

29/10/18

Lec-21

- Multi-class classification
 - One vs rest

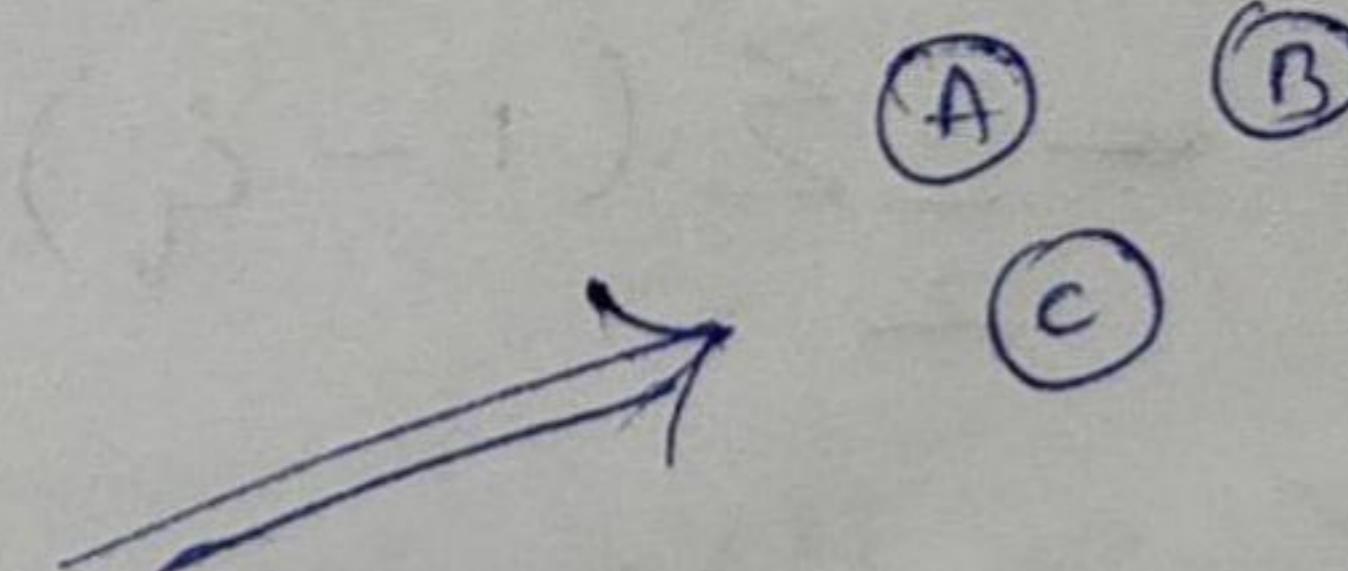
Eg MNIST

- 0 vs 9 else
- 1 vs else
- .
- .
- .

↓ If one of them says 1, then
no problem

↓ If 2 say, use some measure confidence

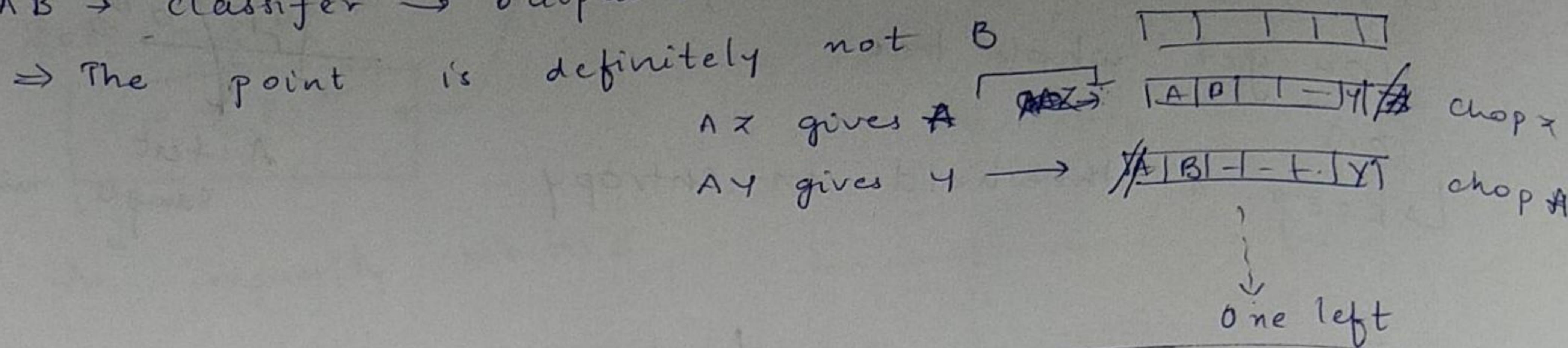
- Pairwise-classifiers:-
 - n_{C_2} classifiers
 - voting
 - Majority



