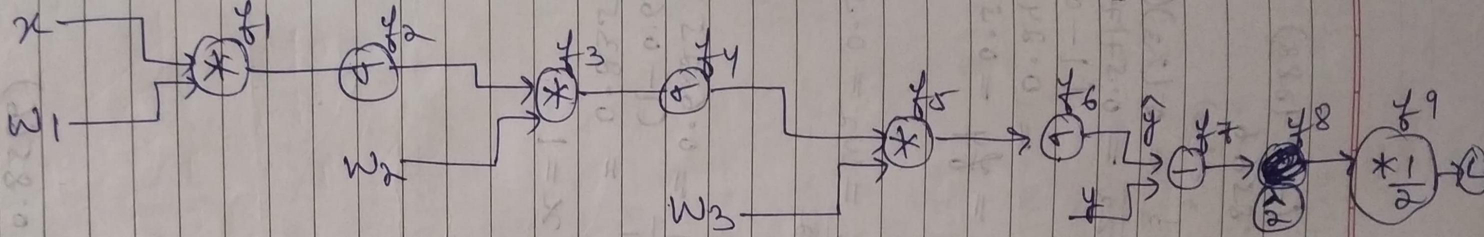
Q1 (a) Perform backpropagation (sigmoid function), given $x=1$, $w=0.5$, $y=1$ (true label)

(VANISHING GRADIENT)



1) Forward pass :-

$z_1 = f1 = x \times w_1 = 1 \times 0.5 = 0.5$

$a_1 = f2 = \sigma\, f1 = \sigma(z_1) = \frac{1}{1+e^{-0.5}} = 0.6225$

$\Rightarrow \frac{\partial f2}{\partial f1} = f1(1-f1)$

$z_2 = f3 = w_2 \cdot f2 = 0.5 \times 0.6225 = 0.3113$

$a_2 = f4 = \sigma(z_2) = \frac{1}{1+e^{-0.3113}} = 0.5772$

$z_3 = f5 = w_3 \cdot f4 = 0.5 \times 0.5772 = 0.2886$

$a_3 = f6 = \sigma(z_3) = \frac{1}{1+e^{-0.2886}} = 0.5717$

or $\hat{y}$

$f7 = a_3 - y = 0.5717 - 1 = -0.4283$

$f8 = (f7)^2 = (-0.4283)^2 = 0.1834$

$f9 = \frac{1}{2} \times f8 = \frac{0.1834}{2} = 0.0917$

or L

Page No.

Date

2) Find local derivatives :-

$$\frac{\partial f_9}{\partial f_8} = \frac{1}{2}, \quad \frac{\partial f_8}{\partial f_7} = 2f_7 = 2 \times (-0.4283) = -0.8566$$

$$\frac{\partial f_7}{\partial \underset{(f_6)}{a_3}} = 1, \quad \frac{\partial f_7}{\partial y} = 1, \quad \frac{\partial \overset{\sigma(z_3)}{a_3}}{\partial z_3} = \sigma(z_3)(1-\sigma(z_3)) = 0.5717 \times (1-0.5717) = 0.2448$$

$$\frac{\partial f_5}{\partial f_4} = w_3 = 0.5, \quad \frac{\partial f_5}{\partial w_3} = f_4 = 0.5772,$$

$$\frac{\partial \overset{(f_4)}{a_2}}{\partial \underset{(f_3)}{z_2}} = 0.5772 \times (1-0.5772) = 0.2440, \quad \frac{\partial f_3}{\partial f_2} = w_2 = 0.5,$$

$$\frac{\partial f_3}{\partial w_2} = f_2 = 0.6225, \quad \frac{\partial a_1}{\partial z_1} = 0.6225 \times (1-0.6225), = 0.2350.$$

$$\frac{\partial f_1}{\partial x} = w_1 = 0.5, \quad \frac{\partial f_1}{\partial w_1} = x = 1$$

3) Backpropagation —

$$\frac{\partial L}{\partial f_9} = 1$$

$$\frac{\partial L}{\partial f_8} = \underset{\text{(global)}}{1} \times \underset{\text{(local)}}{\frac{\partial f_9}{\partial f_8}} = \frac{1}{2} = 0.5$$

$$\frac{\partial L}{\partial f_7} = 0.5 \times \frac{\partial f_8}{\partial f_7} = 0.5 \times (-0.8566) = -0.4283$$

