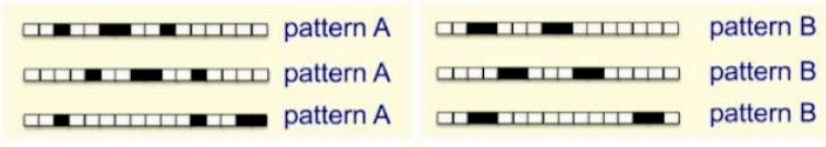
Question 1



Consider the above classification of patterns . The training set consists of patterns A and B in all possible translations. Consider a neural network that consists of a 1D convolution layer with a linear activation function, followed by a linear layer with a logistic output. Can such an architecture perfectly classify all of the training examples? Why or why not?

Add file

Question 2

Write the core difference of RNN ,Peephole-LSTM and GRU? What is the basic difference of LSTM forget gate and GRU's reset gate. Show the matrix workflow of the LSTM.

Add file

Question 3

If you have a 64*64 binary image at input in a CNN network with 7 filters(size of 5*5) stride of 2 and no padding of 0 and apply 3 sets of Conv and max pool(size of 2*2) what will be the number of nodes in the flattening layer? Show each steps after conv and max pool layers happen.

Add file

Question 2

Write the core difference of RNN ,Peephole-LSTM and GRU? What is the basic difference of LSTM forget gate and GRU's reset gate. Show the matrix workflow of the LSTM.

Add file

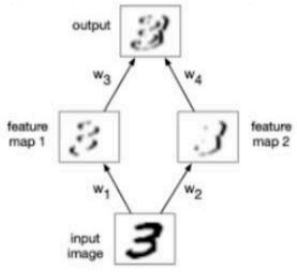
Question 3

If you have a 64*64 binary image at input in a CNN network with 7 filters(size of 5*5) stride of 2 and no padding of 0 and apply 3 sets of Conv and max pool(size of 2*2) what will be the number of nodes in the flattening layer? Show each steps after conv and max pool layers happen.

Add file

Question 4

Explain the shared concept of CNN? You will design a convolutional network to detect vertical boundaries in an image. The architecture of the network is as shown below.



The ReLU activation function is applied to the first convolution layer. The output layer uses the linear activation function. Design two convolution filters for the rst layer, of size 3 x 3. One of them should detect dark/light boundaries, and the other should detect light/dark boundaries.

Add file

Submit

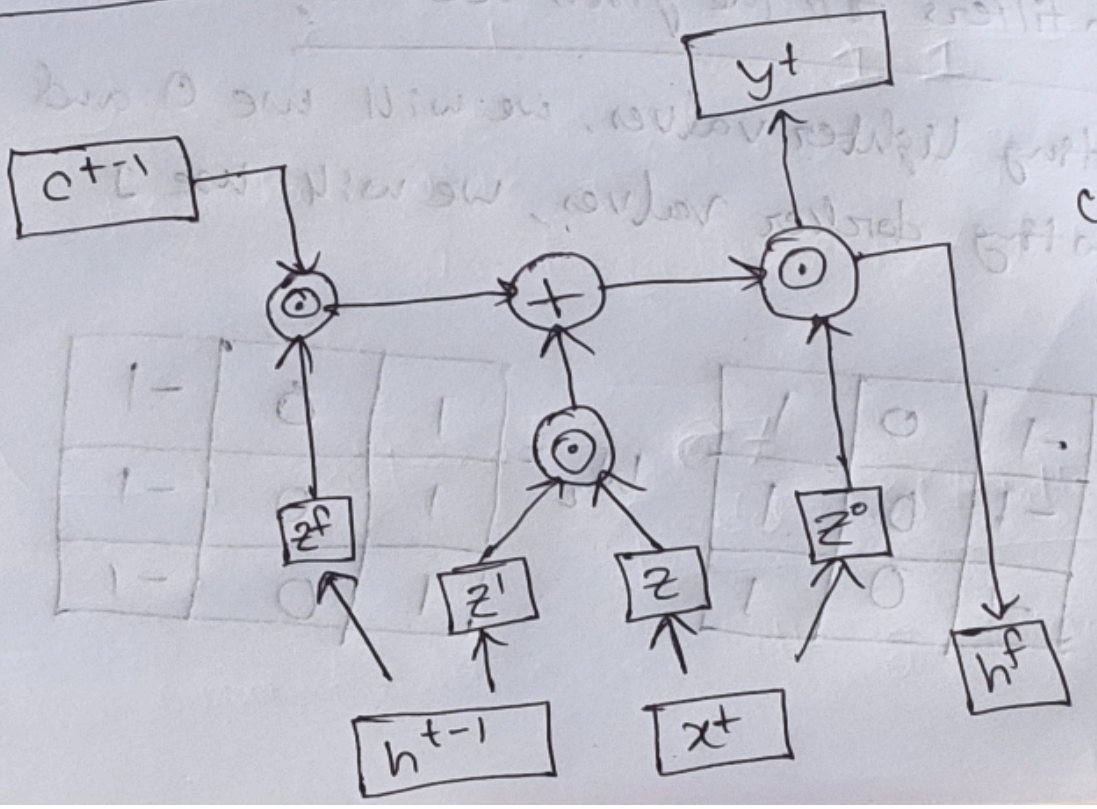
Clear form

has two doors. Moreover, LSTMs has two activation functions σ_1, σ_1 and $\sigma_2 \sigma_2$ while GRU has only one activation function.

