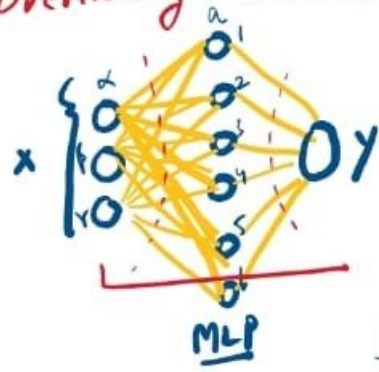


#CNNs y Project plan ✓

29 December 2024 17:42
Overview of last Lecture



$$a = \text{ReLU}[WX + b]$$

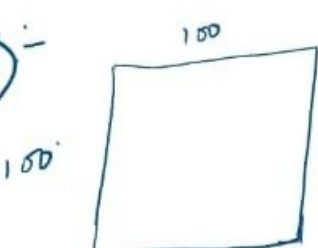
$$= \text{ReLU} \left[\begin{bmatrix} w_1^x & w_1^y & w_1^z \\ w_2^x & w_2^y & w_2^z \\ w_3^x & w_3^y & w_3^z \\ w_4^x & w_4^y & w_4^z \end{bmatrix} \begin{bmatrix} x^x \\ x^y \\ x^z \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \right]$$

↳ Universal approx. theorem

then $Y = [w^1 \ w^2 \ \dots \ w^6] \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$

Approximate any f^n via "Piecewise Linear function"

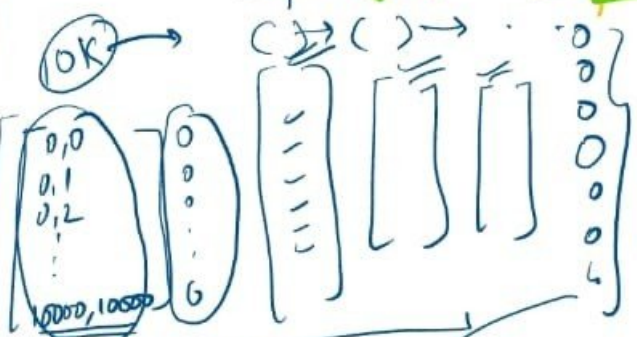
CNN



223 →



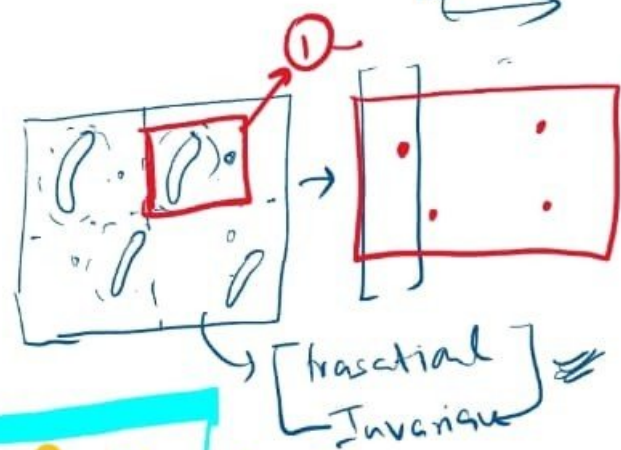
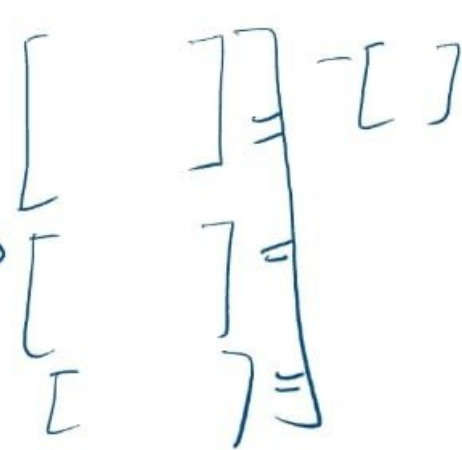
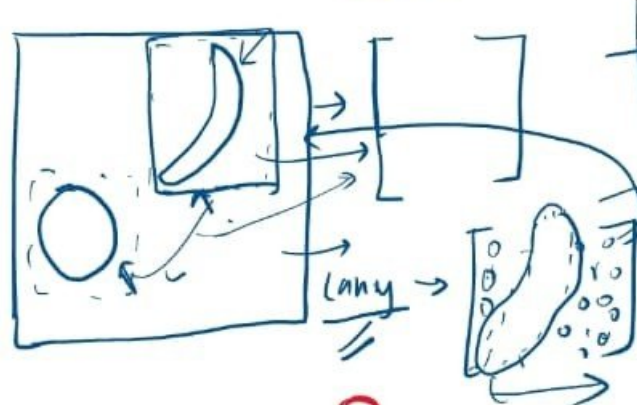
[1000 pixels]



