Solved Practise Questions

Based on CNN and BPPT

Atharv Shendage

November 30, 2023

Q(1)

Given below is some info of CNN model layers used for classification , fill in the blanks and also add 2 more layers such that we can classify the input in one of the 10 classes.

Layer	Type	Maps	Size	Kernel Size	Stride	Activation
Input	Input	1	32×32	–	–	–
C_1	Conv	6	–	5×5	1	tanh
S_2	Avg. pool	6	–	2×2	2	tanh
C_3	Conv	16	–	5×5	1	tanh
S_4	Avg pool	16	–	2×2	2	tanh
C_5	Conv	120	–	5×5	1	–

Ans

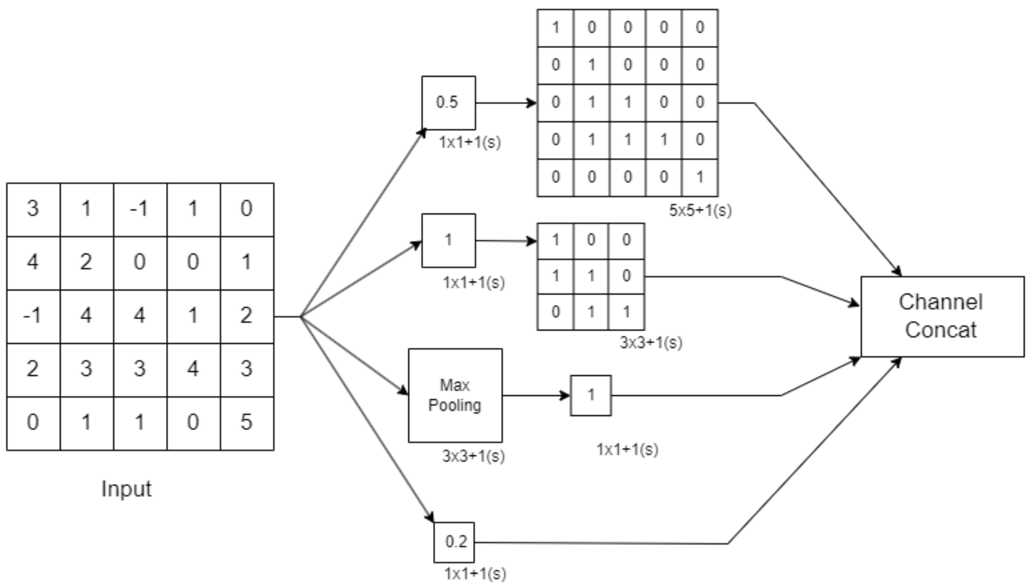
We can fill in the blanks using the formula ,$n_{out} = \frac{n_{in}+2\times padding-k}{stride} + 1$.
And for adding 2 layers , we know the final output size needs to be 10 and after flattening C_5 we have size of 120, taking the average of 120 and 10 , we get ≈ 64 and both of those are fully connected layers.

So the completed model summary looks like :

Layer	Type	Maps	Size	Kernel Size	Stride	Activation
Input	Input	1	32×32	–	–	–
C_1	Conv	6	28×28	5×5	1	tanh
S_2	Avg. pool	6	14×14	2×2	2	tanh
C_3	Conv	16	10×10	5×5	1	tanh
S_4	Avg pool	16	5×5	2×2	2	tanh
C_5	Conv	120	1×1	5×5	1	$flatten$
D_6	FC	64	1×1	5×5	1	$softmax$
D_7	FC	10	1×1	5×5	1	$softmax$

Q2

Inception Module - Giver below is an image of 5x5x1 and passed through Google's Inception module v1. The kernel sizes and strides are given below. Find out the output at last stage after concatenating all the individual results



Ans. Let's break it down in 4 paths and then concatenate the results :

Path 1 : Just 1*1 Conv

Input:

3	1	-1	1	0
4	2	0	0	1
-1	4	4	1	2
2	3	3	4	3
0	1	1	0	5

0.2

1x1+1(s)

0.6	0.2	-0.2	0.2	0
0.8	0.4	0	0	0.2
-0.2	0.8	0.8	0.2	0.4
0.4	0.6	0.6	0.8	0.6
0	0.2	0.2	0	1

O1

Path 2 : 3*3 max pool followed by 1*1 conv

3	1	-1	1	0
4	2	0	0	1
-1	4	4	1	2
2	3	3	4	3
0	1	1	0	5

max pool

3x3+1(s)

