \end{pmatrix}$ циркулянт.

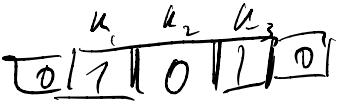
$h = \theta x$

$h_i = \sum_{j=0}^{2} k_{(0+2i) \bmod 3} \cdot x_j = [\theta * x]_i$

$$\Theta = \begin{pmatrix} k_0 & k_1 & 0 \\ 0 & k_0 & k_1 \\ k_1 & 0 & k_0 \end{pmatrix}$$

$$\frac{\Theta^T * X}{\sqrt{G(\theta)\, G(X)}}$$

$X =$ 

$k =$ 

$X * k = [8.3, 0, -4, \ldots]$

$k = [0, 0, 0, 0, 10]$

$X * k = [$