AB → A
AC → C
BC → C }
Winner = c

Tree - Based

$A \cup B \rightarrow$ classifier \rightarrow output A



Current story

Classifiers:

Many are there

can inputs be combined to give output?

$$h_1(x), h_2(x), h_3(x)$$

Decision tree

Bayesian Multi SVM 90% 98%

One soln

\Rightarrow Find the best classifier, by finding error-rate
and choose the best

Can we find anything good out of the not so
good classifiers ???

• Multiple classifiers \rightarrow find the
best answer

① Majority Vote

② \hookrightarrow can cause ties too

eg: 3 \downarrow $\boxed{A \rightarrow B}$ { so tie

② Weighted voting

y_k // class output by each classifier

$$y_{\text{final}} = f(y_i, \epsilon_i)$$

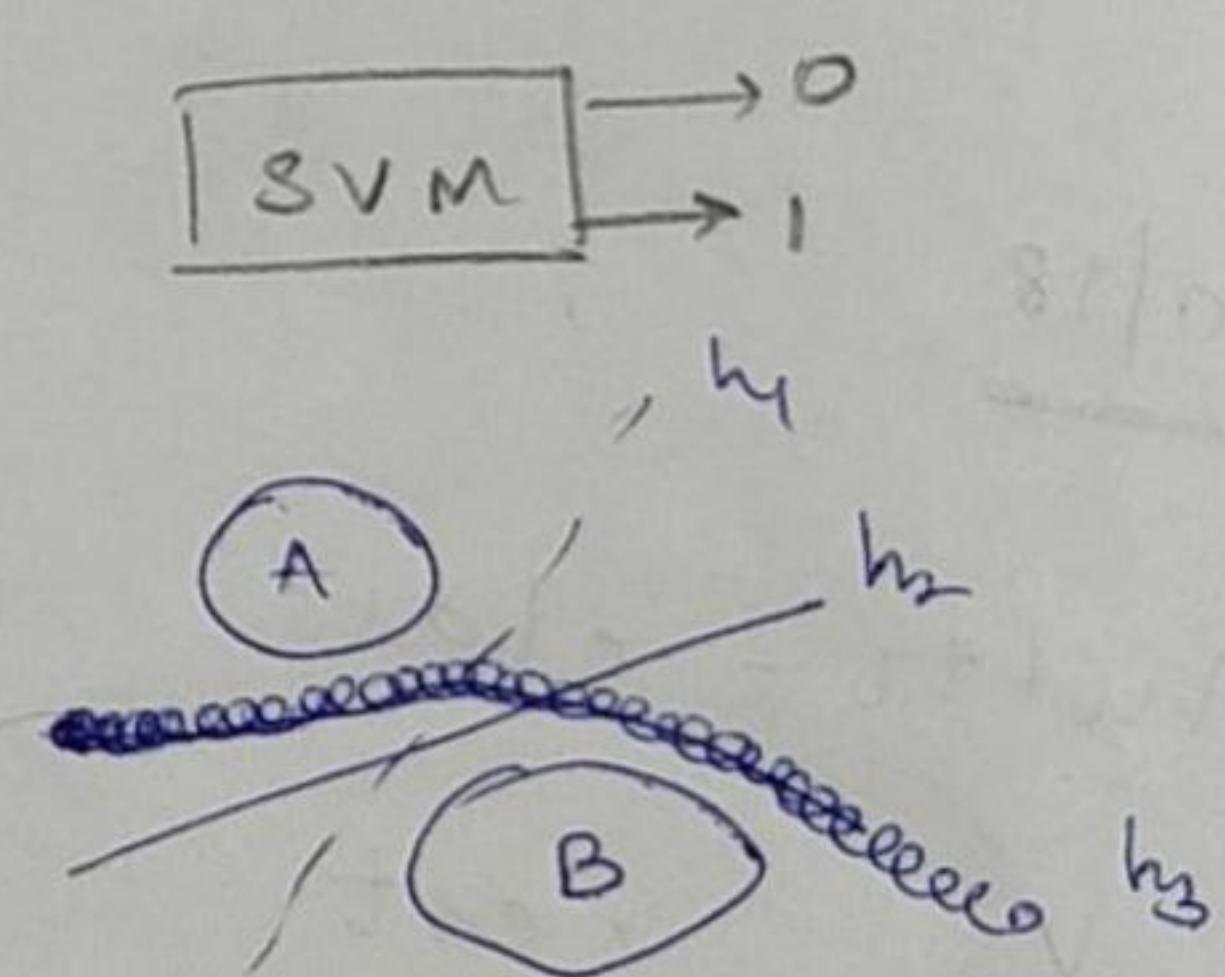
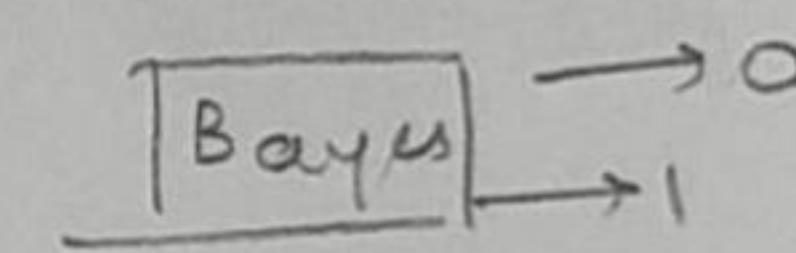