4	4	2	1	1
4	4	4	4	2
4	4	4	4	4
4	4	4	5	5
3	3	4	5	5

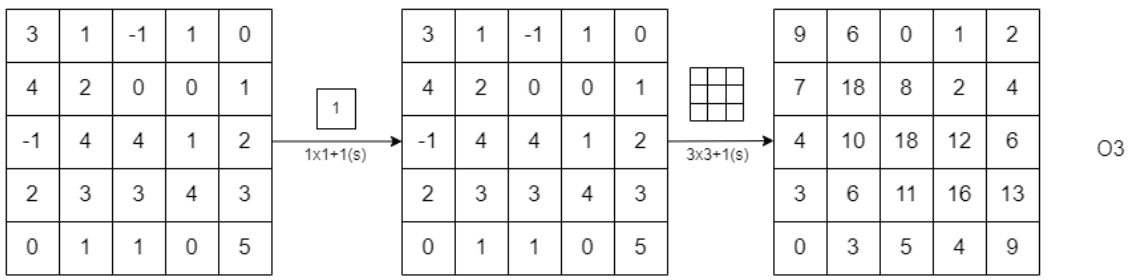
1

1x1+1(s)

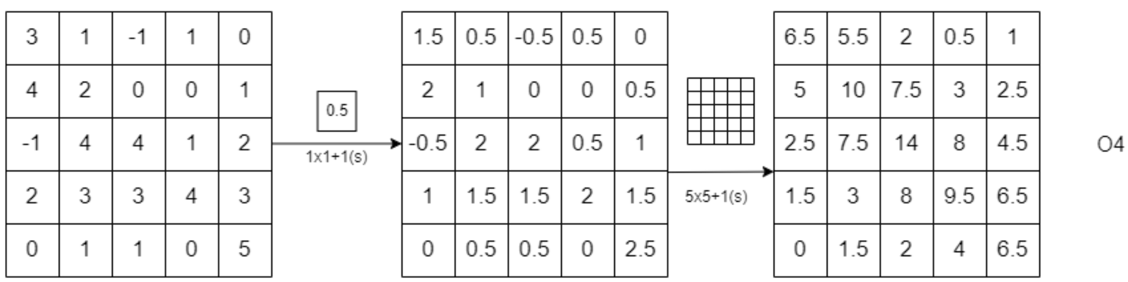
4	4	2	1	1
4	4	4	4	2
4	4	4	4	4
4	4	4	5	5
3	3	4	5	5

O2

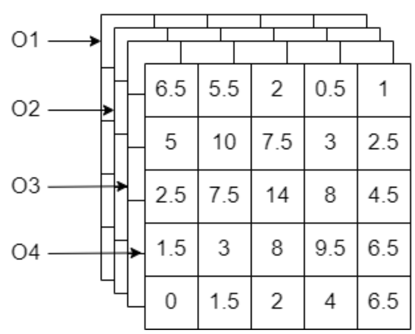
Path 3 : 1*1 conv followed by 3*3 conv



Path 4 : 1*1 conv followed by 5*5 conv



Concatenate (stacking) it all up



Q(3)

You have given that ,

$$y(t) = \omega_3 \cdot h(t)$$

$$\& \quad h(t+1) = \omega_1 \cdot h(t) + \omega_2 \cdot x(t)$$

Initial values of $\omega_1, \omega_2, \omega_3 = (0.5, 0.5, 1)$
And $h(0) = 0$
Input Sequence $\Rightarrow (100, 110, 120, 130)$ And
Ground Truth $\Rightarrow (-, 112, 118, 129, 140)$
So, Compute Loss, Gradient and update the parameters.

Ans

Let's calculate hidden state predicted values at each time step first and then calculate the gradients.
So, at t=1 ;

$$\begin{aligned} h(1) &= \omega_1 \cdot h(0) + \omega_2 \cdot x(0) \\ &= 0 + 0.5 \times 100 \\ &= 50 \\ \therefore y(1) &= 50 \end{aligned}$$

Similarly,

$$\begin{aligned} h(2) &= 0.5 \times 50 + 0.5 \times 110 \\ &= 80 \\ \therefore y(2) &= 80 \end{aligned}$$

$$\begin{aligned} h(3) &= 0.5 \times 80 + 0.5 \times 120 \\ &= 100 \\ \therefore y(3) &= 100 \end{aligned}$$

$$\begin{aligned} h(4) &= 0.5 \times 100 + 0.5 \times 120 \\ &= 110 \\ \therefore y(4) &= 110 \end{aligned}$$

