21.09.18 dl III

Методы инициализации нейросети

$z = Wx$, $f(z)$

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = W^T \frac{\partial f}{\partial z}$$

$$\mathbb{D} z_i = \mathbb{D}\left(\sum_j w_{ij} x_j\right) = \sum_j \mathbb{D}(w_{ij} x_j) = \sum_j \mathbb{D} w_{ij}\, \mathbb{D} x_j =$$

$$= n_{input}\, \mathbb{D} w_{ij}$$

$$\mathbb{D}\frac{\partial f}{\partial x_i} = n_{output}\, \mathbb{D} w_{ij}, \text{ потребуем: } \begin{cases} \mathbb{D} z_i = \mathbb{D} x_i \\ \mathbb{D}\left(\frac{\partial f}{\partial x_i}\right) = \mathbb{D}\left(\frac{\partial f}{\partial z_i}\right) \end{cases}$$

$$\text{тогда: } \begin{cases} n_{input}\, \mathbb{D} w_{ij} = 1 \\ n_{output}\, \mathbb{D} w_{ij} = 1 \end{cases} \approx \mathbb{D} w_{ij} = \frac{2}{n_{input} + n_{output}}$$

(Xavier) Glorot инициализация

① $w_{ij} \sim \mathcal{N}\left(w_{ij} \mid 0, \frac{2}{n_{input} + n_{output}}\right)$

$w_{ij} \sim \mathcal{U}\left[-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right]$, $\mathbb{E} w_{ij}^2 = \int_{-\frac{1}{\sqrt{n}}}^{\frac{1}{\sqrt{n}}} w_{ij}^2 \frac{\sqrt{n}}{2}\, dw_{ij} = \frac{\sqrt{n}}{2} \left.\frac{w_{ij}^3}{3}\right|_{-\frac{1}{\sqrt{n}}}^{\frac{1}{\sqrt{n}}} =$

$$= \frac{\sqrt{n}}{2}\,\frac{2}{3n\sqrt{n}} = \frac{1}{3n} = \frac{2}{n_{input} + n_{output}}$$

② $w_{ij} \sim \mathcal{U}\left[-\frac{\sqrt{6}}{\sqrt{n_{input} + n_{output}}}, \frac{\sqrt{6}}{\sqrt{n_{input} + n_{output}}}\right]$

③ Ортогональная матрица (с ортогональными строками)

$W \in \mathbb{R}^{n_{out} \times n_{inp}}$: $W^T W = I$

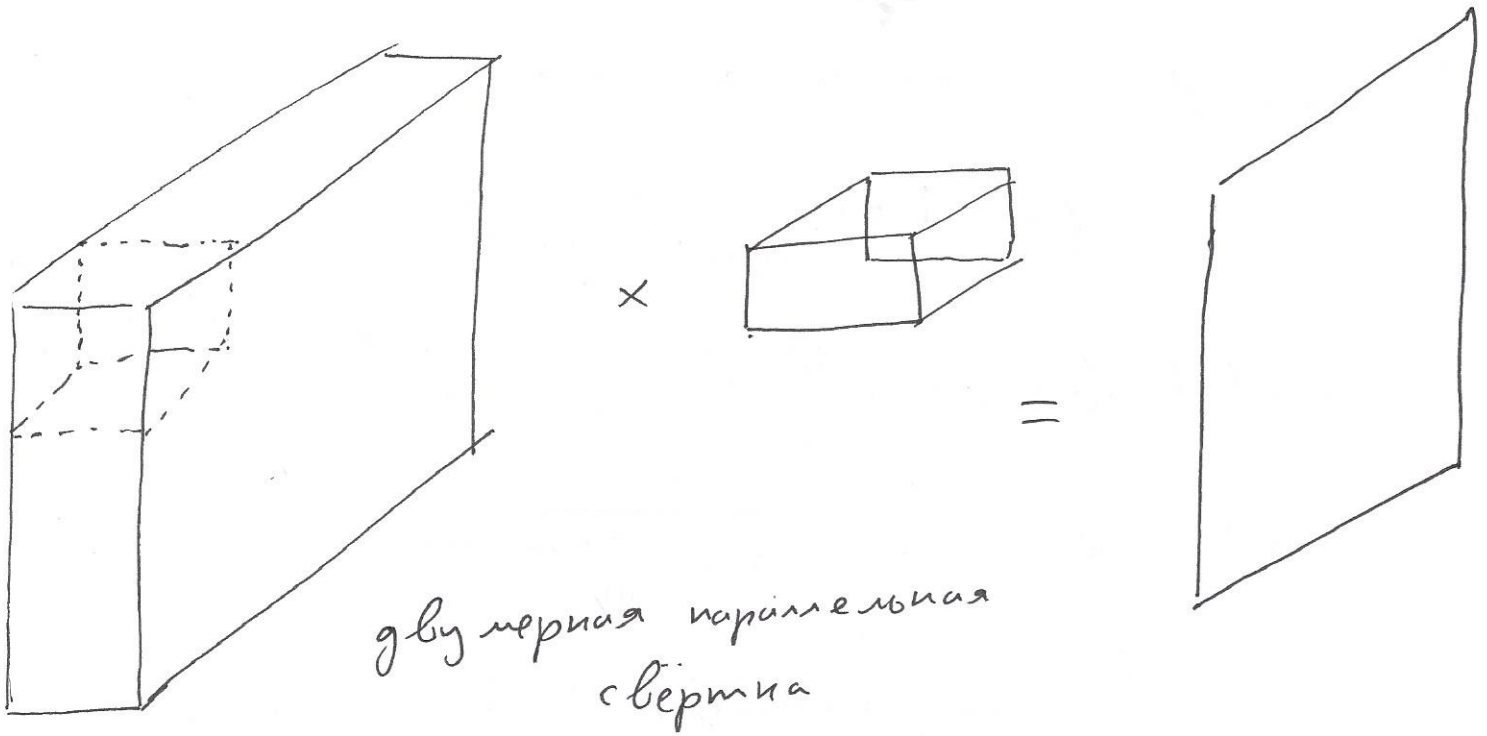
$z = Wx$, $\|z\|_2 = \|Wx\|_2 = \|x\|_2$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} k_1 & k_2 & 0 & k_3 & k_4 & 0 & \dots & 0 \\ 0 & k_1 & k_2 & 0 & k_3 & k_4 & \dots & 0 \\ 0 & 0 & 0 & k_1 & k_2 & 0 & k_3 k_4 & 0 \\ 0 & 0 & 0 & 0 & k_1 & k_2 & 0 & k_3 k_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_9 \end{bmatrix}$$