Page No.
Date

$$\frac{\partial L}{\partial f_6} = -0.4283 \times \cancel{\frac{\partial f_7}{\partial f_6}}$$
$$= -0.4283$$

$$\frac{\partial L}{\partial f_5} = -0.4283 \times 0.2448$$
$$= -0.105$$

$$\frac{\partial L}{\partial f_4} = -0.105 \times \frac{\partial f_5}{\partial f_4}$$
$$= -0.105 \times 0.5$$
$$= -0.0525$$

$$\frac{\partial L}{\partial f_3} = -0.0525 \times \frac{\partial f_4}{\partial f_3}$$
$$= -0.0525 \times 0.2440$$
$$= -0.0128$$

$$\frac{\partial L}{\partial f_2} = -0.0128 \times \frac{\partial f_3}{\partial f_2}$$
$$= -0.0128 \times 0.5$$
$$= -0.0064$$

$$\frac{\partial L}{\partial f_1} = -0.0064 \times \frac{\partial f_2}{\partial f_1}$$
$$= -0.0064 \times 0.2350$$
$$= -0.0015$$

$$\frac{\partial L}{\partial x} = -0.0015 \times \frac{\partial f_1}{x}$$
$$= -0.0015 \times 0.5$$
$$= -0.00075$$

$$\frac{\partial L}{\partial w_1} = -0.0015 \times \frac{\partial f_1}{\partial w_1}$$
$$= -0.0015$$

$\therefore$ For layer-1, $\frac{\partial L}{\partial z_1} = -0.0015$

For layer-2, $\frac{\partial L}{\partial z_2} = -0.0128 \approx -0.013$

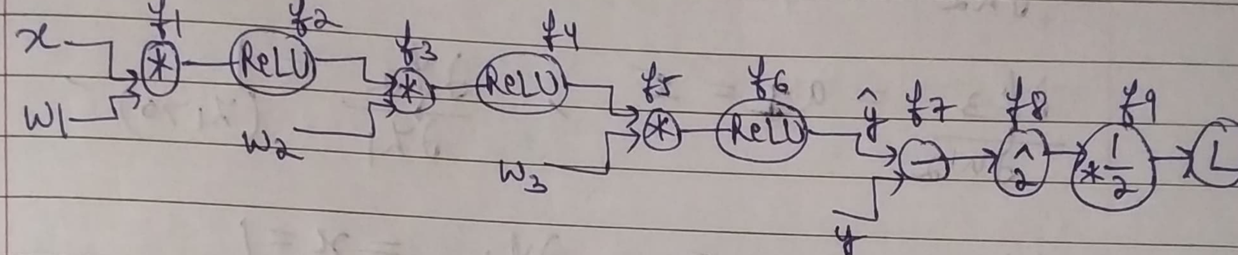
For layer-3; $\frac{\partial L}{\partial z_3} = -0.105$

Page No.

Date

Q1 (b) Do backpropagation to show the exploding gradient using ReLU function ; given that $x = 1$, $w = 5$, $y = 1$.

Ans :-



1) Forward pass :-

$z_1 = f_1 = x \times w_1 = 1 \times 5 = 5$

$a_1 = f_2 = \text{ReLU}(z_1) = \max(0, z_1) = \max(0, 5) = 5$

$z_2 = f_3 = a_1 \times w_2 = 5 \times 5 = 25$

$a_2 = f_4 = \max(0, 25) = 25$

$z_3 = f_5 = a_2 \times w_3 = 25 \times 5 = 125$

$\hat{y} = a_3 = f_6 = \max(0, 125) = 125$

$f_7 = \hat{y} - y = 125 - 1 = 124$

$f_8 = (f_7)^2 = (124)^2 = 15376$

$f_9 = \frac{1}{2} \times f_8 = \frac{15376}{2} = 7688$

or

L

2) Compute local ~~gradients~~ derivatives :-

$\frac{\partial f_9}{\partial f_8} = \frac{1}{2} = 0.5$, $\frac{\partial f_8}{\partial f_7} = 2 f_7 = 2 \times 124 = 248$

$\frac{\partial f_7}{\partial \hat{y}} = 1$, $\frac{\partial f_7}{\partial y} = 1$, $\frac{\partial f_6}{\partial z_3} = 1$ (since, $z_3 > 0$)

$$\frac{\partial f_5}{\partial a_2} = w_2 = 5 \;,\; \frac{\partial f_5}{\partial w_2} = a_2 = 25 \;,$$

$$\frac{\partial f_4}{\partial z_2} = 1 \;\text{(since } z_2 > 0\text{)} \;,\; \frac{\partial f_3}{\partial a_1} = w_2 = 5 \;,$$

$$\frac{\partial f_3}{\partial w_2} = a_1 = 5 \;,\; \frac{\partial f_2}{\partial z_1} = 1 \;(z_1 > 0) \;,$$

$$\frac{\partial f_1}{\partial x} = w_1 = 5 \;,\; \frac{\partial f_1}{\partial w_1} = x = 1$$