The basic difference of LSTM forget gate and GRU, reset gate is forget gate controls what is kept forgotten from the previous cell state. It will decide how much information from the previous state should kept and forget remaining.

On contrast, output gate controls which parts of the cell are output to the hidden state, it will determin what the next hidden state will be.

Matrix Workflow of LSTM:



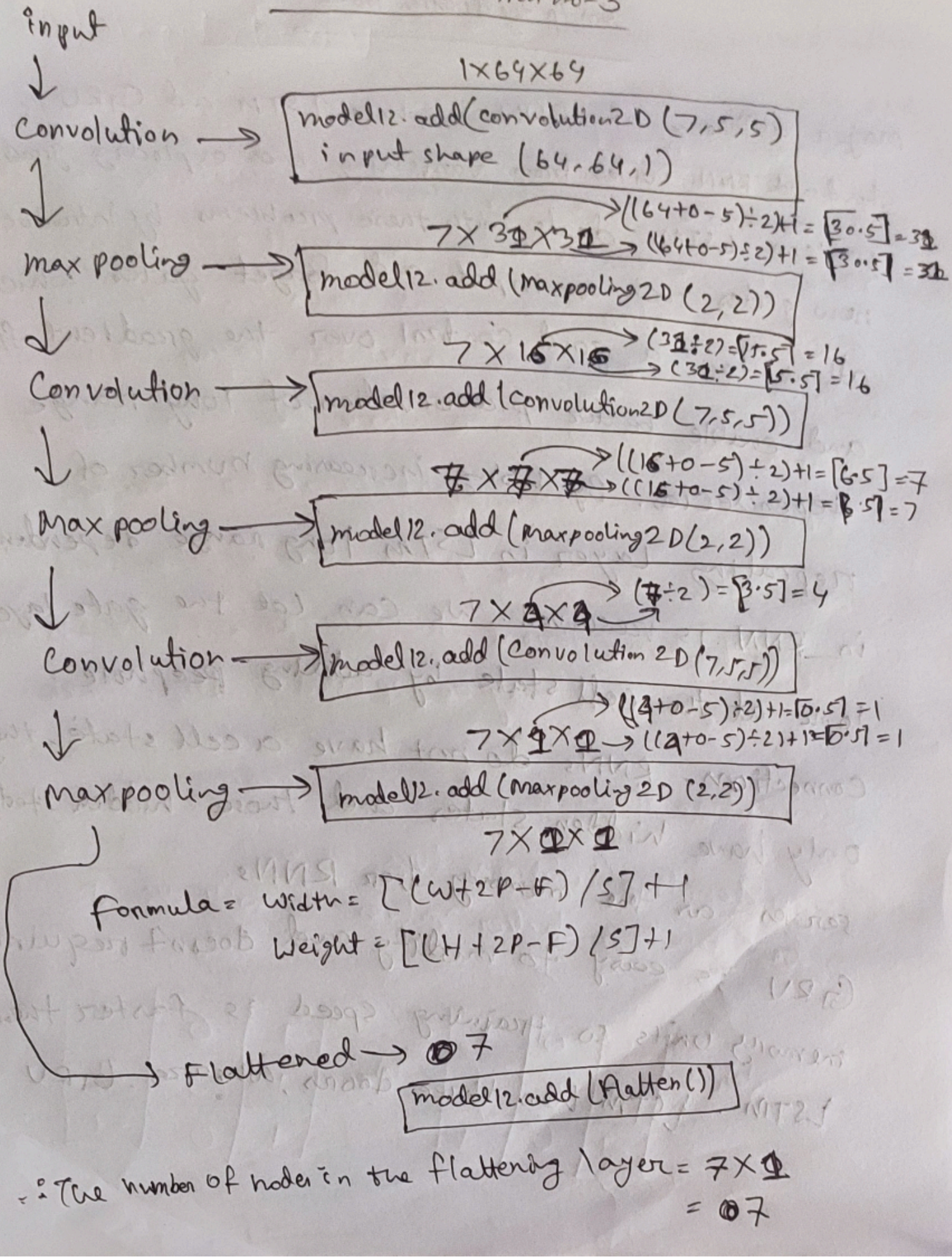
$$c^t = z^f \odot c^{t-1} + z^i \odot z$$
$$h^t = z^o \odot \tanh(c^t)$$
$$y^t = \sigma(W h')$$

Ans to the Question no-2

Major difference of RNN, Peephole-LSTM and GRU is standard RNN suffer from vanishing and exploding gradient problems. LSTMs deals with these problems by introducing new gates such as input and forget gates which allows for a good control over the gradient flow and enable better preservation of long range dependecies. By using increasing number of repeating layer in LSTM long range dependency in RNN is resolved. We can let the gate layers look at the cell state by adding peephole connections. RNNs do not have a cell state, they only have hidden states and those hidden states server as the memory for RNNs.

GRU is easy to modify and doesn't require memory units so training speed is faster then LSTM. LSTM has three doors, whereas GRU

Ans to the Question no-3



∴ The number of node in the flattening layer = 7×1
= 07

Ans to the question no- 4

Shared concept of CNN:

To reiterate parameter sharing occurs when a feature map is generated from the result of the convolution between a filter and input data from a unit within a plane in the convolution layer. All units within this layer plane share the same weights; hence it is called weight/parameter sharing. It means that the same parameters will be used to represent two different transformations in the system. So, same matrix may be updated multiple times.

Convolution filters for the given scenario:

For denoting lighter values, we will use 0 and for denoting darker values, we will use 1

$W_1 =$

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

$W_2 =$

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