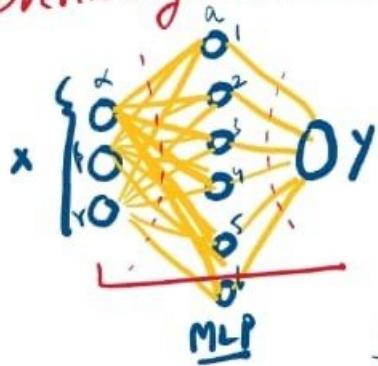


# \*CNNs & Project plan ✓

10 December 2024 17.3 Overview of last lecture



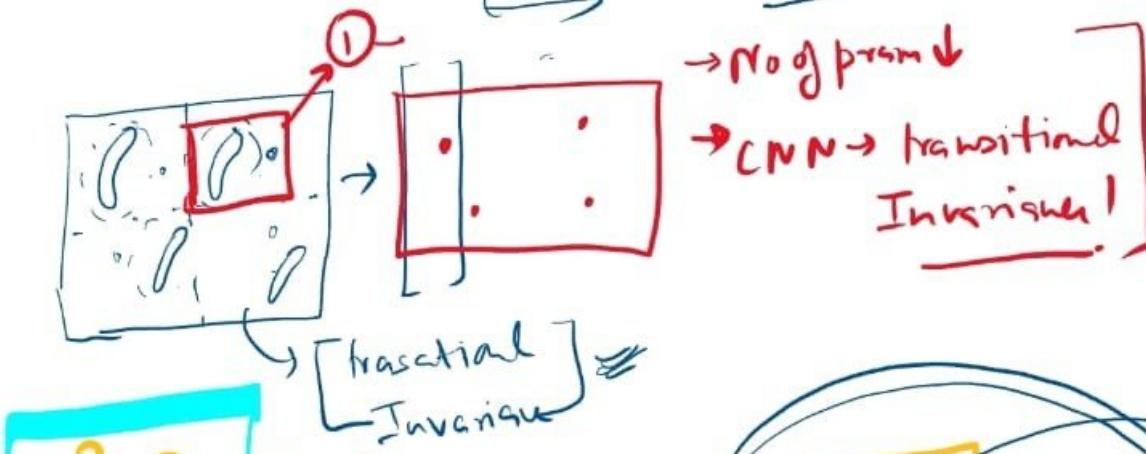
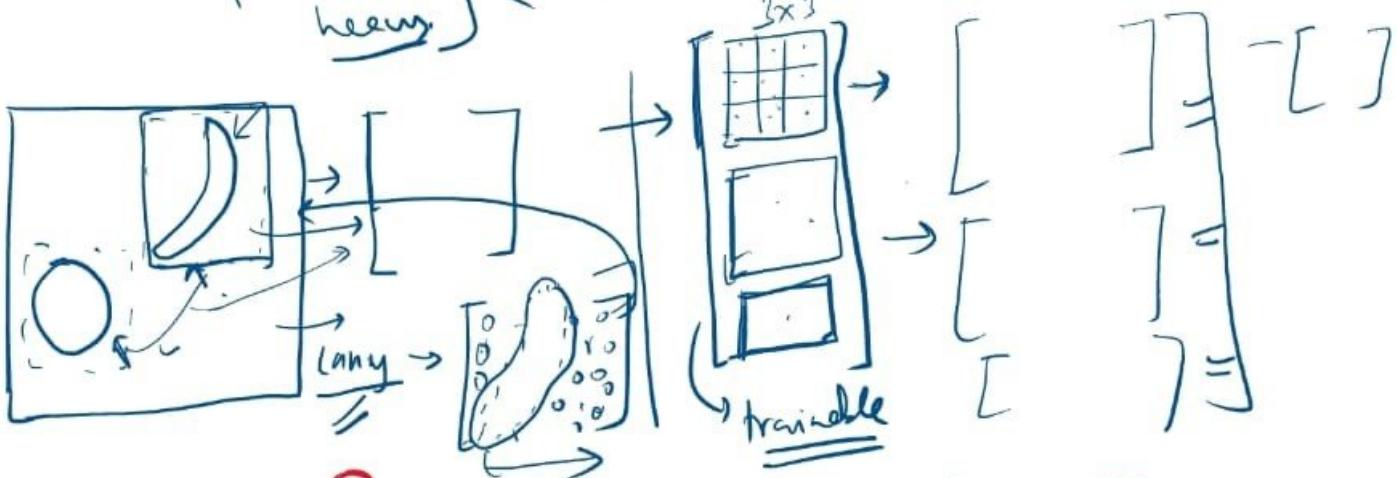
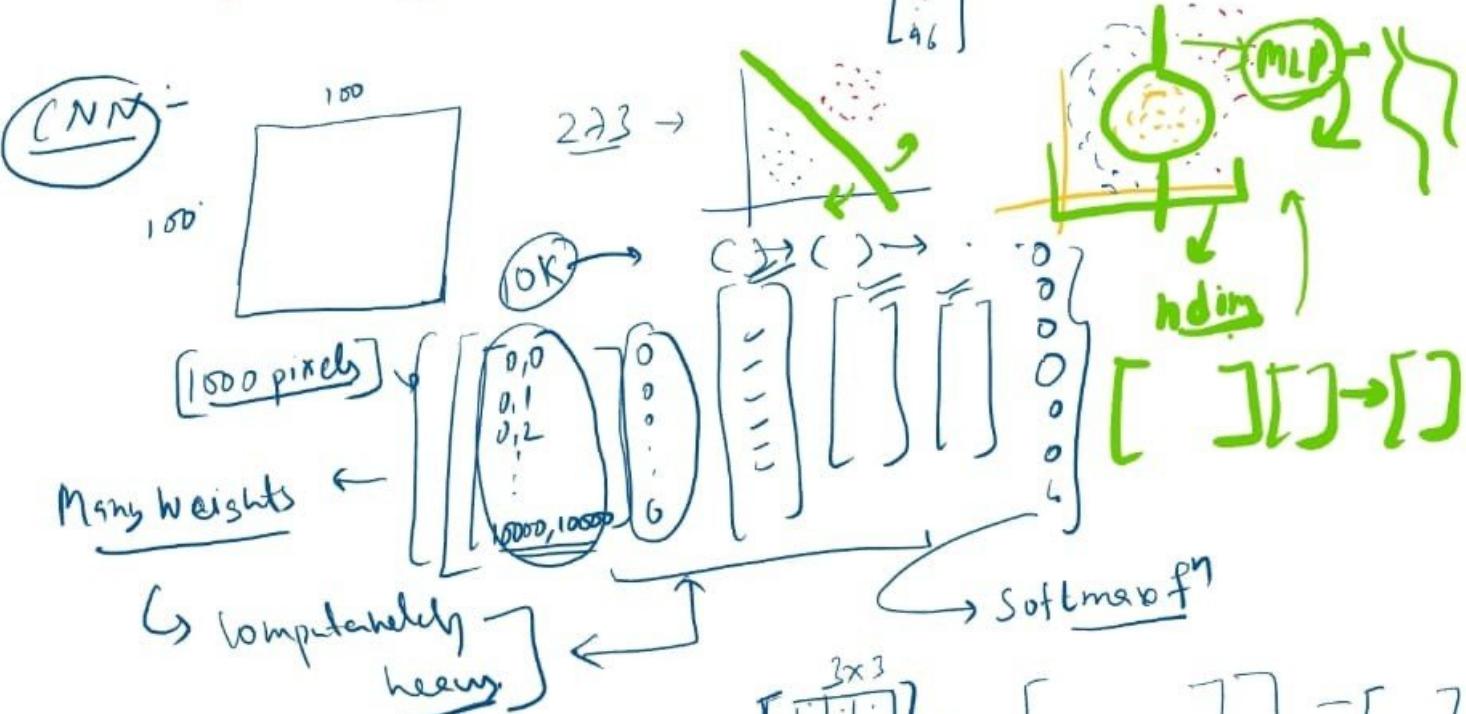
$$a = \text{ReLU}[Wx + b]$$