3) **Backpropagation :-**

$$\frac{\partial L}{\partial f_9} = \frac{\partial f_9}{\partial f_9} = 1$$

$$\frac{\partial L}{\partial f_8} = 1 \times \frac{\partial f_9}{\partial f_8} = 1 \times 0.5$$

$$\frac{\partial L}{\partial f_7} = 0.5 \times \frac{\partial f_8}{\partial f_7} = 0.5 \times 248 = 124$$

$$\frac{\partial L}{\partial f_6} = 124 \times \frac{\partial f_7}{\partial f_6} = 124 \times 1 = 124$$

$$\frac{\partial L}{\partial f_5} = 124 \times \frac{\partial f_6}{\partial f_5} = 124 \times 1 = 124$$

$$\frac{\partial L}{\partial f_4} = 124 \times \frac{\partial f_5}{\partial f_4} = 124 \times 5 = 620$$

Date

$$\frac{\partial L}{\partial f_3} = 620 \times \frac{\partial f_4}{\partial f_3} = 620 \times 1 = 620$$

$$\frac{\partial L}{\partial f_2} = 620 \times \frac{\partial f_3}{\partial f_2} = 620 \times 5 = 3100$$

$$\frac{\partial L}{\partial f_1} = 3100 \times \frac{\partial f_2}{\partial f_1} = 3100 \times 1 = 3100$$

$$\frac{\partial L}{\partial x} = 3100 \times \frac{\partial f_1}{\partial x} = 3100 \times 5 = 15500$$

$$\frac{\partial L}{\partial w_1} = 3100 \times \frac{\partial f_1}{\partial w_1} = 3100$$

$\therefore$ Gradients

For layer-1 $\Downarrow$ $\frac{\partial L}{\partial x_1} = 3100$

For layer-2 $\Downarrow$ $\frac{\partial L}{\partial x_2} = 620$

For layer-3 $\Downarrow$ $\frac{\partial L}{\partial x_3} = 124$

$\longleftarrow$ Gradients explode.

Q1 z output of a particular layer —

$$z = [4, 8, 6, 10] \rightarrow \text{a batch of train set}$$

$$\mu = \frac{4+8+6+10}{4} = \frac{28}{4} = 7$$

$$\sigma = \frac{1}{4} \sum_{i=1}^{4} (z - \mu)^2$$

no. of samples in a batch

$$\sigma = \frac{(4-7)^2 + (8-7)^2 + (6-7)^2 + (10-7)^2}{4 + \epsilon}$$

$\epsilon \rightarrow 10^{-5}$ → To avoid division error (undefined)

$\sigma = 5$

(vector) $\hat{z} = \dfrac{z - \mu}{\sigma + \epsilon}$

$$\boxed{\tilde{z} = \gamma \hat{z} + \beta}$$ *** → applied for each layer

→ γ and β are learnable parameters ~~at~~ which normalise the data at the place where it is shifted.

e.g.:- $\gamma = 2$, $\beta = 1$

$\hat{z}_1 = -1.3416$

$\tilde{z}_1 = 2(-1.3416) + 1$

$\tilde{z}_1 = -1.6832$

$\tilde{z}_2 = 1.8944$

$\tilde{z}_3 = 0 \ldots$

∴ Weights, γ and β are updated batchwise.

$$\underset{(\text{Batch-2})}{W_{new}} = \underset{(\text{Batch-1})}{W_{old}} - \alpha \frac{\partial L}{\partial W}$$

So, at each batch loss will ↓.

e.g.:- Dropout

$$1 \quad x_1 \xrightarrow[p=0.5]{w_1=2} \bigcirc$$

↳ 50% of time neuron has seen the samples.

Train:- Sample-1 $\Rightarrow$ O/P $= 1 \times 2 = 2$
Sample-2 $\Rightarrow$ O/P $= 0$ (dropout)
Sample-3 $\Rightarrow$ O/P $= 0$ (")
Sample-4 $\Rightarrow$ O/P $= 1 \times 2 = 2$

Total O/P $\Rightarrow 2+0+0+2 = 4$

Test:- Sample-1 $\Rightarrow$ O/P $= 1 \times 2 = 2$
Sample-2 $\Rightarrow$ O/P $= 2$
Sample-3 $\Rightarrow$ O/P $= 2$
Sample-4 $\Rightarrow$ O/P $= 2$

Total O/P $\Rightarrow 8$

Since test $\neq$ train, so we will divide the train total O/P by P (Keep_prob.), i.e., $4/0.5 = 8$; thus test O/P $\approx$ train O/P.