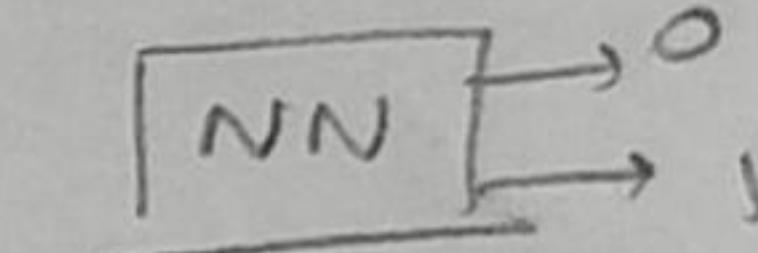
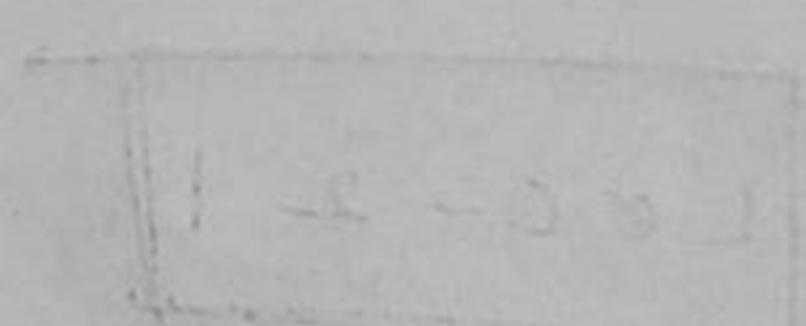
$$= \begin{cases} \frac{\sum (1 - \epsilon_i) y_i}{\sum (1 - \epsilon_i)} > 0.5 \Rightarrow \text{class } 1 \\ < 0.5 \Rightarrow \text{class } 0 \end{cases}$$

$$\sum (1 - \epsilon_i) y_i - \frac{1}{2} \sum (1 - \epsilon_i) > 0 \Rightarrow c1$$

$< 0 \quad c2$

$$= \frac{\text{floor}(\quad) + 1}{\text{round}(\quad)}$$

a round fn instead of if.



• Errors that h_1 makes \neq errors that h_2 makes, (even tho' $E_1 > E_2$)
Because, they ~~can~~ make different types of errors //

• Sample specific confidence measure:

NN - directly give

$$y_k \rightarrow \frac{c_k}{\sum c_k}$$

$$\frac{y_k c_k}{\sum c_k}$$

② Feature fusion:-

Assuming the classifiers use different features
find a combined vec and classify --