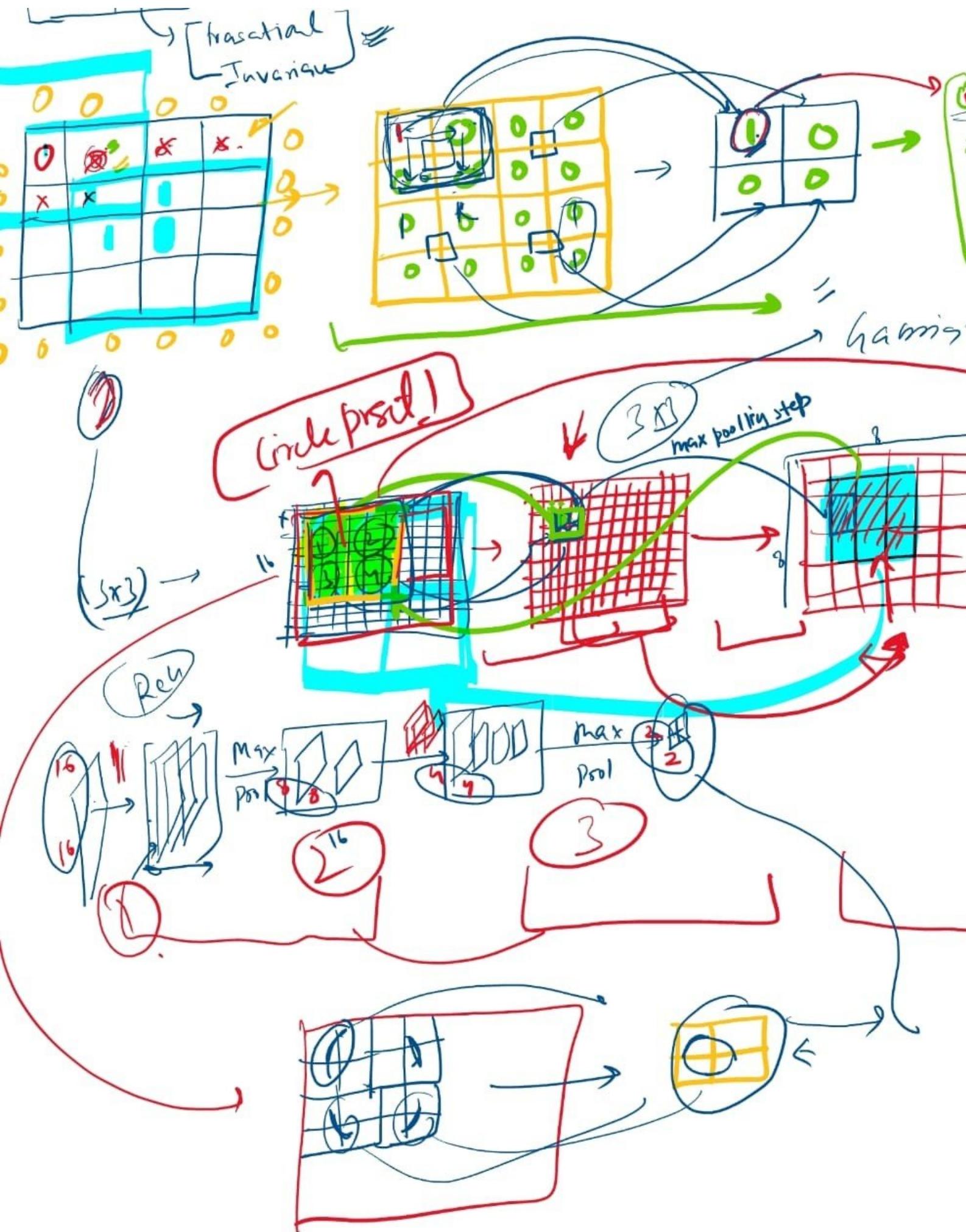
$$= \text{ReLU} \left[ \begin{bmatrix} w_1^1 & w_1^2 & w_1^3 \\ w_2^1 & w_2^2 & w_2^3 \\ w_3^1 & w_3^2 & w_3^3 \end{bmatrix} \begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \right]$$