↳ So, each neuron O/P, i.e., 'a' is multiplied by $1/P$.

L2 + dropout $\Rightarrow$ Regularization (for every layer)

• CONVOLUTION OPERATION —

* e.g.:-

3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

n = Raw I/P (6 X 6) image
↳ sq. matrix

*
↓
convolution operator

Model parameters (How to choose?)

1	0	-1
1	0	-1
1	0	-1

↓
Filter or Kernel
$f = (3 \times 3)$

=

(4X4)

↳ Extracted (O/P) features of image

1) Superimpose the filter on raw image
2) Multiply the px value with filter value.
(Convolution operation)

3×1	0×0	1×-1	2	7	4
1×1	5×0	8×-1	9	3	1
2×1	7×0	2×-1	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

$3\times1+0\times0+1\times-1+1\times1+5\times0+8\times-1+2\times-1$
$+7\times0+2\times-1 = -5$

3) Put this value in O/P matrix.

-5			

4) Now, slide the filter by one position and repeat steps - 1 to 3.

	0×1	1×0	2×-1		
	5×1	8×0	9×-1		
	7×1	2×0	5×-1		

$0\times1+1\times0+2\times-1+5\times1+8\times0+9\times-1+7\times1+2\times0$
$+5\times-1=-4$

-5	-4		

5)

		$1^{\times 1}$	$2^{\times 0}$	$7^{\times -1}$	
		$8^{\times 1}$	$9^{\times 0}$	$3^{\times -1}$	
		$2^{\times 1}$	$5^{\times 0}$	$1^{\times -1}$	

$$1\times1+2\times0+7\times-1+8\times1+9\times0+3\times-1+2\times1$$
$$+5\times0+1\times-1=0$$

-5	-4	0	

6)

			2	7	4
			9	3	1
			5	1	3

$2 \times 1 + 7 \times 0 + 4 \times -1 + 9 \times 1 + 3 \times 0 + 1 \times -1 + 5 \times 1$
$+ 1 \times 0 + 3 \times -1 = 8$

$\downarrow$

-5	-4	0	8

7) Now, slide the filter one step down.

1	5	8			
2	7	2			
0	1	3			

$1 \times 1 + 5 \times 0 + 8 \times -1 + 2 \times 1 + 7 \times 0 + 2 \times -1 + 0 \times 1$
$+ 1 \times 0 + 3 \times -1 = -10$

$\downarrow$

-5	-4	0	8
-10			

and so on

$\downarrow$

Final o/p $\Rightarrow$
(features vector)

-5	-4	0	8
-10	-2	2	3
0	-2	-4	-7
-3	-2	-3	-16

Q1

2	3	7	4	6
6	6	9	8	7
3	4	8	3	8
7	8	3	6	6
4	2	1	8	3

$*$

1	-1
0	-1

$= ?$

$f = 2 \times 2$

$n = 5 \times 5$

$p = 1 \,;\, s = 2$

Ans:-

0	0	0	0	0	0	0
0	2	3	7	4	6	0
0	6	6	9	8	7	0
0	3	4	8	3	8	0
0	7	8	3	6	6	0
0	4	2	1	8	3	0
0	0	0	0	0	0	0

- $\cancel{0 \times 1} + \cancel{0 \times -1} + \cancel{0 \times 0} + 2 \times -1 = -2$ (O/P)
- Shift by 2 positions –

  $\cancel{0 \times 1} + \cancel{0 \times -1} + \cancel{3 \times 0} + 7 \times -1 = -7$ (O/P)
- Again shift by 2 positions –

  $\cancel{0 \times 1} + \cancel{0 \times -1} + \cancel{4 \times 0} + 6 \times -1 = -6$ (O/P)
- Shift 2 positions down –

  $\cancel{0 \times 1} + 6 \times -1 + \cancel{0 \times 0} + 3 \times -1 = -9$ (O/P)

- Shift 2 pos. to right —
  $1\times 6 + -1\times 9 + \cancel{4\times 0} + 8\times -1 = -11$ (O/P)

- Shift by 2 pos. to right —

  $8\times 1 + 7\times -1 + \cancel{3\times 0} + 8\times -1 = -7$ (O/P)