features
positions

No use //
of this

We can learn the fn $f \rightarrow$ called

STACKING

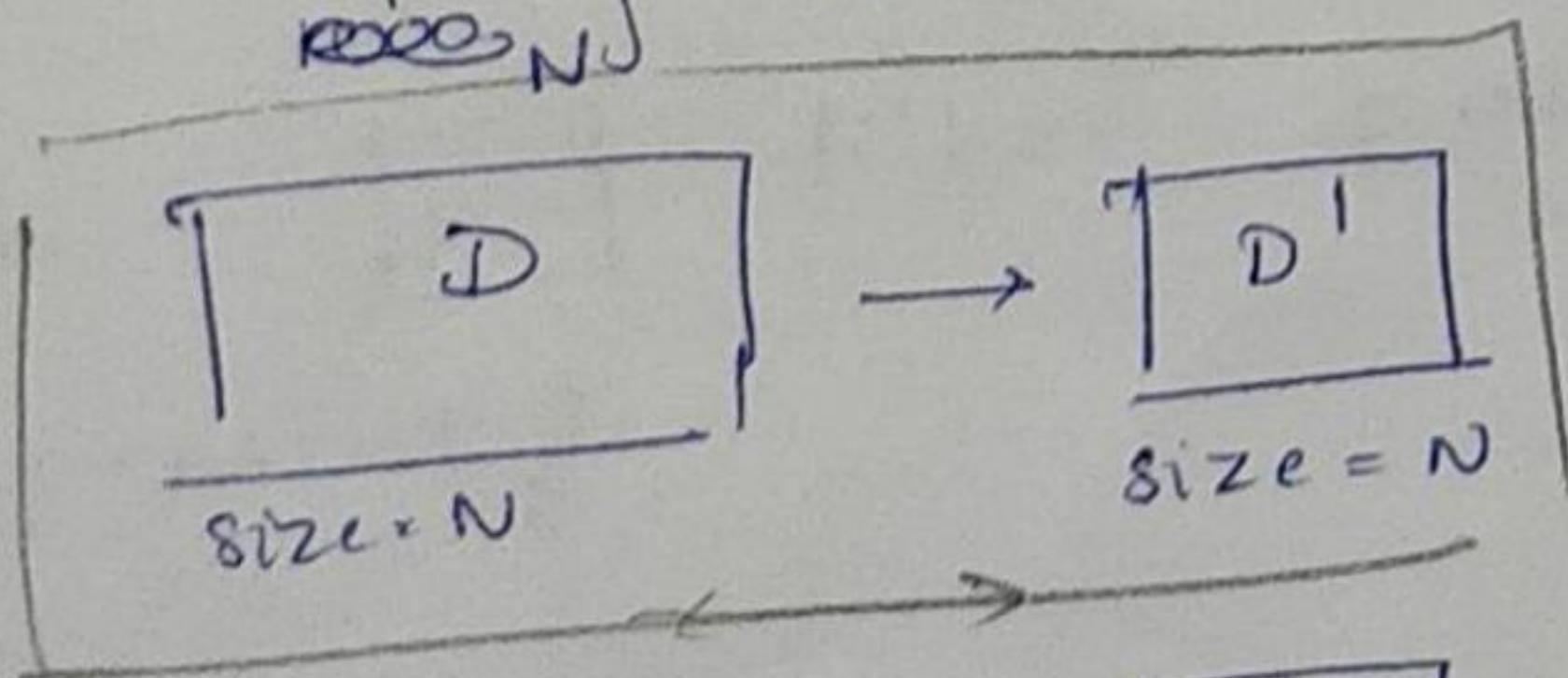
Tweaking classifier params for better advantage

Strong \rightarrow ~90% accuracy
weak \rightarrow ~50% accuracy
all classifiers
If $\frac{1}{2} \rightarrow$ everything, no improv
If $\frac{1}{2} + \delta \rightarrow$ Improvement