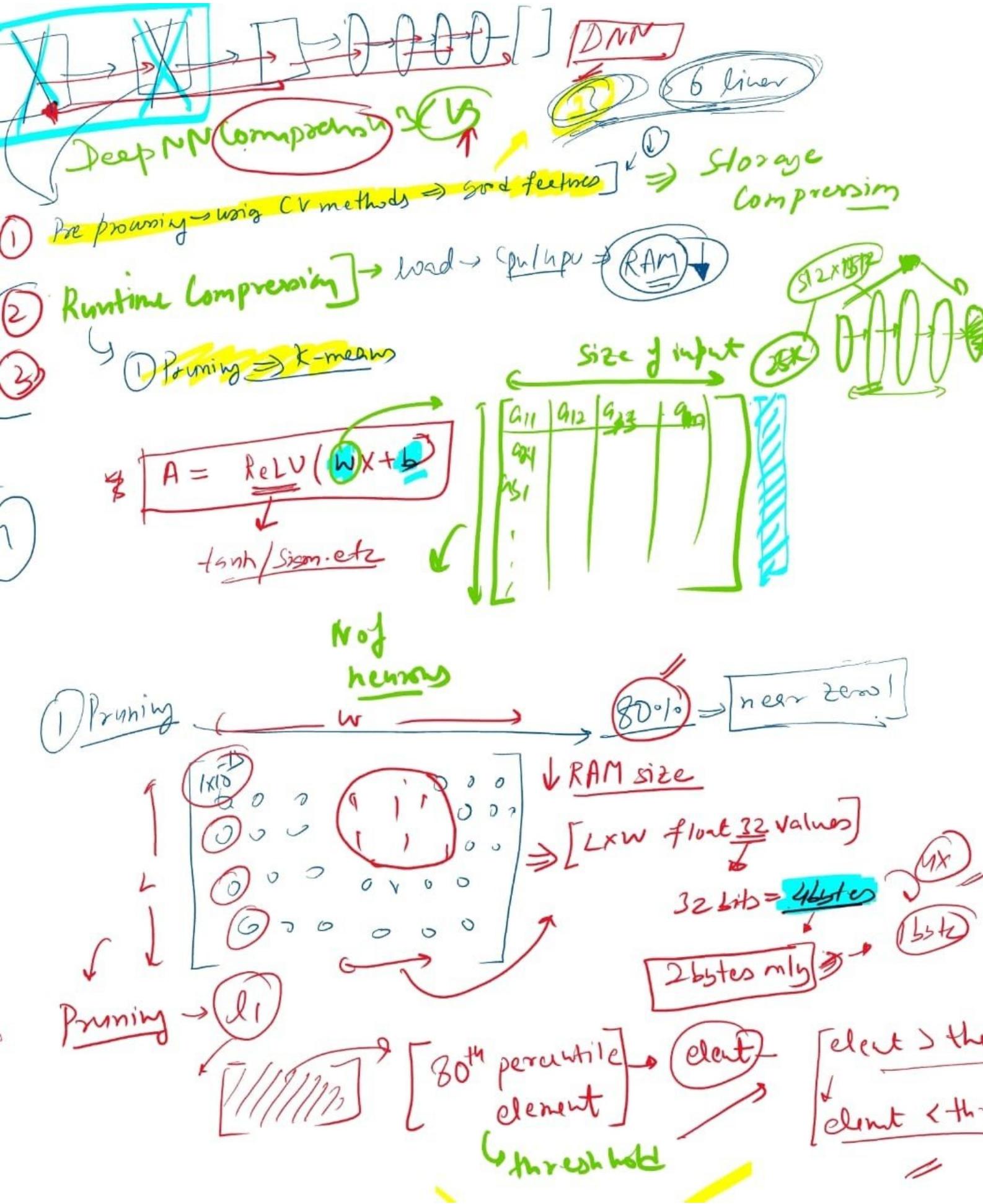
→ Approximate any f via  
"Piecewise Linear function"