Predicted output sequence $= (0, 50, 80, 100, 110)$ and
Ground truth is given as $(-, 112, 118, 129, 150)$
We have,

$$\begin{aligned} \text{Loss} &= \frac{1}{4} \sum_{t=1}^{4} (y_t - \hat{y}_t)^2 \\ &= \frac{(y_4 - \hat{y}_4)^2}{4} + \cdots + \frac{(y_1 - \hat{y}_1)^2}{4} \\ &= \left(62^2 + 38^2 + 29^2 + 30^2\right)/4 \\ &= 1757.25 \end{aligned}$$

For easier understanding let's write as ,

$$L=\sum_{t=1}^{4} L_t = L_4 + L_3 + L_2 + L_1$$

Now let's calculate $\frac{\partial L}{\partial \omega_3}$ we have,

$$\frac{\partial L_4}{\partial \omega_3} = \frac{\partial L_4}{\partial \hat{y}_4} \cdot \frac{\partial \hat{y}_4}{\partial \omega_3}$$

$$= \frac{-\left(y_4-\hat{y}_4\right)}{2} \cdot h_4$$

adding up all the losses we get

$$\begin{aligned} \therefore \quad \frac{\partial L}{\partial \omega_3} &= -\frac{1}{2} \cdot \sum_{t=1}^{4} \left(y_t - \hat{y}_t\right) \cdot h_t \\ &= -\frac{1}{2}(62 \cdot 50 + 38 \cdot 80 + 29 \times 100 + 30 \times 110) \\ &= -6170 \end{aligned}$$

Though calculation of $\frac{\partial L}{\partial \omega_3}$ was easy , Now we need to calculate each hidden state's derivative w.r.t. $w_1$to calculate $\frac{\partial L}{\partial \omega_1}$which in turn depends on derivative w.r.t w_1 of the previous state , as shown below ;

$$h_4 = \omega_1 \cdot h_3 + w_2 \cdot x_3$$

$$\frac{\partial h_4}{\partial \omega_1} = h_3 + \omega_1 \cdot \frac{\partial h_3}{\partial \omega_1}$$

Similarly,

$$\frac{\partial h_3}{\partial \omega_1} = h_2 + \omega_1 \cdot \frac{\partial h_2}{\partial \omega_1}$$

$$\frac{\partial h_2}{\partial w_1} = h_1 + w_1 \cdot \frac{\partial h_1}{\partial \omega_1}$$

$$\frac{\partial h_1}{\partial \omega_1} = h_0 + w_1 \cdot \frac{\partial h_0}{\partial \omega_1}$$

$$\frac{\partial h_1}{\partial \omega_1} = 0 + 0.5 \cdot 0 = 0$$

...($h(0)$ is given 0 and $\frac{\partial h_1}{\partial \omega_1}$ is assumed to be 0)

Putting these values back we get ,

$$\frac{\partial h_2}{\partial w_1} = 50 + 0 \cdot 5 + 0 = 50$$

$$\frac{\partial h_3}{\partial \omega_1} = 80 + 0.5 \times 50 = 105$$

$$\frac{\partial h_4}{\partial \omega_1} = 100 + 0.5 + 105 = 152.5$$

Now We have ,

$$\frac{\partial L_4}{\partial \omega_1} = \frac{\partial L_4}{\partial \hat{y}_4} \cdot \frac{\partial \hat{y}_4}{\partial h_4} \cdot \frac{\partial h_4}{\partial \omega_1}$$

Aggregating all losses we have,

$$\begin{aligned}
\therefore \frac{\partial L}{\partial \omega_1} &= \sum_{t=1}^{4} \frac{\partial L_t}{\partial \hat{y}_t} \cdot \frac{\partial \hat{y}_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial \omega_1} \\
&= \frac{-1}{2} \sum_{t=1}^{4} (y_t - \hat{y}_t) \cdot \omega_3 \cdot \frac{\partial h_t}{\partial \omega_1} \\
&= -\frac{1}{2}(62 \cdot 0 + 38 \cdot 50 + 29 \cdot 105 + 30 \times 152.5) \\
&= -4760
\end{aligned}$$

Now we need to calculate each hidden state's derivative w.r.t. $w_2$which is very similar as the case with $w_1$because $\frac{\partial L}{\partial \omega_2}$for state h_t in turn depends on $\frac{\partial L}{\partial \omega_2}$ of stateh_{t-1}