Many weights

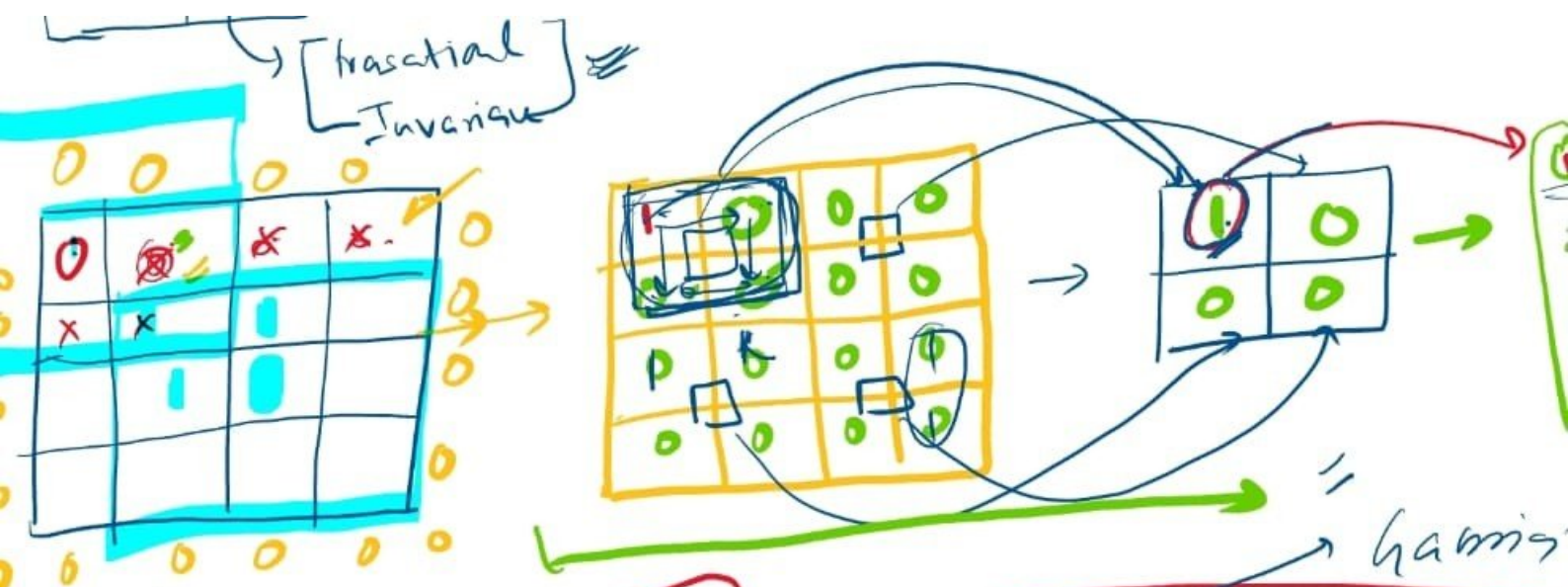
↳ computationally heavy

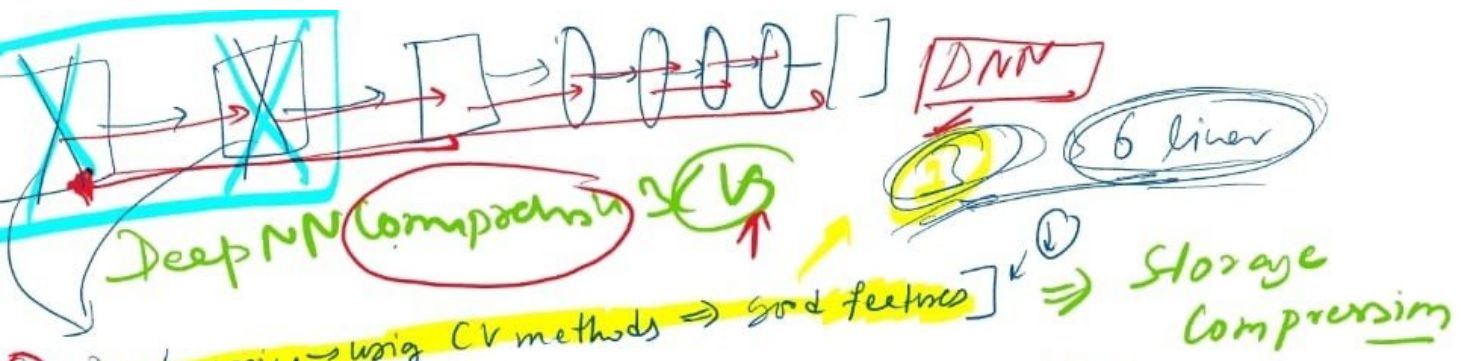
Softmax f^n