↳ Universal  
approx. theory

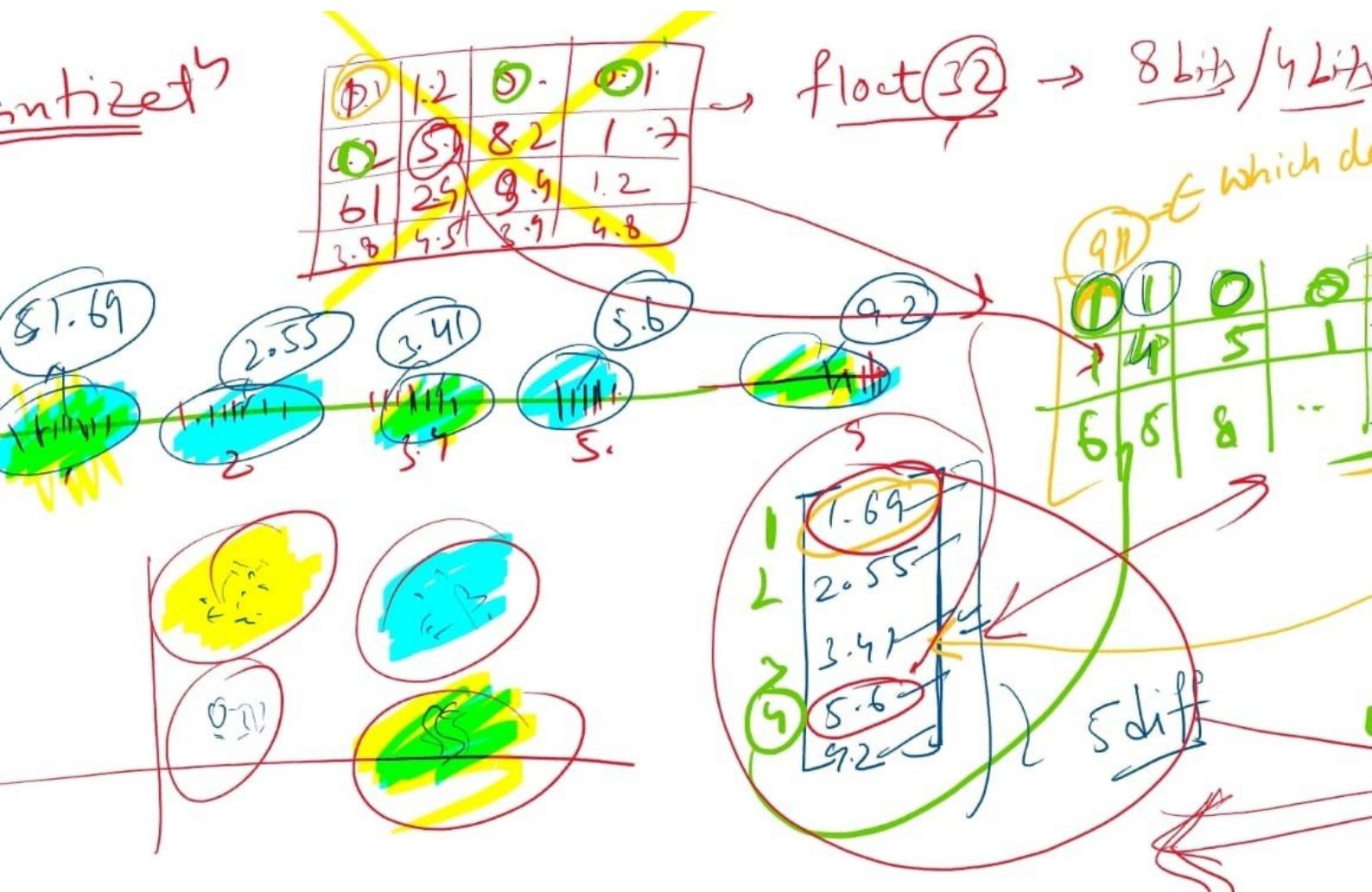
$$\underline{y} \text{ then } Y = [w^1 \ w^2 \dots w^6] \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix}$$







intized

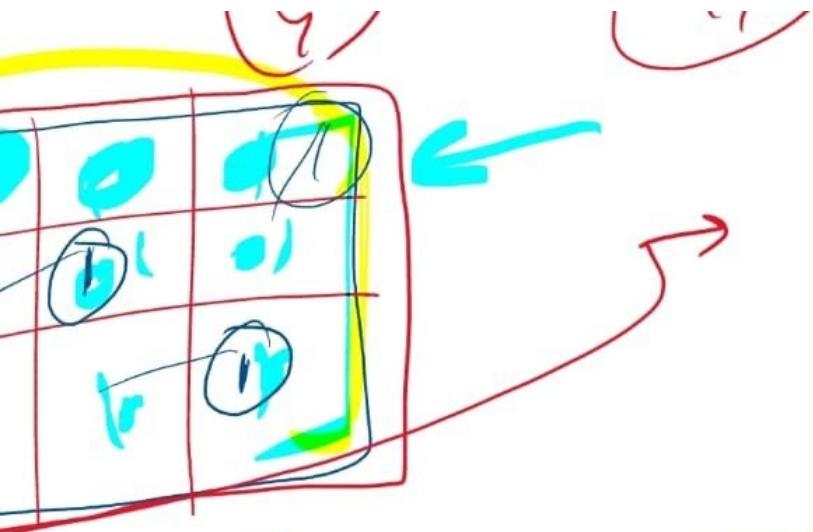


$n \times n$

$$\frac{(n \times n) \times 8 + n \times 4}{(n \times n) \times 32} = 0.075$$

Hash table

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



$$\frac{df}{dt} = \left( \frac{\partial f}{\partial x} \times \frac{dx}{dt} \right)$$

Diagram illustrating backpropagation:

$\frac{\partial f}{\partial y}$

$\frac{\partial y}{\partial t}$

$\frac{\partial f}{\partial z}$

gradient  
tells

- ① find values for the client  $\rightarrow$    
 1st cluster
- ② Add their gradients  $\rightarrow$  1

new gradient  
for mem-1  $\Rightarrow \sum C$

RAM  $\propto$  [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