

Class Plan

Reminder

Recap

Convolutional NN

Residual NN

Reminders

• 24/26 ☺

• Homework due 5pm

• All problems posted

• No office hours Thursday

• Today's problem more
"free form"

Yesterday's Topics

- GPU - matrix mul

↳ batches

- Variants of SGD

↳ momentum

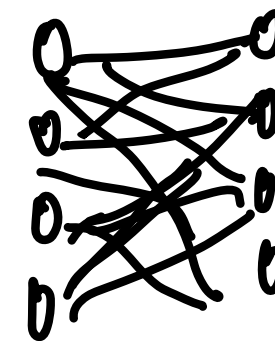
↳ adaptive learning rate

↳ Adam

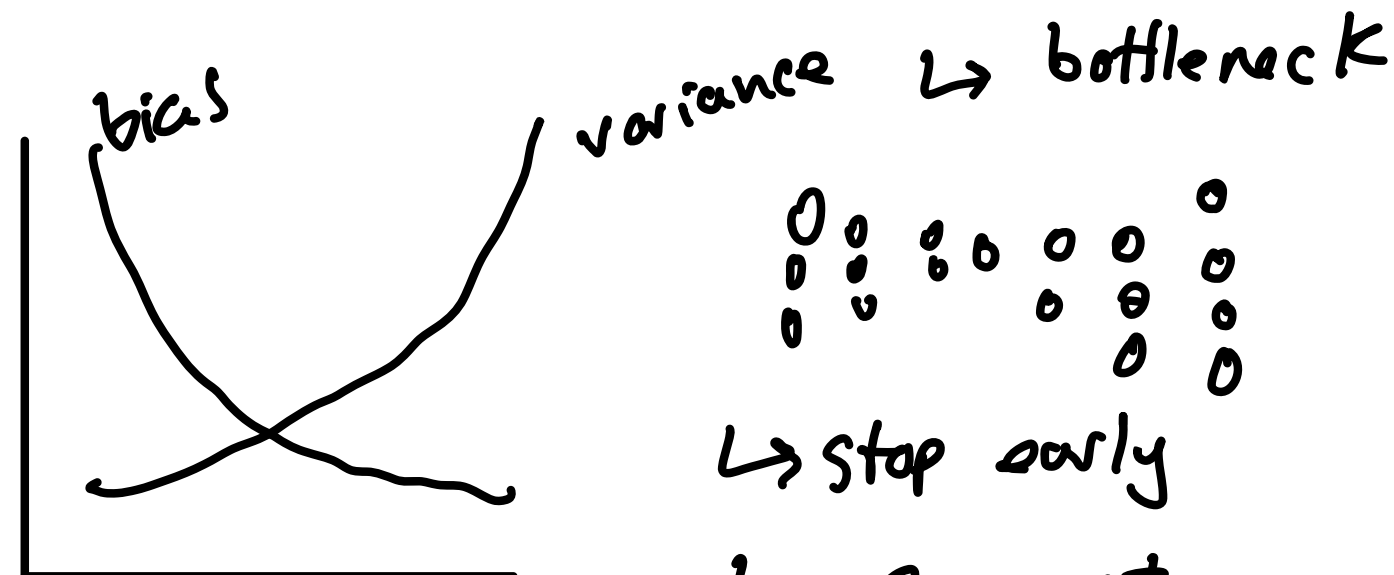
- Initialization

↳ exploding / vanishing

↳ symmetry



- Generalization



↳ stop early

↳ augment

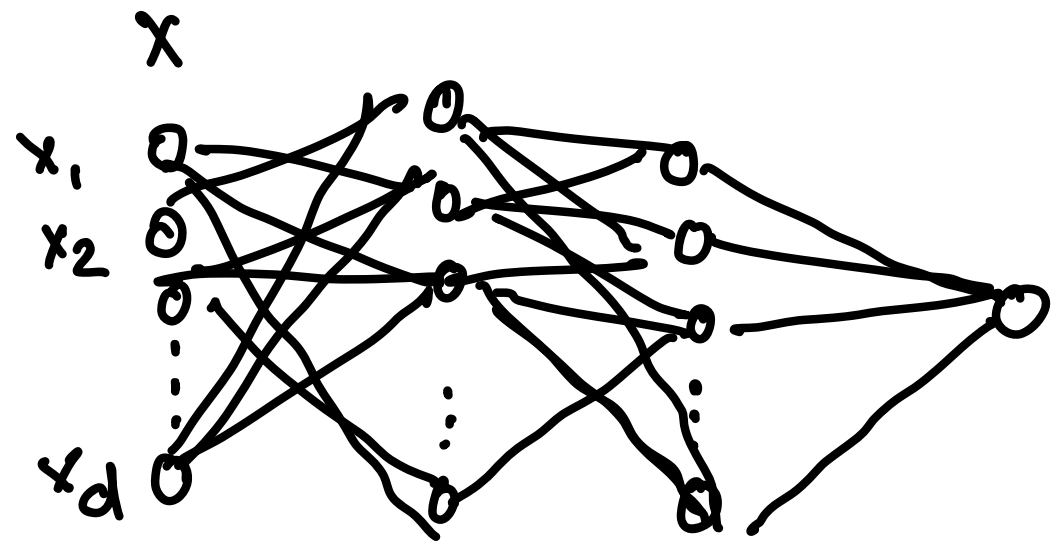
↳ drop out

↳ transfer