- Shift 2 pos. down —

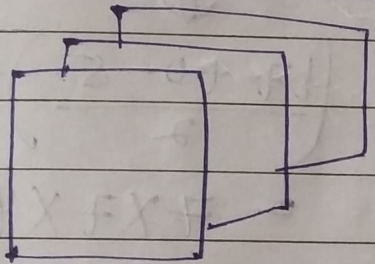
  ~~$7\times 1 + 8\times -1 + 4\times 0 + 2\times -1 = -3$ (O/P)~~
  $\cancel{0\times 1} + 7\times -1 + \cancel{0\times 0} + 4\times -1 = -11$ (O/P)
- $8\times 1 + 3\times -1 + \cancel{2\times 0} + 1\times -1 = 4$ (O/P)

- $\cancel{6\times 1} + \cancel{6\times -1} + \cancel{8\times 0} + 3\times -1 = -3$ (O/P)

output matrix —

-2	-7	-6
-9	-11	-7
-11	4	-3

Construct convolutional Network :- (For 1 sample)



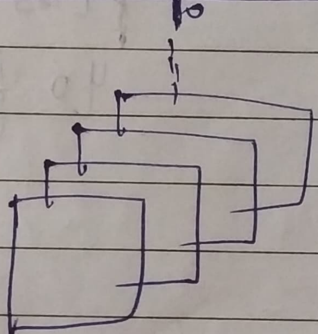
$39 \times 39 \times 3$

$f^{[1]} = 3$

$s^{[1]} = 1$

$p^{[1]} = 0$

10 filters



$37 \times 37 \times 10$

$\left(\frac{n+2p-f+1}{s}\right)$

1st layer

⇓

~~each filter has~~ 10 channels

Convolo operation

+

bias

+

ReLU

⇓

1 ConvNet (1st layer)

1st layer $\longrightarrow$ 2nd layer

$(37 \times 37 \times 10)$

$f^{[2]} = 5$

$s^{[2]} = 2$

$p^{[2]} = 0$

20 filters

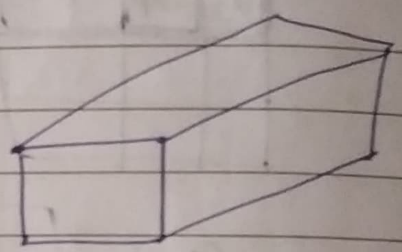
$\left(\frac{n+2p-f}{s}+1\right.$

$= \frac{37+0-5}{2}+1$

$= 34/2 = 17)$

$\Downarrow$

$17 \times 17 \times 20$



*** (Thus, at each ↑ing conv. layer, size reduces but channels ↑).

(we try to do this)

2nd layer $\longrightarrow$ 3rd layer

$(17 \times 17 \times 20)$

$f^{[3]} = 5$

$s^{[3]} = 2$

$p^{[3]} = 0$

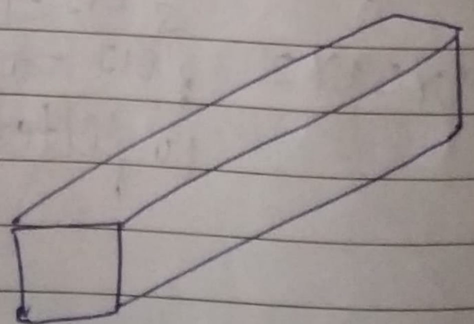
40 filters

$\left(\left(\frac{n+2p-f}{s}\right)+1\right)$

$\Downarrow$

$\left(\frac{17+0-5}{2}\right)+1$

$= 7 \times 7 \times 40$



learns 40 diff types of features

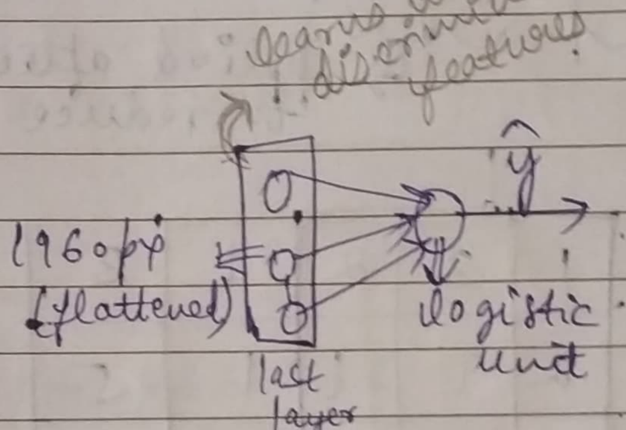
3rd layer
$(7 \times 7 \times 40 = 1960$ i/p$)$ $\longrightarrow$ flattened this

↓

passed to the last layer & apply logistic ϕ. (func)

