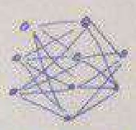
Коллоквиум 1. Токова Александра. Билет 3.

1. Свёрточные нейронные сети

До этого у нас были полносвязные сети.



У них были недостатки: 1) требуется огромное кол-во нейронов
2) серьёзное переобучение (т.к. если столько параметров, то сеть может просто запомнить выборку)

Сверт. нейр. слои – ~~[illegible]~~ одно из решений этих проблем.

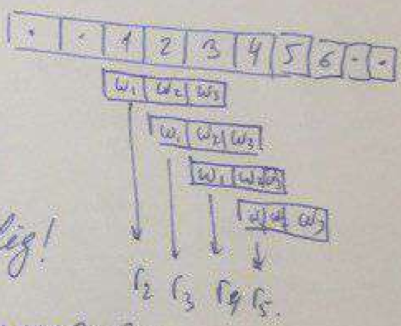
Сверт. слои – они хороши тем, что ловят внутренние инварианты картинки, и не зависят от того, где находится этот кусок картинки, который этот слой распознает. Т.е. полносв. слой, если на конт. [illegible] сдвинем – разрушится его распознавание, а свёрточный слой нет.

Опр. свёртки:

$n = m * a$ (массив, ядро)

1-мерная

$$n_k := \sum_{i=-\omega}^{\omega} m[k+i]\,a[-i].$$



! Свёрточное ядро пишется задом наперёд!

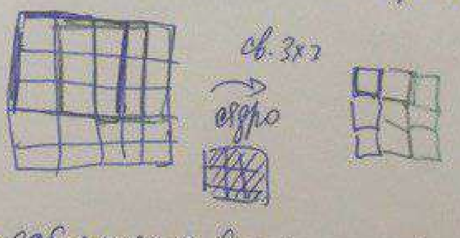
Т.е. свёрточное ядро как бы проезжает по массиву!

Padding – это сколько доп. клеток по краям добавлять.

~~[illegible]~~ Например, если padding = 1 (т.е. по 1 клетке с обеих сторон), а ядро длины 3 – то размер не изменится.

Вообще padding используется часто в режиме valid или same.

2-мерная свёртка:



Примеры ядер: ядро собеля для выделения границ, или такое:

0 1 0
1 -4 1
0 1 0

– часто свёртки 1×1, 3×3, 5×5.

Как использовать: nn.Conv2D(in_ch, out_ch, 3x3, padding=1)

② Найти $\frac{\partial(w^T A w)}{\partial w}$; $w \in \mathbb{R}^n$; $A \in \mathbb{R}^{n \times n}$.

~~$\frac{\partial}{\partial w}(w^T A w) = \frac{\partial (w^T)}{\partial w} A w + w^T \frac{\partial}{\partial w}(Aw) = \dots$~~

сейчас A – постоянн.

$$\underbrace{d(w^T A w)}_{df} = d(w^T) \cdot Aw + w^T \cdot d(Aw) = (dw)^T A w + w^T \overset{0}{\cancel{(dA)}} w + w^T A dw$$

(т.к. A = const сейчас)

но df – число ∨ и оба слагаемые – числа => если его транспонировать – будет оно же

$$\Rightarrow df = ((dw)^T \cdot Aw)^T + w^T A dw = w^T A^T dw + w^T A dw = w^T (A^T + A) dw$$

$$\Rightarrow \frac{\partial}{\partial w}(w^T A w) = w^T (A^T + A) = (A + A^T) w.$$

транспонируем чтобы размеры были правильные

$$\boxed{\text{Ответ: } \frac{\partial}{\partial w}(w^T A w) = (A + A^T) w}$$

③ Для лин. слоя $y = \sigma(\widetilde{w^T x})$ выразите $\frac{\partial L}{\partial w}$ через $\frac{\partial L}{\partial y}$; $x \in \mathbb{R}^n$; $y \in \mathbb{R}^m$.