model complexity

↳ weight decay

Dense Neural Network

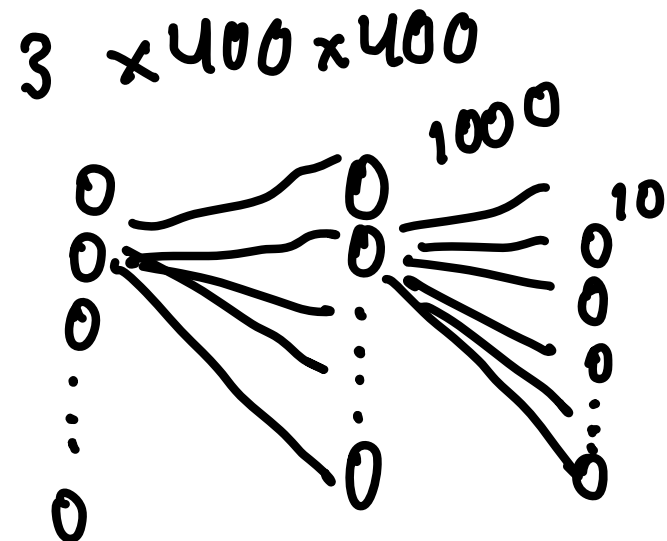


$$W'' \sigma(W' \sigma(Wx))$$

Problems:

Computational cost

①

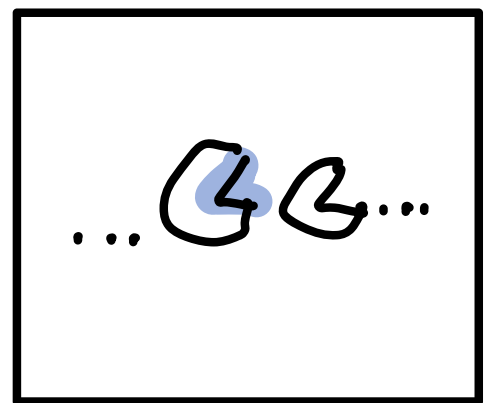
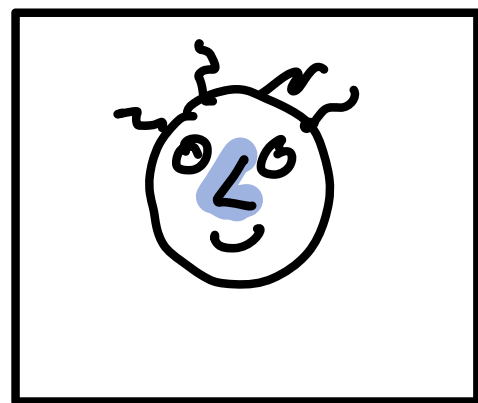


$$\begin{aligned} \# \text{ param} &= \\ &= 3 \cdot 400 \cdot 400 \cdot 1000 \\ &+ 1000 \cdot 10 \\ &= 480,010,000 \end{aligned}$$

$$O(L d_{\max}^2)$$

↳ need a lot data
(for our many parameters)

② Loss of context

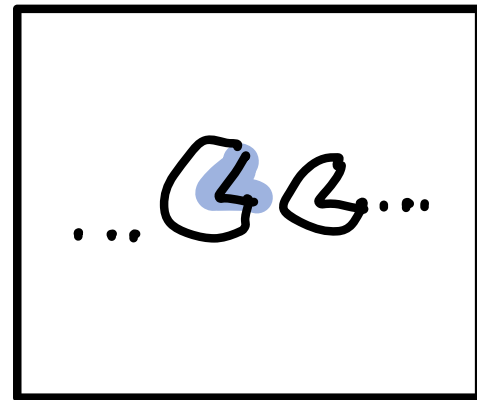
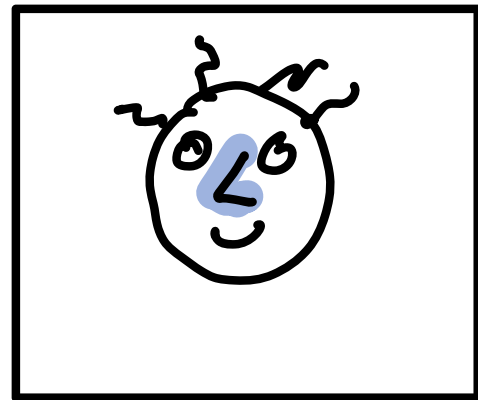