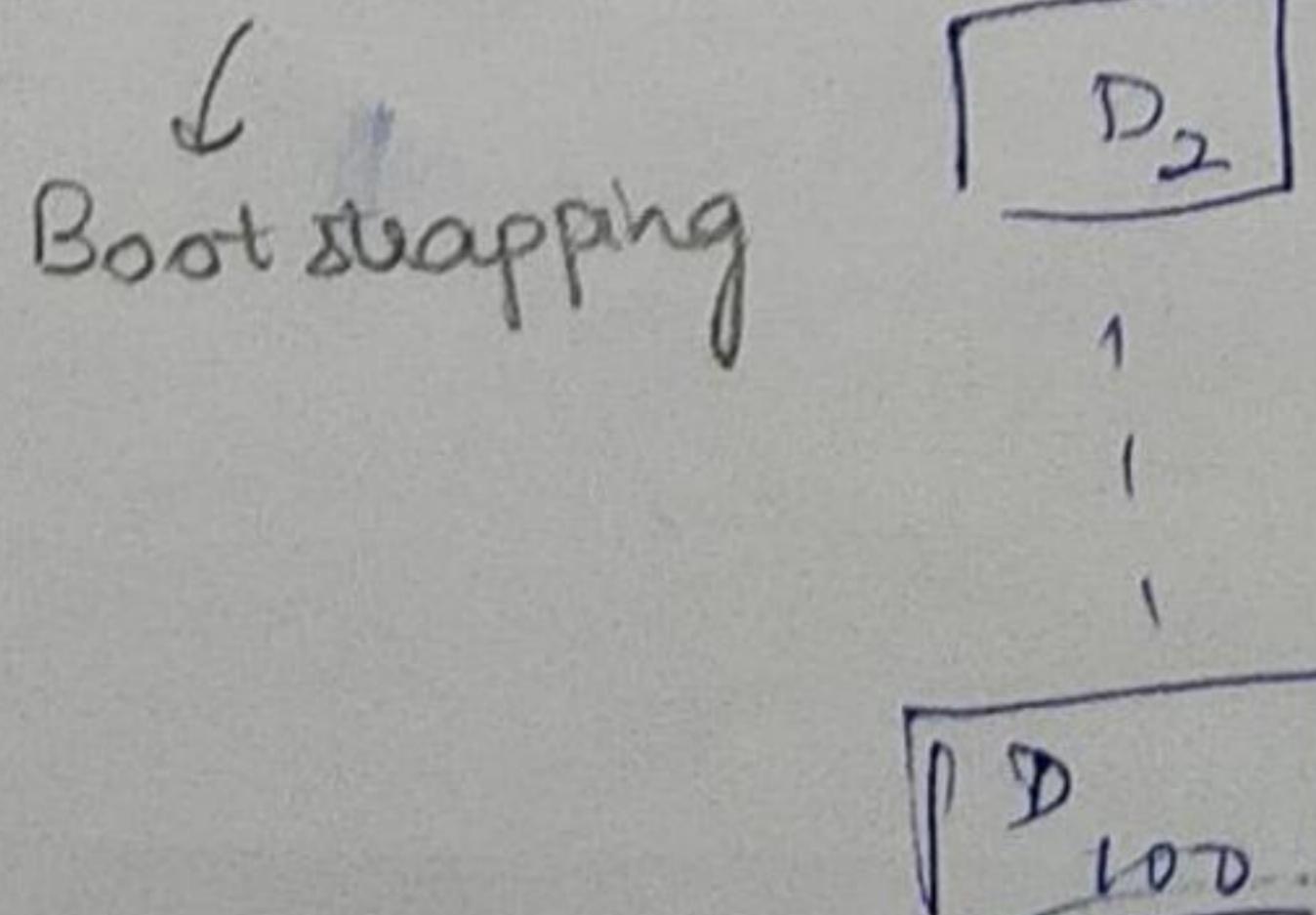
Ensemble methods

↳ Bootstrap aggregation

• Sample $\xrightarrow{\text{size } N} 100$ times \rightarrow with replacement



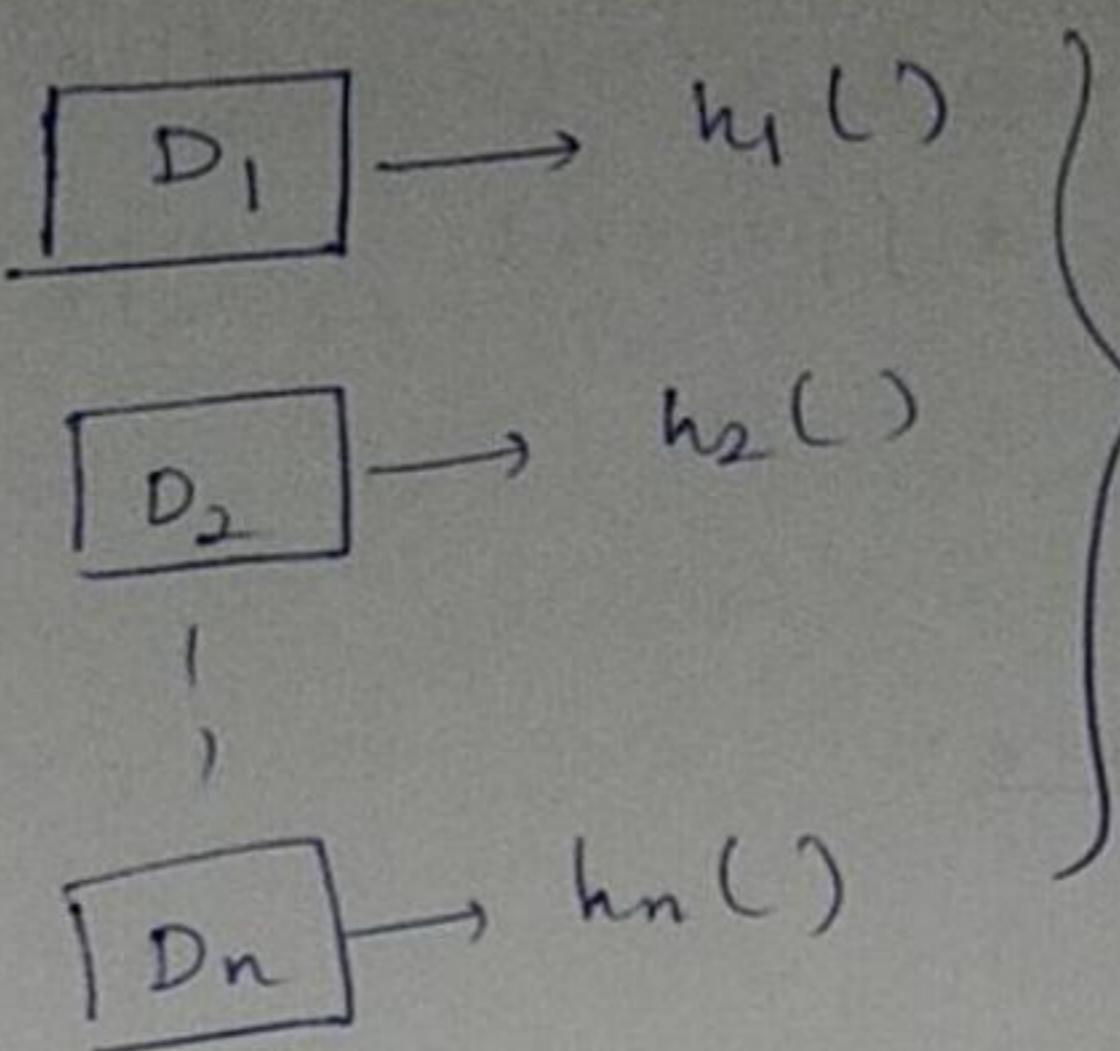
Some elems $\in D'$
and repeat
Some are not in D' which belong to D .



[It was found that when $N' < N \Rightarrow$ very high variability in D_1, D_2, D_3, \dots]

and

$N' \gg N \Rightarrow$ very $D' \approx D$
(100N) distros //.



Parallelizable
on different
computers.

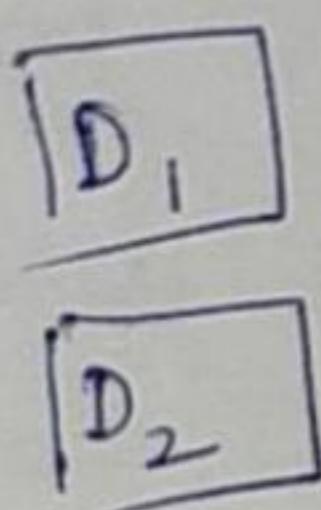
all h_i 's are generally
same.

→ all this is to
create variation
in data.

→ Boosting:
Non-uniform sampling

$s_i(x)$
 $D_i \xrightarrow{\frac{1}{n}}$ for all x_i
 $s_i(x) \rightarrow$ associate a weight to each x_i

$$s_i(x) = \begin{cases} \frac{1}{n} & \rightarrow \text{correctly classified} \\ e^{-\alpha_i} & \\ e^{\alpha_i} & \rightarrow \text{incorrectly classified} \end{cases}$$