$$\underset{1\times n}{\frac{\partial L}{\partial w}} = \underset{1\times 1}{\frac{\partial L}{\partial y}} \cdot \underset{1\times n}{\frac{\partial y}{\partial w}} = \underset{1\times 1}{\frac{\partial L}{\partial y}} \cdot \underset{1\times n}{\frac{\partial(\sigma(\widetilde{wx}))}{\partial w}} = \underset{1\times 1}{\frac{\partial L}{\partial y}} \cdot \sigma'(wx) \cdot \underset{1\times n}{x^T} = \frac{\partial L}{\partial y} \cdot \sigma(wx) \cdot (1 - \sigma(wx)) \cdot x^T$$

$\sigma'(z) = \sigma(z)(1 - \sigma(z))$

$$\boxed{\text{Ответ: } \frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \cdot \sigma(wx) \cdot (1 - \sigma(wx)) \cdot x^T}$$

④ $x \to z_1 = g(x, w)$; $x \to z_2 = f(x, w)$; $z_1, z_2 \to y = z_1 + z_2 \to L$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial w} = \frac{\partial L}{\partial y} \cdot \frac{\partial (z_1 + z_2)}{\partial w} = \frac{\partial L}{\partial y}\left(\frac{\partial z_1}{\partial w} + \frac{\partial z_2}{\partial w}\right) = \frac{\partial L}{\partial y}\left(\frac{\partial g}{\partial w} + \frac{\partial f}{\partial w}\right)$$

$$\boxed{\text{Ответ: } \frac{\partial L}{\partial w} = \frac{\partial L}{\partial y}\left(\frac{\partial g}{\partial w} + \frac{\partial f}{\partial w}\right)}$$

⑤ Напишите ф-лу градиента для convolution слоя

~~Возьмем все веса в~~ ~~возьмем и слой в вектор~~,

Возьмем 1-мерный сверт. слой.

Паддинга нет.

$x = (x_1; x_2; \dots x_n)$

$y = (y_1; \dots y_{n-k+1})$

$$y_i = \sum_{t=i}^{k+i-1} x_t \cdot w_{t-i+1}$$

т.е. $f(w_1 \dots w_k) = (x_1 w_1 + \dots + x_k w_k;\ x_2 w_1 + \dots + x_{k+1} w_k;\ \dots;\ x_{n+k-1} w_1 + \dots + x_n w_k)$

$\Rightarrow \frac{\partial f}{\partial \omega_1} = (x_1; x_2; \dots x_{n+k-1})$

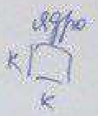
$\frac{\partial f}{\partial \omega_2} = (x_2; x_3; \dots x_{n+k})$

$\vdots$

$\frac{\partial f}{\partial \omega_k} = (x_k; x_{k+1} \dots ; x_n)$

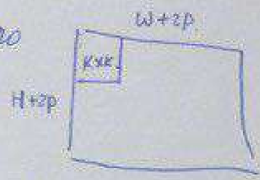
$$\Rightarrow \boxed{\nabla_{\bar{\omega}}(y) = \begin{pmatrix} x_1 & x_2 & \dots & x_{n+k-1} \\ x_2 & x_3 & \dots & x_{n+k} \\ \vdots & & & \\ x_k & x_{k+1} & \dots & x_n \end{pmatrix}}$$ ответ

(6)

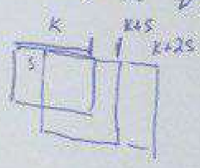


padding $p \times p$ stride $s \times s$ На выходе C' каналов.

Имеем: после паддинга стало



Сколько раз поместится K в $(W+2p)$ с шагом s?



Ну $\underline{(W+2p-K)//s+1}$ раз.

$\Rightarrow$ на выходе будет

$$\boxed{\begin{aligned} H_{новое} &= (H+2p-K)//s+1 \\ W_{новое} &= (W+2p-K)//s+1 \\ C_{новое} &= C' \end{aligned}}$$ — размеры выходного изобр.

Тензор ядра свертки: $\boxed{\underbrace{C}_{in\_ch} \times \underbrace{C'}_{out\_ch} \times \underbrace{K \times K}_{kernel\text{-}size}}$