audio, language

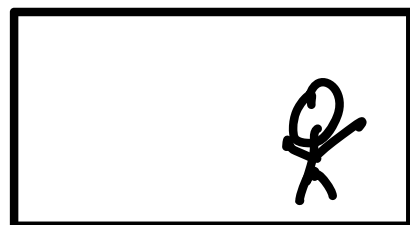
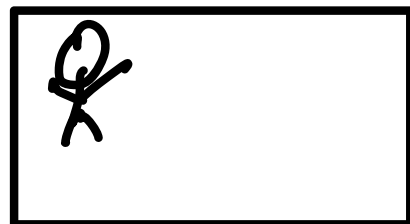
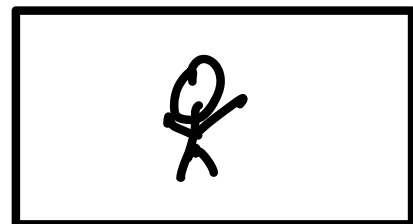
③ confounding factors

Convolutional Layers!

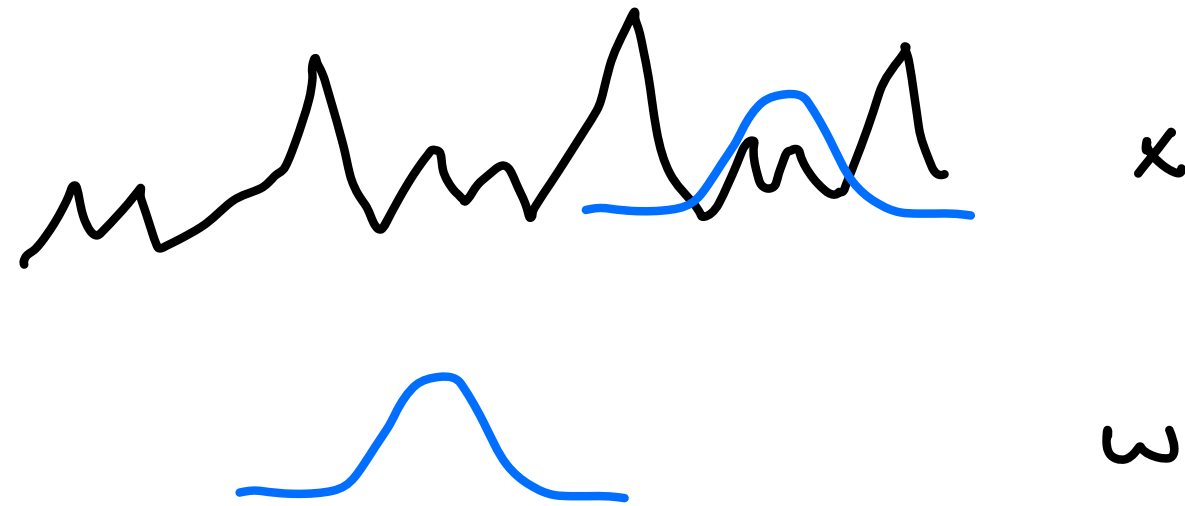
↳ Principle 1: Locality



↳ Principle 2: Shift Invariance



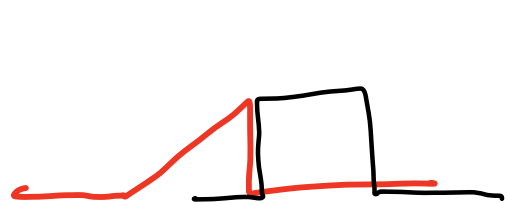
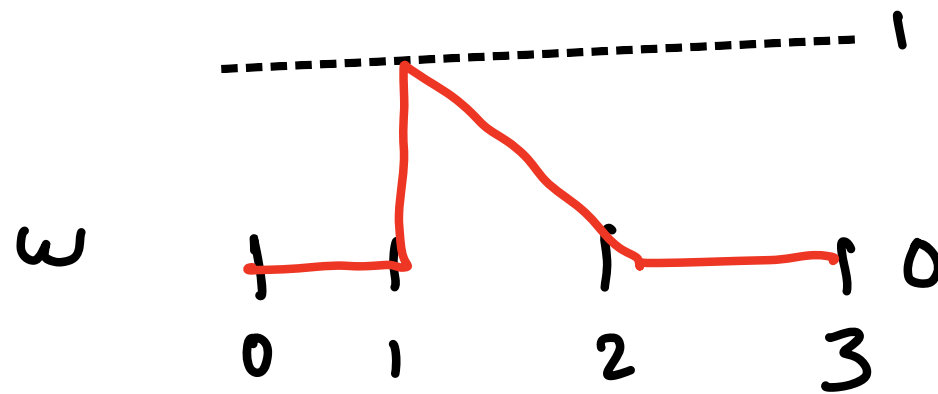
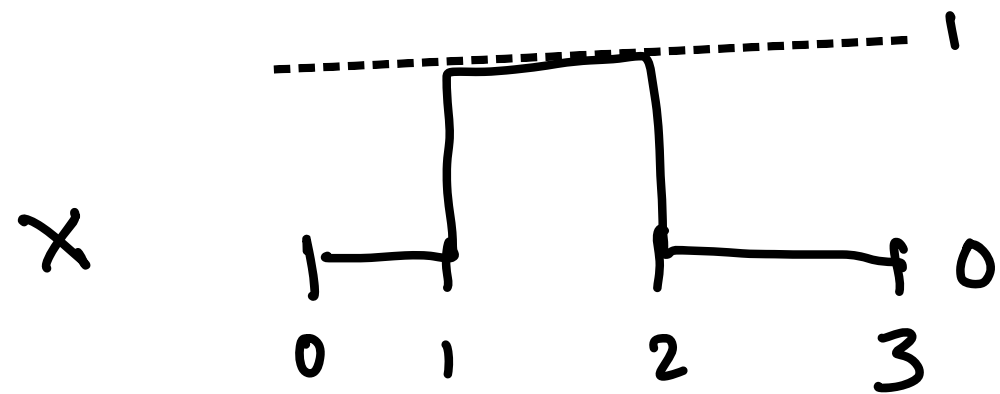
"Moving Average"