↓

get $\hat{y}$



***

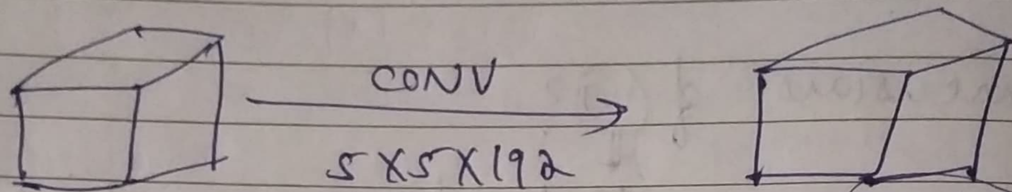
:: $\boxed{\text{I/P — CONV — CONV — CONV — O/P}}$

Q1 filter size ~~i/p~~ $= 3 \times 3 \times 3$, 10 filters } in 1 layer of convnet

Ans: $(3 \times 3 \times 3) + 1$ (bias) $= 27 + 1 = 28$ (for 1 filter)

for 10 filters $= 28 \times 10 = 280$ parameters in this layer

Q1



(a) 28×28×192 filters = 32 → 28×28×32

same CONV $(p=1)$ → I/P size = O/P size

$n+2p-f+1$
$28+2-5+1$

(b) Total operations (multiplications) = ?

Ans :-

1 filter = $5\times5\times192 = 4800$ (for one px value, perform 4800 conv.)

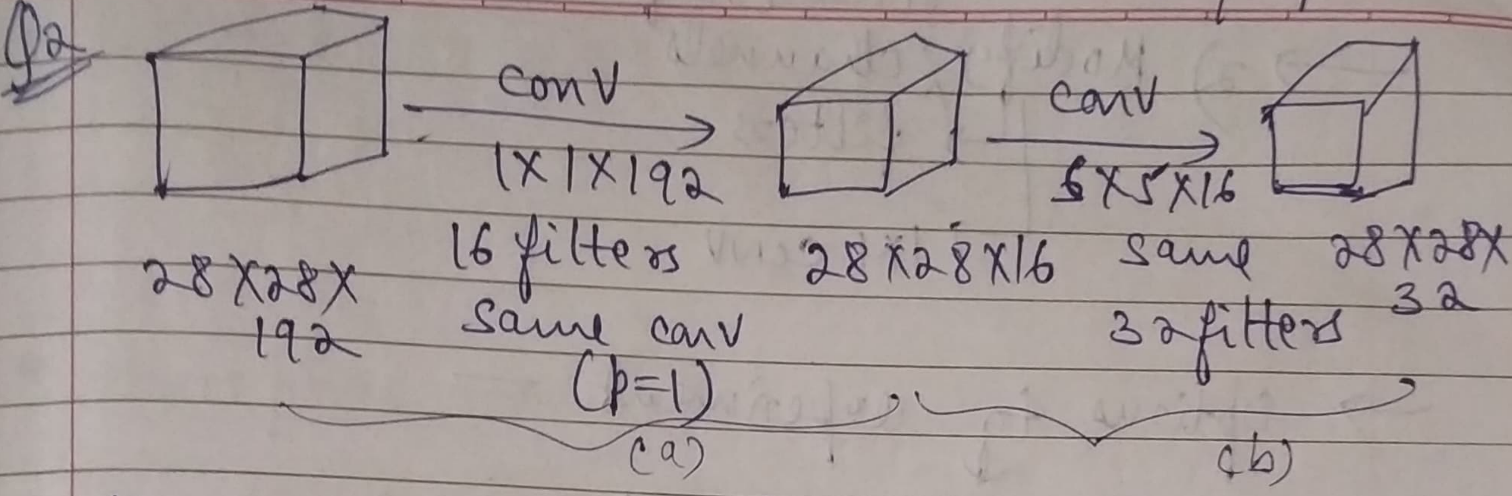
O/P size = $28\times28\times32$

→ 1 filter $(28\times28\times1) = 4800\times28\times28$
$= 3763200$

→ for remaining filters $(28\times28\times32)$
$= 32\times3763200$
$= 120{,}422{,}400$
$\approx$ 120 Million (multiplications)

Date 28/08/25

Q2



How many multiplications do you need?

Ans :-

(a) 1^{st} filter = 192

$28 \times 28 \times 192 \times 16 = 2,408,448 \approx 2.4$ million

(b) 2nd filter = $5 \times 5 \times 16 = 400$

$28 \times 28 \times 400 \times 32 = 10,035,200$

≈ 10 million

<u>Total multiplications</u> $\Rightarrow$ 2408448

+

10035200

= 12,443,648

$\approx$ <u>12 million</u>

$\therefore$ We can infer that — computational complexity has been reduced from 120 to 12 million.