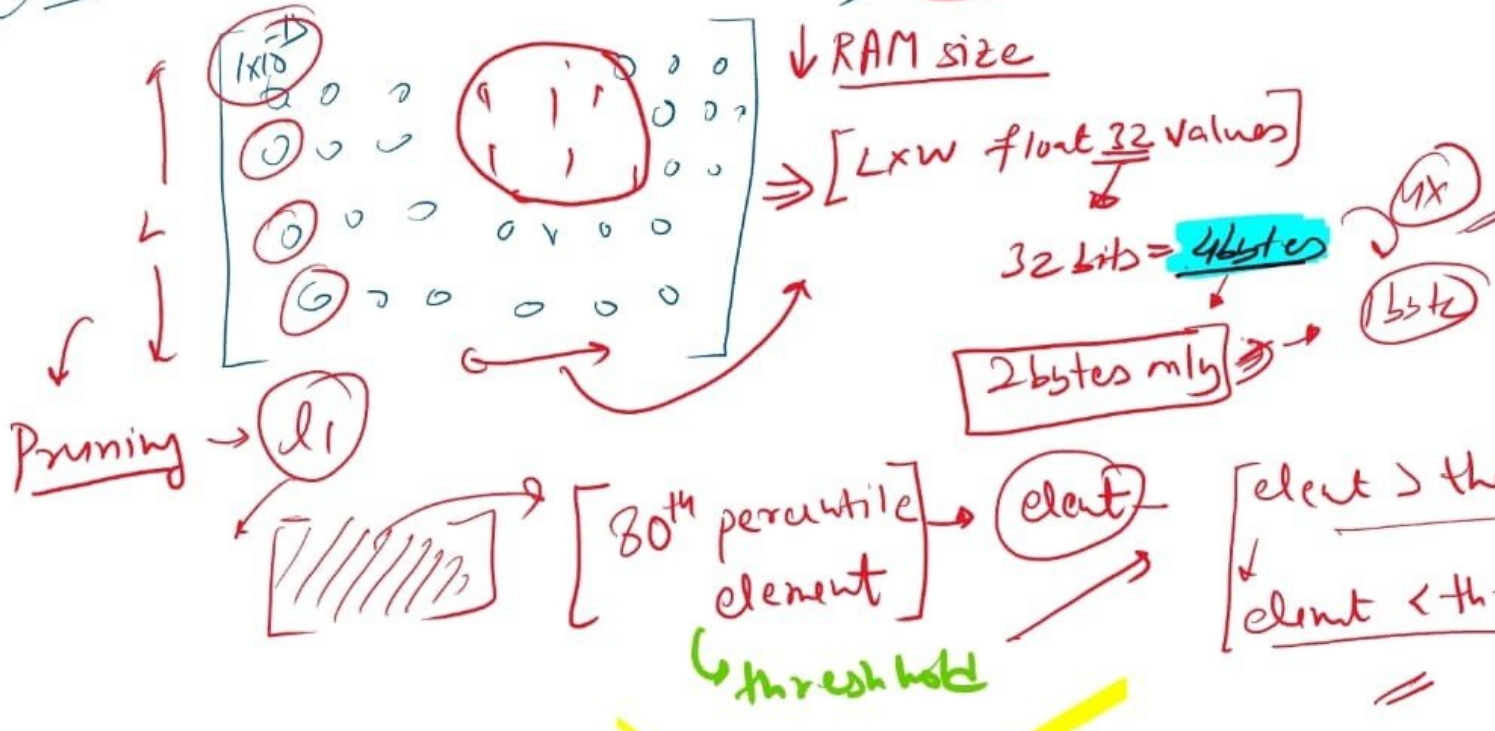
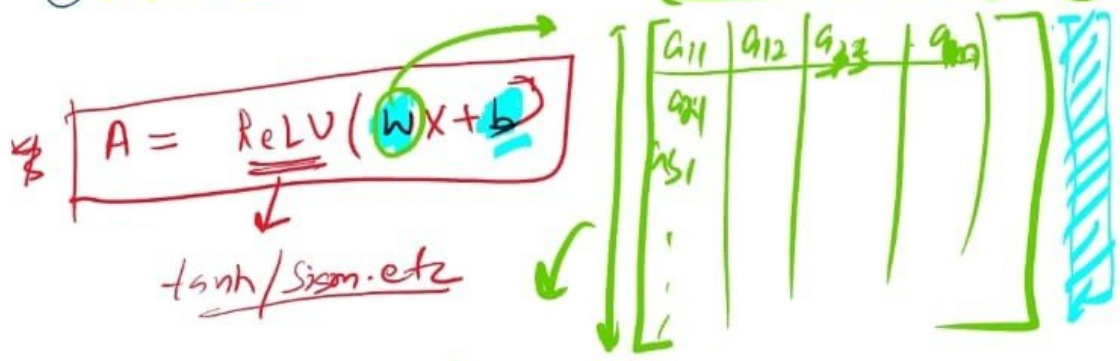
→ No of param ↓
→ CNN → transitional Invention!