$x * w$

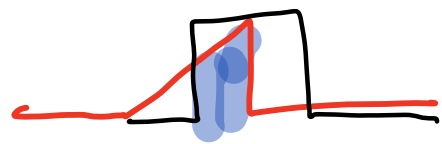


$$(x \star w)[t] = \sum_{\tau} x[\tau] w[t-\tau]$$

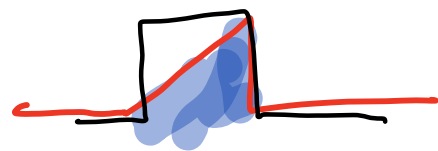


$$0 + \frac{1}{2} \cdot 0 + 1 \cdot 0 + 0 \cdot 1$$

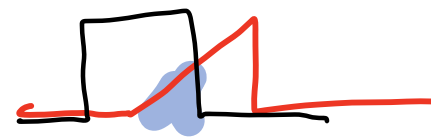
0



$\frac{1}{4}$



$\frac{1}{2}$



$\frac{1}{8}$



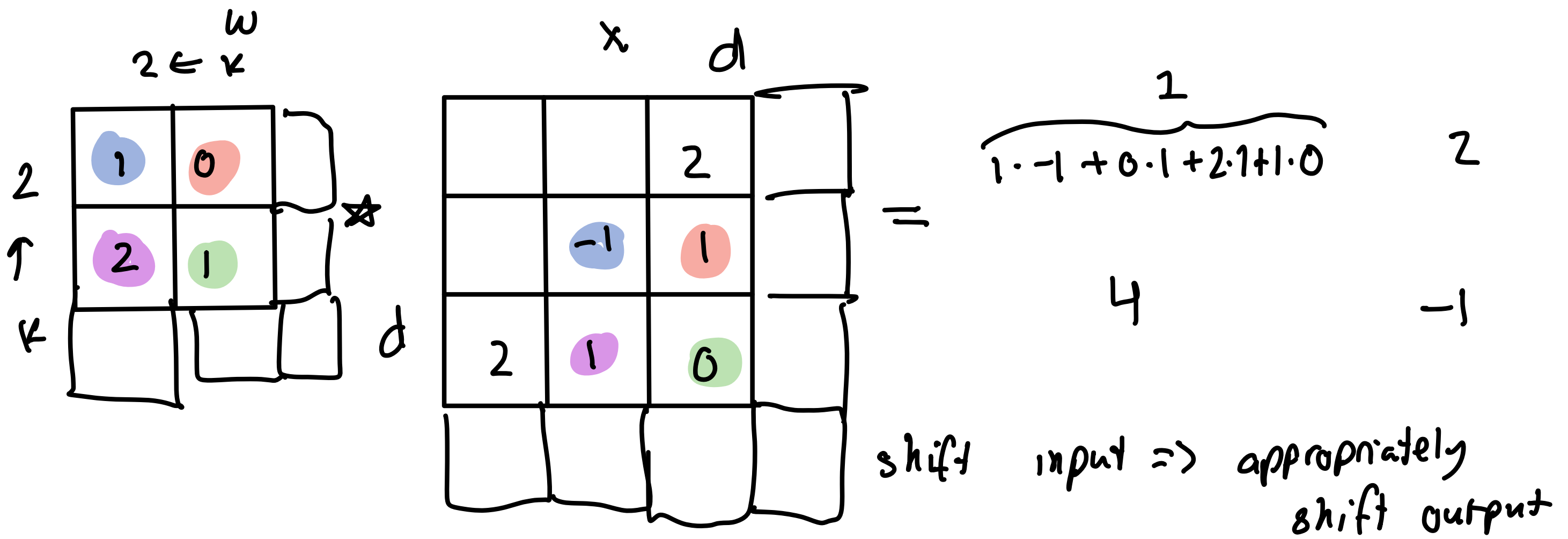
0



$x \star w$

$$(X \star w)[\Delta, t] = \sum_{\sigma} \sum_{\tau} x[s-\sigma, t-\tau] w[\sigma, \tau]$$