Q1 $T = 2$ (two time steps)

$X^{(1)}: x_1 \in \mathbb{R}^2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ (representation vector), $x_2 \in \mathbb{R}^2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

x_t

Input dimension $= 2 \Rightarrow x_t \in \mathbb{R}^2$

Hidden state dimension $= 3 \Rightarrow h_t \in \mathbb{R}^3$

$h_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3\times 1}$

$w_1 \quad w_2 \quad w_3$

Ex. I am Yashpriya

$w_2 \rightarrow$ Gen Emb $\rightarrow [\ ]_{2\times 1}$

$x_1 \rightarrow$ RNN $\rightarrow [\ ]$ $D\times 1$, 1024×1, 512

$y_t \in \mathbb{R}^2$

(a) W_{hh} dim. = ? $= 3\times 3$

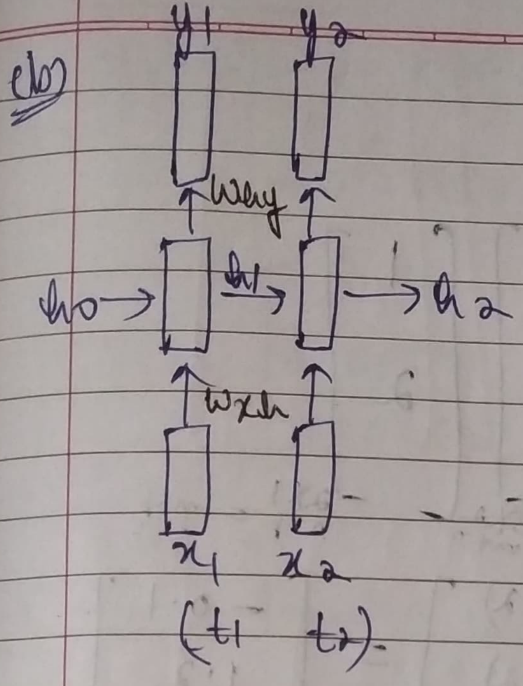
W_{xh} dim. = ? $= 3\times 2$

W_{hy} dim. = ? $= 2\times 3$

$$\underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}_{3\times 1}}_{h_t} = \underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}}_{\substack{W_{hh} \\ 3\times 3}} \underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}_{3\times 1}}_{\substack{h_{t-1} \\ \text{or} \\ h_0}} + \underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}}_{\substack{W_{xh} \\ 3\times 2}} \underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}_{2\times 1}}_{x_t}$$

$$\underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}_{2\times 1}}_{y_t} = \underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}}_{\substack{W_{hy} \\ 2\times 3}} \underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}_{3\times 1}}_{h_t}$$

(b)



compute h_1, y_1
h_2, y_2.

Ans:- use known — (given)

$$W_{xh} = \begin{bmatrix} 0.5 & -0.3 \\ 0.8 & 0.2 \\ 0.1 & 0.4 \end{bmatrix}_{3\times 2}$$

$$W_{hh} = \begin{bmatrix} 0.1 & 0.4 & 0.0 \\ -0.2 & 0.3 & 0.2 \\ 0.05 & -0.1 & 0.2 \end{bmatrix}_{3\times 3}$$

$$W_{hy} = \begin{bmatrix} 1.0 & -1.0 & 0.5 \\ 0.5 & 0.5 & -0.5 \end{bmatrix}$$

$$\underset{(h_1)}{h_t} = \tanh\left((W_{hh}\,\underbrace{h_{t-1}}_{h_0}) + (W_{xh}\,\underbrace{x_t}_{(x_1)})\right)$$

t=1:- $W_{hh} \cdot h_{t-1} = \begin{bmatrix} 0.1 & 0.4 & 0.0 \\ -0.2 & 0.3 & 0.2 \\ 0.05 & -0.1 & 0.2 \end{bmatrix} \cdot \underset{(h_0)}{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}}$

$$W_{hh} \cdot h_{t-1} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$W_{xh} \cdot x_1 = \begin{bmatrix} 0.5 & -0.3 \\ 0.8 & 0.2 \\ 0.1 & 0.4 \end{bmatrix}_{\substack{3\times2 \\ m\times n}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{\substack{2\times1 \\ n\times p}} = \substack{m\times p \\ 3\times1}$$