Ada(ptive)-
Boost //

→ Badly classified probs [An expert]
Highly specialized in a small region

Now, each classifier is given weight
If $\epsilon_i = 0$ (D_{100})

$$\frac{1 - \epsilon_i}{\epsilon_i}$$

$$\frac{1 - 0}{0} \sim \infty$$

$$\log(\infty) \rightarrow \text{high wt}$$

$$y_{\text{out}} = \sum \alpha_t h_t(x)$$

$$h_t(x) = \pm 1, \text{ not } (0, 1)$$

If $y_{\text{out}} > 0 \rightarrow \text{class 1}$

$y_{\text{out}} < 0 \rightarrow \text{class 0}$

Cross entropy = softmax + NLL

- SimpleMLP II

, CNN

exp. Image info

$$28 \times 1$$

A hand-drawn diagram consisting of a vertical stack of approximately ten horizontal lines on the left side. From the right side, several lines radiate outwards towards the stack, creating a fan-like or beam-like effect.

each imp
case is lost!!

so spatial info is lost //

cNN operated over Volumes

The diagram illustrates a convolutional layer's input and a single filter. On the left, a vertical stack of three rectangular boxes represents the input volume. The top box is labeled "32x32". The middle box is labeled "3" and contains a small circle with a cross, indicating it is the active receptive field of a specific neuron. The bottom box is also labeled "32x32". To the right of the input is a single rectangular box representing a filter, labeled "6x5x3". An arrow points from the input volume to the text "Neurons * filters".

• Less number of params.

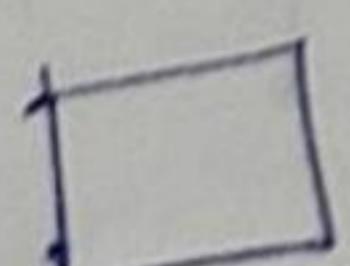
$$32 \times 32 \rightarrow \boxed{51}^5$$

$$4 - 2 + 1 = 3$$

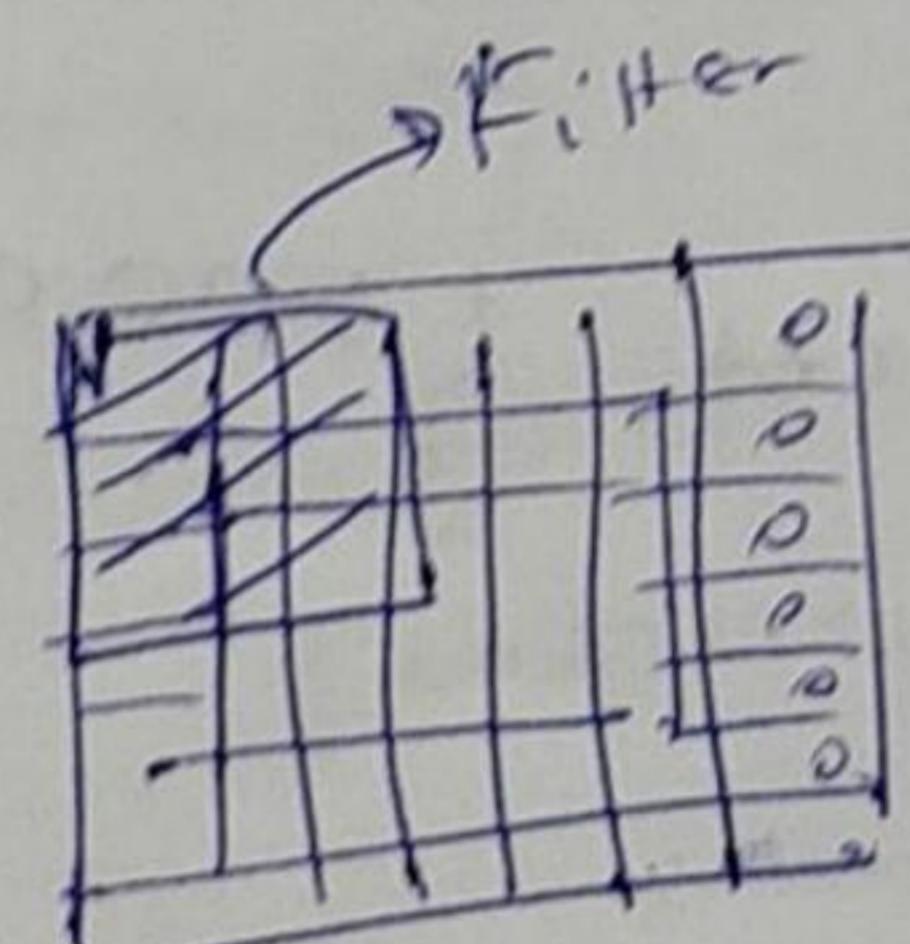
padding

is done

Others



is present //



it sees how much line