$\leftarrow \text{sigma}$ $d - k + 1$

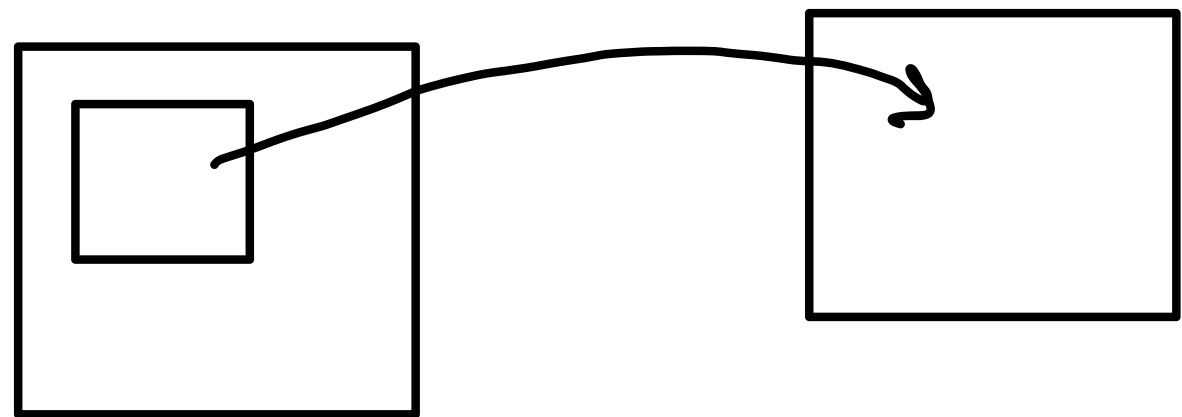


\hookrightarrow Locality

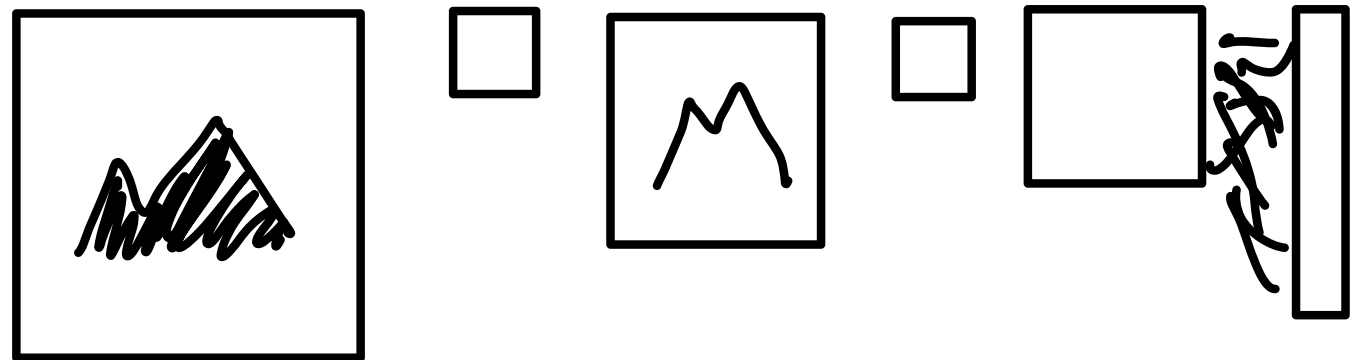
\hookrightarrow Shift Invariance

$[0, 1] \rightarrow [0, 9]$ divide by 9