We have,

$$\begin{aligned}
\frac{\partial L_4}{\partial \omega_2} &= \frac{\partial L_4}{\partial \hat{y}_4} \frac{\partial \hat{y}_4}{\partial h_4} \cdot \frac{\partial h_4}{\partial \omega_2} \\
&= \frac{\partial L_4}{\partial \hat{y}_4} \cdot \frac{\partial \hat{y}_4}{\partial h_4} \cdot \left[x_4 + \frac{\partial (\omega_1 \cdot h_3)}{\partial \omega_2} \right] \\
\frac{\partial L_4}{\partial \omega_2} &= \frac{\partial L_4}{\partial \hat{y}_4} \frac{\partial \hat{y}_4}{\partial h_4} \cdot \left[x_4 + \omega_1 \cdot \frac{\partial h_3}{\partial \omega_2} \right]
\end{aligned}$$

for getting $\frac{\partial h_4}{\partial \omega_2}$, let's calculate $\frac{\partial h_4}{\partial \omega_1}$ first.

$$\begin{aligned}
\frac{\partial h_1}{\partial \omega_2} &= x_1 + \omega_1 \cdot \frac{\partial h_0}{\partial \omega_2} = x_1 = 100 \\
\therefore \frac{\partial h_2}{\partial \omega_2} &= x_2 + \omega_1 \cdot \frac{\partial h_1}{\partial \omega_2} = 110 + 0.5 \times 100 = 160 \\
\therefore \frac{\partial h_3}{\partial \omega_2} &= 120 + 0.5 \times 160 = 200 \\
\therefore \frac{\partial h_4}{\partial \omega_2} &= 130 + 0.5 \times 200 = 230
\end{aligned}$$

Aggregating all the losses ,

$$\frac{\partial L}{\partial \omega_2} = \sum_{t=1}^{4} \frac{\partial L_t}{\partial \hat{y}_t} \cdot \frac{\partial \hat{y}_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial \omega_2}$$

$$\begin{aligned}
&= -\frac{1}{2} \cdot \Sigma (y_t - \hat{y}_t) \cdot \omega_3 \cdot \frac{\partial h_t}{\partial \omega_2} \\
&= -\frac{1}{2}(61 \cdot 100 + 38 \times 160 + 29 \times 200 + 30 \times 230) \\
\frac{\partial L}{\partial \omega_2} &= -12,490
\end{aligned}$$

$\therefore$ Now gradients are

$$\frac{\partial L}{\partial \omega_1} = -4760 \quad \frac{\partial L}{\partial \omega_2} = -12,490 \quad \frac{\partial L}{\partial \omega_3} = -6170$$

For Itr_0 $(\omega_1, \omega_2, \omega_3) \equiv (0.5, 0.5, 1)$

Updating parameters with $\alpha = 10^{-5}$ we get,

$$\begin{aligned}\omega_1 &= \omega_1 - \alpha \cdot \frac{\partial L}{\partial \omega_1} \\ &= 0.5 + 4760 \times 10^{-5} \\ &= 0.55\end{aligned}$$

$$\begin{aligned}\omega_2 &= 0.5 + 12490 \times 10^{-5} = 0.63 \\ \omega_3 &= 1 + 6170 \times 10^{-5} = 1.06\end{aligned}$$

Now let's use these updated parameters and predict new $y's$

$$\begin{aligned}h(1) &= 0 + 0.62 + 100 \\ &= 62 \\ y(1) &= 1.06 \times 62 \\ &= 66\end{aligned}$$

$$\begin{aligned}h(2) &= .55 \times 62 + .62 + 110 \\ &= 96.1 \\ y(2) &= 102\end{aligned}$$

$$\begin{aligned}h(3) &= .55 \times 96.1 + .62 \times 120 \\ &= 127.25 \\ y(3) &= 135\end{aligned}$$

$$\begin{aligned}h(4) &= .55 \times 127.25 + .62 \times 130 \\ &= 150.5 \\ y(4) &= 160\end{aligned}$$

$$\begin{aligned}\text{New Loss } &= \frac{50^2 + 8^2 + 6^2 + 20^2}{4} \\ &= 750\end{aligned}$$

We can see that Loss has reduced.
(Note : in calculation I found $\alpha = 10^{-3}$ to be large so I have taken $\alpha = 10^{-5}$)