[translational Invariant]





- ① Pre processing \rightarrow using CV methods \Rightarrow good features \Rightarrow Storage Compression
- ② Runtime Compression \rightarrow load \rightarrow GPU/CPU \rightarrow RAM
- ③ \rightarrow ① Pruning \Rightarrow k-means



quantized

01	1.2	0.1	0.1
02	5.1	8.2	1.7
61	2.9	9.9	1.2
3.8	4.5	3.9	4.8

float(32) → 8 bits / 4 bits

81.69

2.55

3.41

5.6

9.2

which d

01	0	0
4	5	1
6	8	...

0.00

5.6

5.6

1	1.69
2	2.55
3	3.41
4	5.6
5	9.2

5 diff

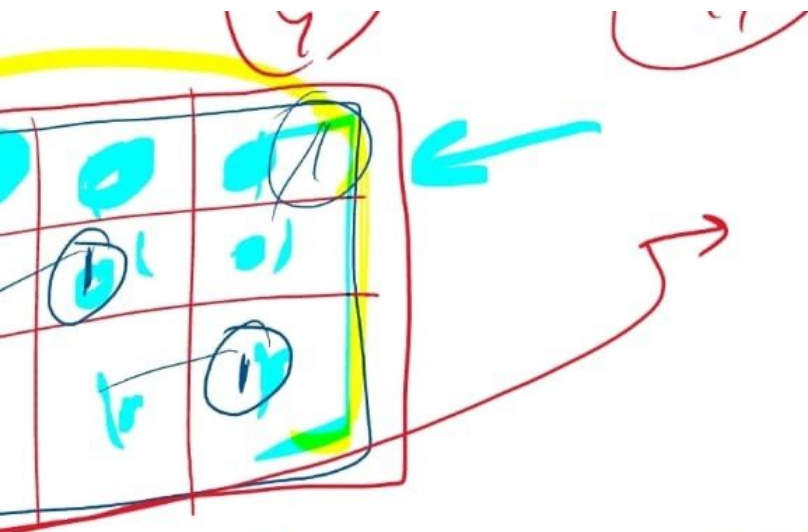
$n \times n$

$$\frac{(n \times n) \times 8 + n \times 9 \times 32}{(n \times n) \times 32} = 0/0 \rightarrow \left(\frac{1}{4} \right)$$

Hash table

$\left(\frac{1}{4} \right) \rightarrow$

4x



$$\frac{df}{dt} = \left(\frac{\partial f}{\partial z} \times \frac{dz}{dt} \right)$$

$$\frac{\partial f}{\partial z} = \left(\frac{\partial f}{\partial z} \right)$$

$$\frac{dz}{dt} = \left(\frac{dz}{dt} \right)$$

assignment
take

→ ① find values for the cluster ~~is~~ also

→ ② 1st cluster
Add their gradient] → 1

→ new gradient = $\sum ()$
for mem-1

RAM & SIZE

Guys small clarification, sorry I forgot why there is $dx/dt = 1$ is there in the chain rule because the mean stored is the average of all the values inside a cluster.....so

Mean = addition (value/N)

so

$d(\text{value})/d(\text{mean}) = 1/N$ which is a constant and will be incorporated in the learning rate

Sorry I said that it is intuition, it is logically sound now

18:47