$[0, 1] \rightarrow [-1, 1]$ $\frac{+1}{2}$



weights are values of filter

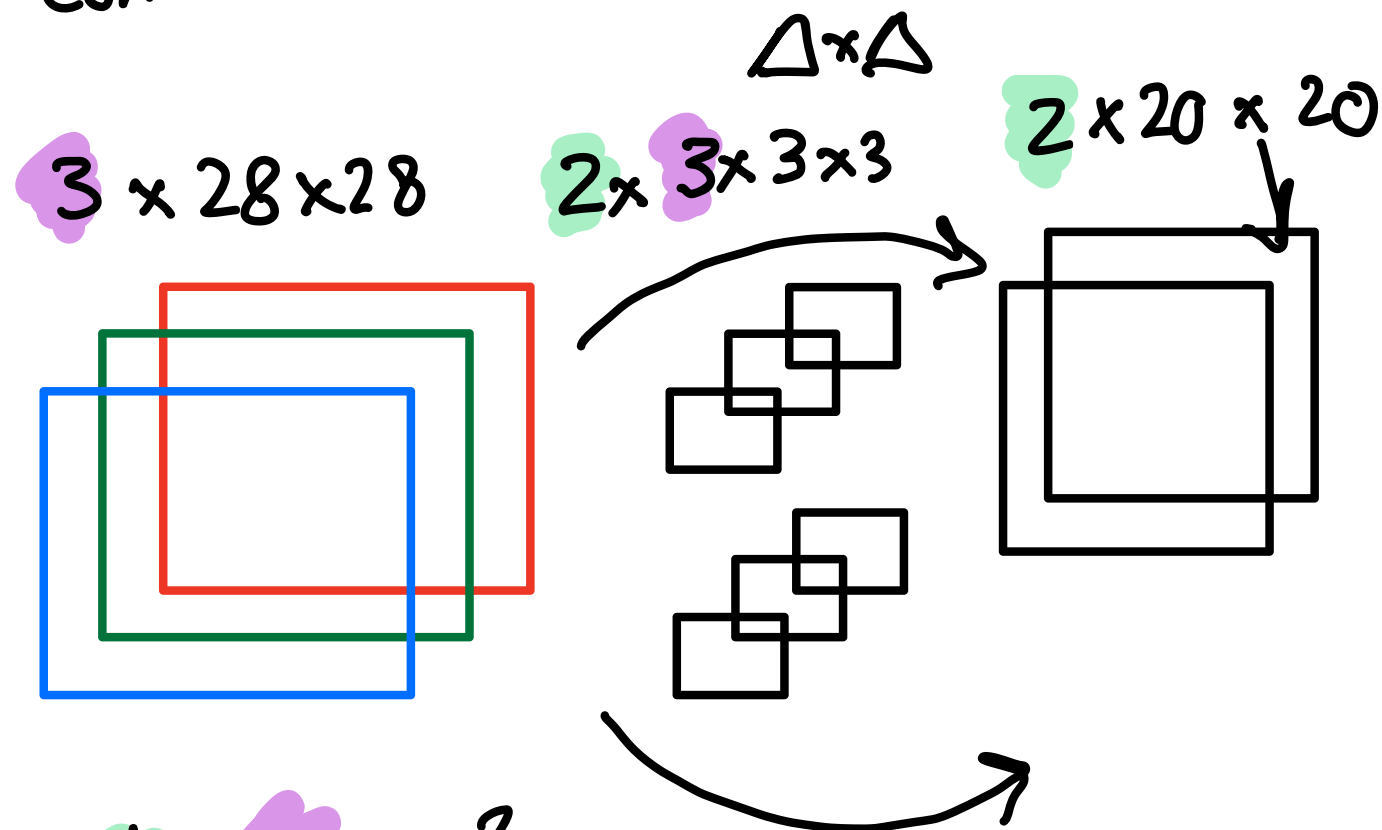


edges

objects

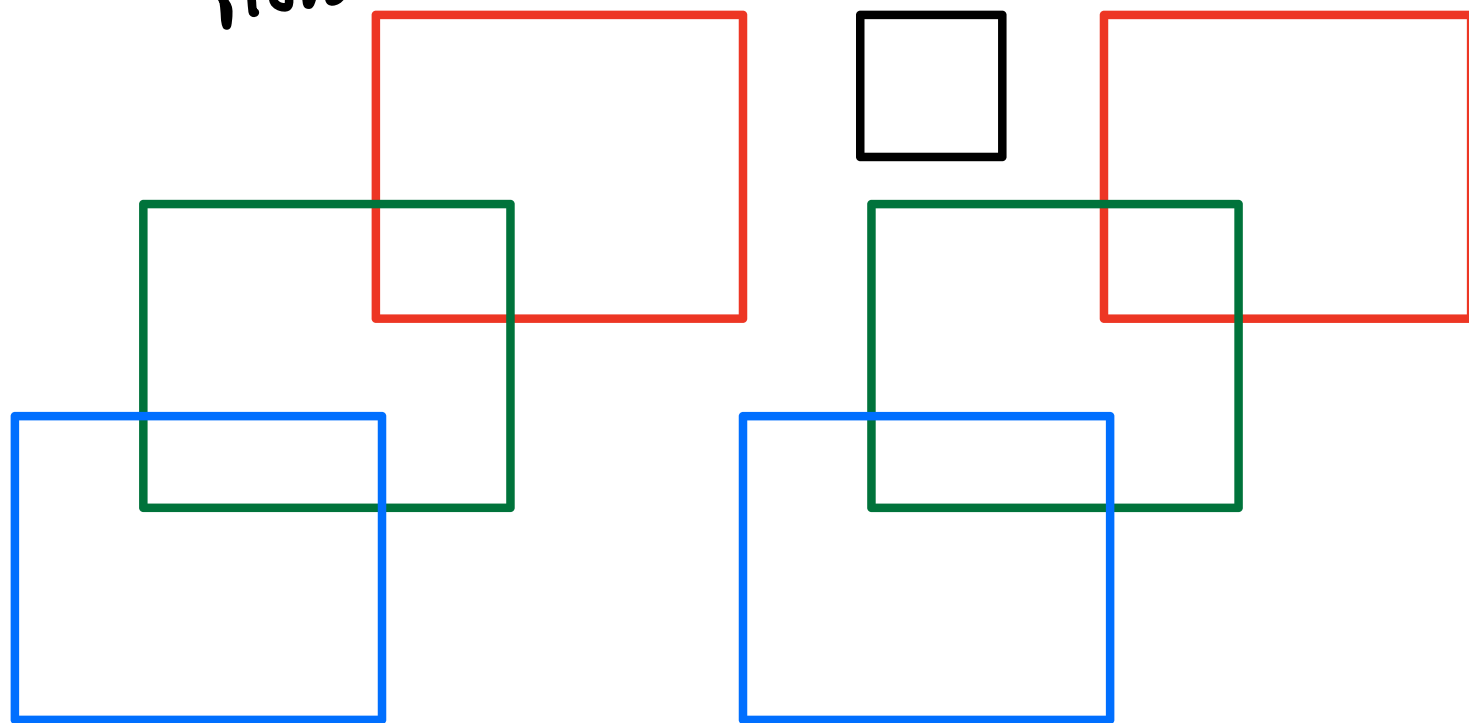
Increase abstract "understanding"