$$W_{xh} \cdot x_1 = \begin{bmatrix} (0.5\times1)+(-0.3\times2) \\ (0.8\times1)+(2\times0.2) \\ (0.1\times1)+(2\times0.4) \end{bmatrix} = \begin{bmatrix} -0.1 \\ 1.2 \\ 0.9 \end{bmatrix}$$

$$h_1 = \tanh \begin{bmatrix} -0.1 \\ 1.2 \\ 0.9 \end{bmatrix} = \begin{bmatrix} \tanh(-0.1) \\ \tanh(1.2) \\ \tanh(1.9) \end{bmatrix}$$

$$h_1 = \begin{bmatrix} -0.099 \\ 0.83 \\ 0.716 \end{bmatrix}$$

$$y_t = W_{hy} \cdot h_t$$

$$y_1 = \begin{bmatrix} 1.0 & -1.0 & 0.5 \\ 0.5 & 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} -0.099 \\ 0.83 \\ 0.716 \end{bmatrix}$$

$$y_1 = \begin{bmatrix} (1.0\times-0.099)+(-1.0\times0.83)+(0.5\times0.716) \\ (0.5\times-0.099)+(0.83\times0.5)+(-0.5\times0.716) \end{bmatrix}$$

$$y_1 = \begin{bmatrix} -0.099 + (-0.83) + 0.358 \\ -0.0495 + 0.415 + (-0.358) \end{bmatrix}$$

$$y_1 = \begin{bmatrix} -0.571 \\ 0.0075 \end{bmatrix}$$

$t=2$:- $h_2 = \tanh\left(W_{hh} h_1 + W_{xh} x_2 \right)$

$$W_{hh} h_1 = \begin{bmatrix} 0.1 & 0.4 & 0.0 \\ -0.2 & 0.3 & 0.2 \\ 0.05 & -0.1 & 0.2 \end{bmatrix} \begin{bmatrix} -0.099 \\ 0.83 \\ 0.716 \end{bmatrix}$$

$$W_{hh} h_1 = \begin{bmatrix} (0.1 \times -0.099) + (0.4 \times 0.83) + 0 \\ (-0.2 \times -0.099) + (0.3 \times 0.83) + (0.2 \times 0.716) \\ (0.05 \times -0.099) + (-0.1 \times 0.83) + (0.2 \times 0.716) \end{bmatrix}$$

$$W_{hh} h_1 = \begin{bmatrix} -0.0099 + 0.332 + 0 \\ 0.0198 + 0.249 + 0.1432 \\ -0.00495 + (-0.083) + 0.1432 \end{bmatrix} = \begin{bmatrix} 0.3221 \\ 0.412 \\ 0.05525 \end{bmatrix}$$

$$W_{xh} x_2 = \begin{bmatrix} 0.5 & -0.3 \\ 0.8 & 0.2 \\ 0.1 & 0.4 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$W_{xh} x_2 = \begin{bmatrix} (0.5 \times 1) + (-0.3 \times 1) \\ (0.8 \times -1) + (0.2 \times 1) \\ (0.1 \times -1) + (0.4 \times 1) \end{bmatrix} = \begin{bmatrix} -0.8 \\ -0.6 \\ 0.3 \end{bmatrix}$$

$$W_{hh} h_1 + W_{xh} x_2 = \begin{bmatrix} 0.3221 \\ 0.412 \\ 0.05525 \end{bmatrix} + \begin{bmatrix} -0.8 \\ -0.6 \\ 0.3 \end{bmatrix}$$

$$W_{hh} h_1 + W_{xh} x_2 = \begin{bmatrix} 0.3221 + (-0.8) = -0.4779 \\ 0.412 + (-0.6) = -0.188 \\ 0.05525 + 0.3 = 0.35525 \end{bmatrix}$$

$$h_2 = \tanh \begin{pmatrix} -0.4779 \\ -0.188 \\ 0.35525 \end{pmatrix} = \begin{bmatrix} -0.444 \\ -0.185 \\ 0.341 \end{bmatrix}$$

$$y_2 = W_{hy} \cdot h_2$$

$$y_2 = \begin{bmatrix} 1.0 & -1.0 & 0.5 \\ 0.5 & 0.5 & -0.5 \end{bmatrix} \begin{bmatrix} -0.444 \\ -0.185 \\ 0.341 \end{bmatrix}$$

$$y_2 = \begin{bmatrix} -0.444 + 0.185 + 0.1705 \\ \cancel{-0.0925} \\ -0.222 - 0.0925 - 0.1705 \end{bmatrix} = \begin{bmatrix} -0.0885 \\ -0.485 \end{bmatrix}$$