Свёрточные нейросети

$$\begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 \end{bmatrix} \times \begin{bmatrix} k_1 & k_2 \\ k_3 & k_4 \end{bmatrix} = \begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \end{bmatrix}$$

$$\tilde{I}(i,j) = \sum_{\delta_1, \delta_2 = -m}^{m} I(i+\delta_1, j+\delta_2) K(\delta_1+m, \delta_2+m)$$

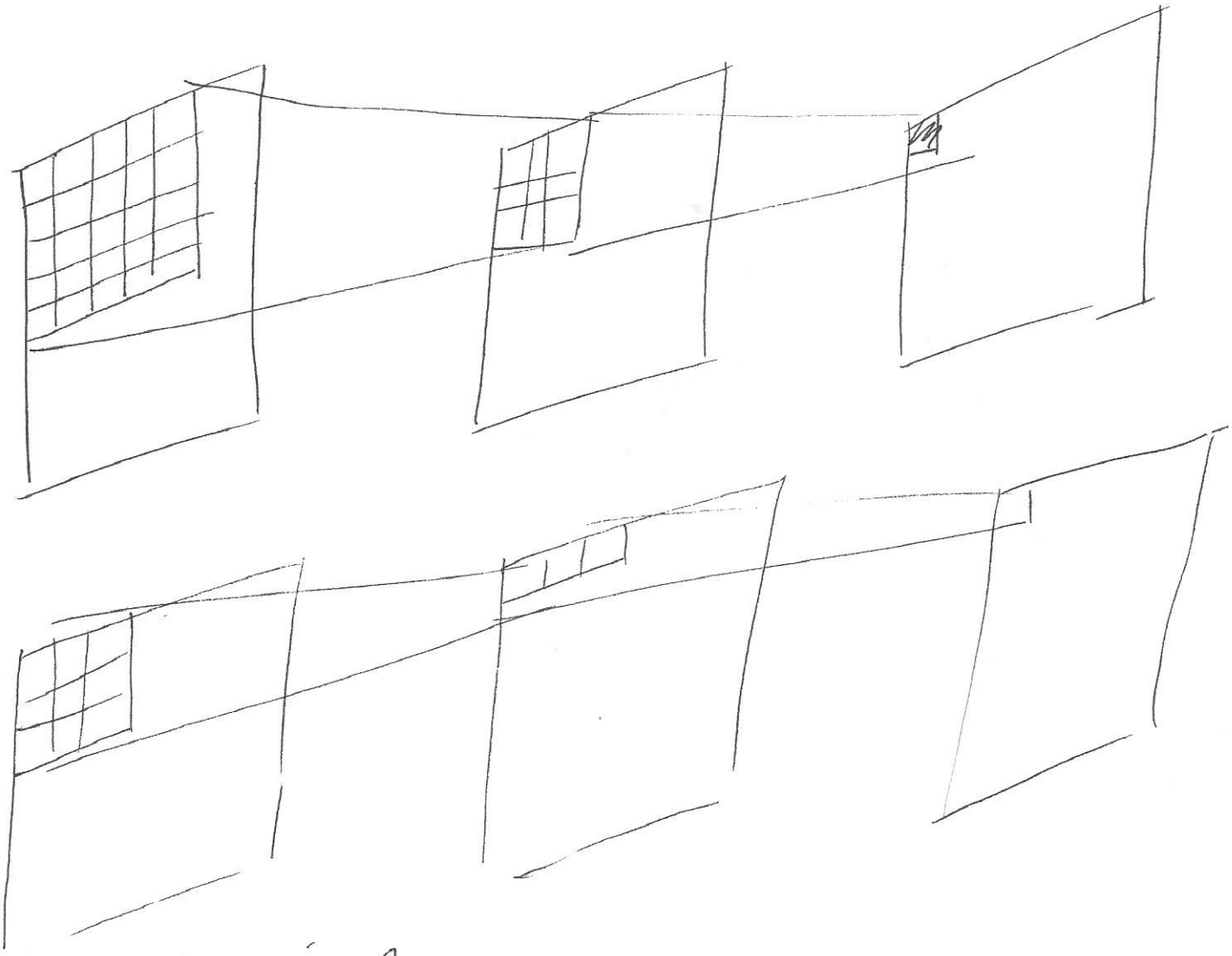


двумерная параллельная свёртка

Пулинг

$$\frac{\partial f}{\partial z}, \quad \frac{\partial f}{\partial x} = \begin{cases} \frac{\partial f}{\partial z}, & \text{если } x = \max(x, y) \\ 0, & \text{иначе} \end{cases}$$

Области видимости



разделение свертки 3×3 на 1×3 → 3×1