Convolutional neural network

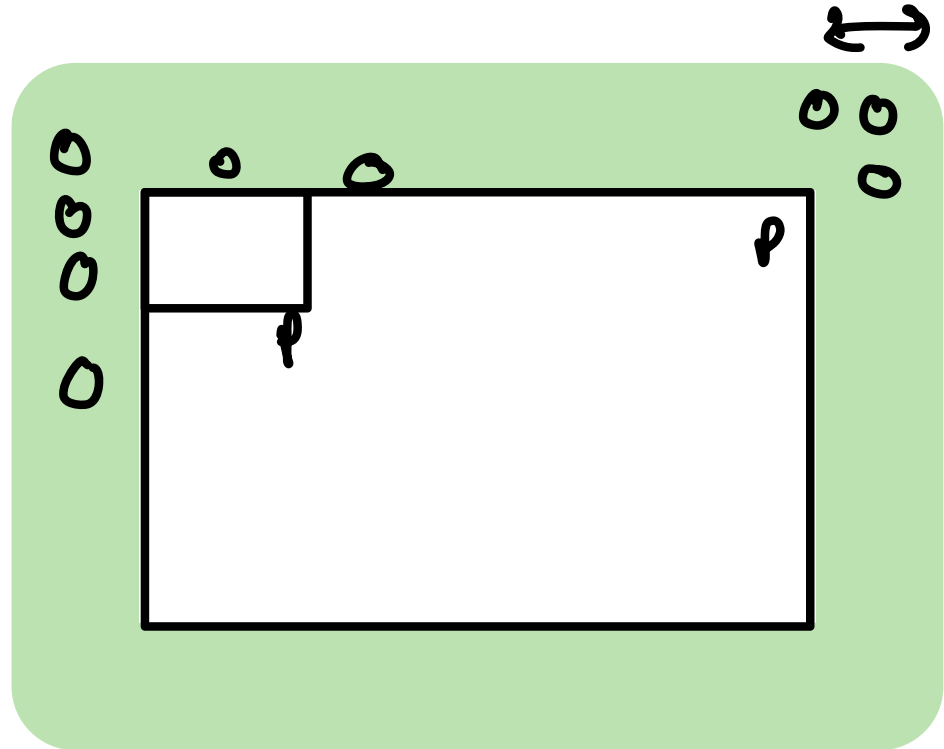


$ch_{out} ch_{in} \Delta^2$

homework

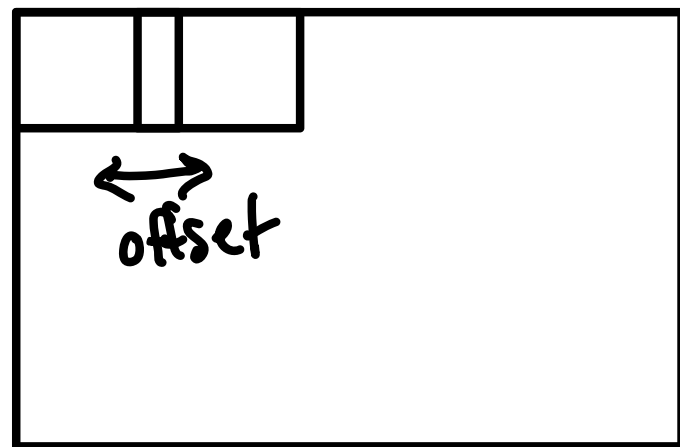


padding with zeros \uparrow pad \Rightarrow \uparrow output

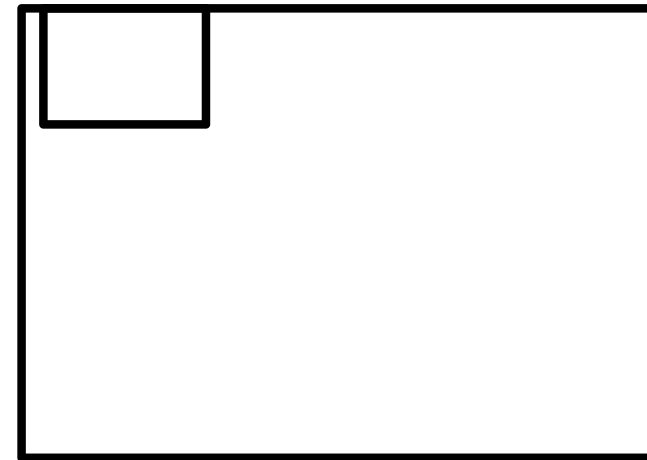


Stride

\uparrow stride \Rightarrow \downarrow output



Pooling



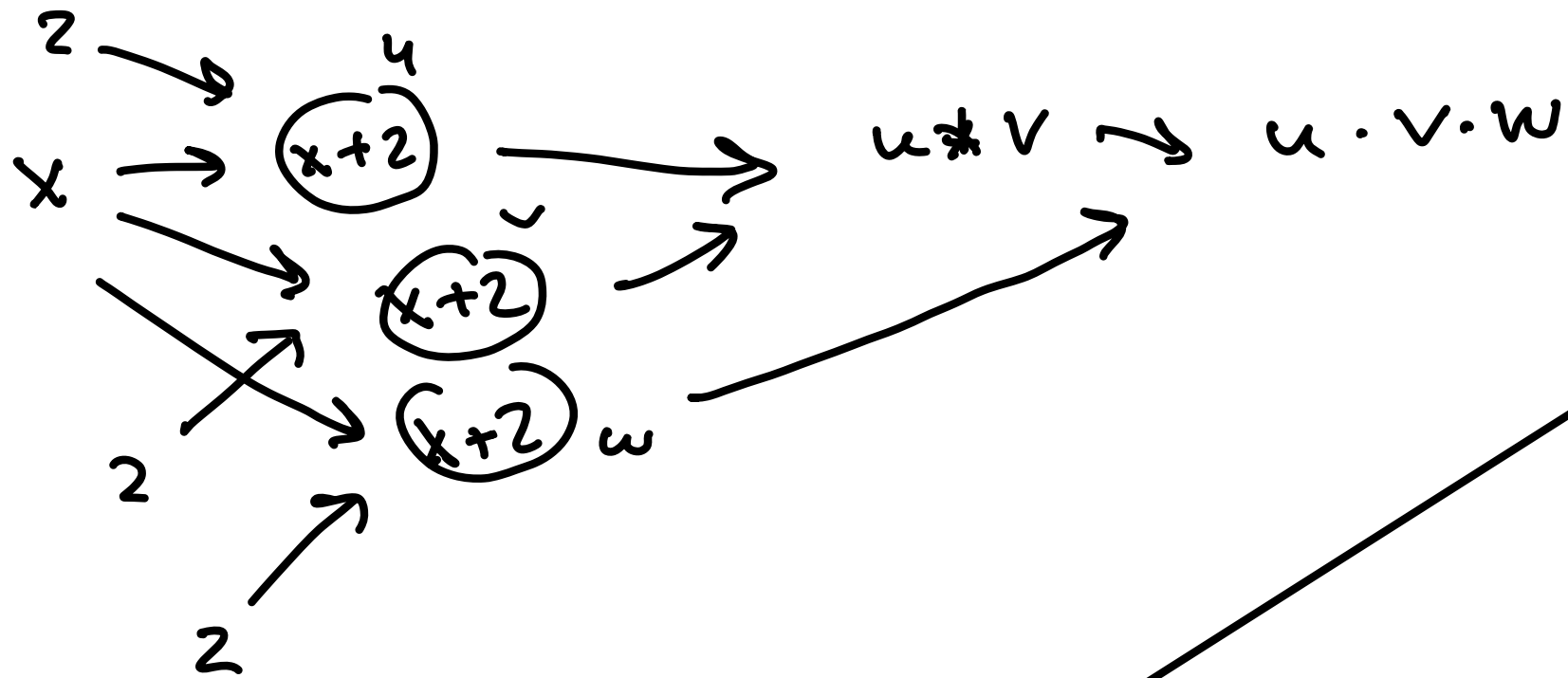
instead of sum,

- max
- min
- average

\hookrightarrow robust

Training ☺

$$(x+2)^* (x+2)^* (x+2)$$



Residual Network

$$h_l \quad h_{l+1} = \sigma(W_l h_l) + h_l$$

