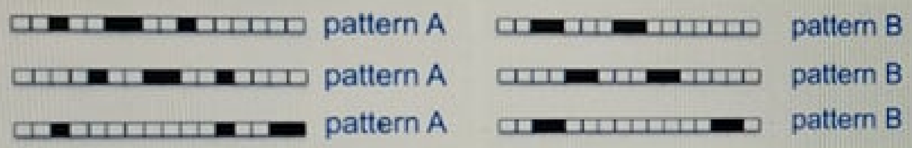
Answer Any 3 Questions

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?

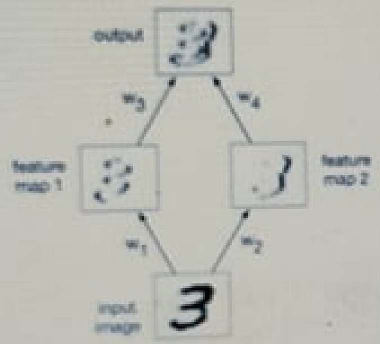
Add file

Question 2

...network to detect vertical

Question 2

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 first layer, of size 3 x 3. One of them should detect black/white boundaries, and the other should detect white/black boundaries.

Add file

Ans to the Qs: 2

Parameter sharing occurs when a feature map is generated from the result of convolution between a filter and input data from a unit in a plane in the convolution layer. This results in all or most of the units in the layer sharing the weights, which is a major feature of CNNs.

CNN for vertical boundaries:

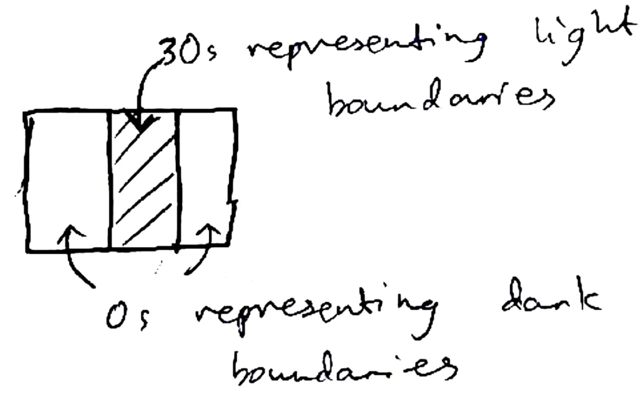
$$\begin{matrix} 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \end{matrix} \times \begin{matrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{matrix} = \begin{matrix} 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \end{matrix}$$

6X6 image

3X3 filter 1

4X4

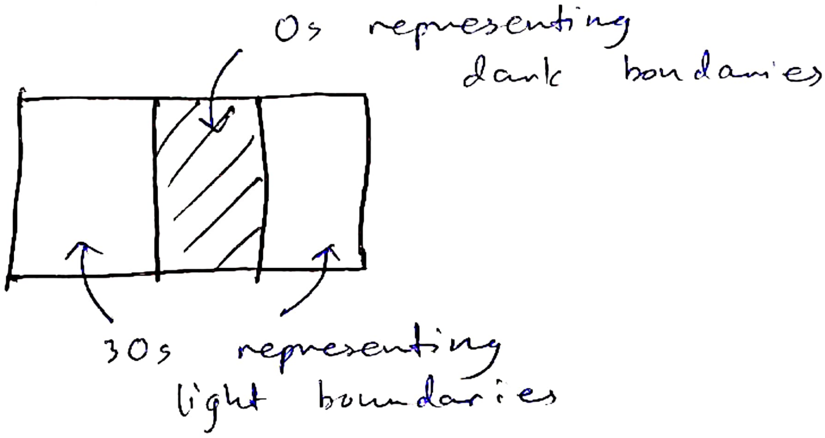
Convolving with this filter results in having the middle part of the 4X4 result in having a dark boundary shown below:



Filter 2 would represent detection of light edges. Therefore, ~~resulti~~ resulting in a 4x4 image having light boundaries.

$$\begin{matrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{matrix}$$

Filter 2



white/black boundaries.

Add file

Question 3

What is update gate and reset gate of GRU(Gated Recurrent Unit)? What is the difference of the work flow of LSTM and GRU? Show your reasons why GRU is faster in compare to RNN??

Add file

Question 4

Ans to the Qs: 3

To solve the vanishing gradient problem of a standard RNN, GRU uses the two gates - update gate and reset gate. Basically, these are two vectors which decide what information should be passed to the output

Update Gate: The update gate helps the model to determine how much of the past information from previous time steps needs to be passed along to the future.

~~Reset gate: The reset gate is used from the~~ model to ~~decide how much~~

For a fully ~~connected~~ gated unit,

$$z_t = \sigma_g (W_z \cdot x_t + U_z \cdot h_{t-1} + b_z)$$

Reset Gate: The reset gate is used from the model to decide how much of the past information to forget.

For a fully gated unit,

$$r_t = \sigma_g (W_r \cdot x_t + U_r \cdot h_{t-1} + b_r)$$

The key differences between GRU and LSTM is given below:

1) GRU has 2 gates - update and reset gates. LSTM has 3 gates - input, output, and forget gates.

2) GRU is less complex than LSTM as it ~~has~~ has less number of gates.

3) If the dataset is small then GRU is preferred. Otherwise, LSTM should be ~~use~~ used for larger dataset.

4) LSTM has cell memory whereas GRU doesn't. LSTM used cell memory to avoid vanishing gradient problem whereas GRU uses computational data flow.

5) GRU improved LSTM by omitting cell memory and using fewer parameters.

6) GRU exposes the complete memory and hidden layers but LSTM doesn't.

RNNs face short-term memory problem. It is caused due to vanishing gradient problem. As RNN processes more steps it suffers from vanishing gradient more than other neural network architectures. RNNs also face the counter-part of vanishing gradient problem called the exploding gradient problem. GRUs do not face such issues.

GRU is faster than RNN as it uses less parameters. GRU can be made even faster using minimal gated variant which uses only one gate which is a combination of update and reset gates of the fully gated variant.

Question 4

Compute the local induced gradient δ_m for the first layer in the picture. For simplicity every node of the previous layer is connected to the next layer node. [Fully connected]

You may assume ,

Output neuron has δ_f as its local induced error and previous layer has δ_h δ_p

Output from output neuron, $y_f = \varphi'(v_f)$

Output from h layer neuron $y_h = \varphi'(v_h)$

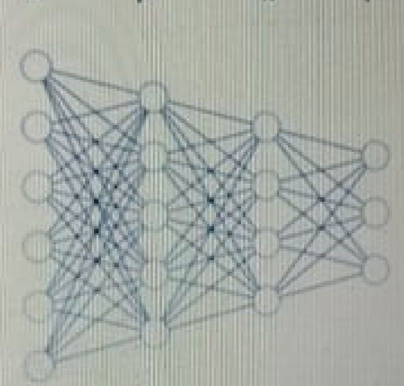
Output from p layer neuron $y_p = \varphi'(v_p)$

Output from m layer neuron $y_m = \varphi'(v_m)$

$v_j = \sum_j w_{ji} y_i$ [For any neuron]

$e_j = d_j - y_j$

Total error of network : $E = \frac{1}{2}\sum_j e_j^2$



↑ Add file

Ans to the Q: 4

~~For k layer,~~
$\cancel{\delta_{h_1} \phi}$

For f - layer,

$$\delta_{f_1} = e_{f_1} \phi'(v_{f_1})$$

$$\delta_{f_2} = e_{f_2} \phi'(v_{f_2})$$

$$\delta_{f_3} = e_{f_3} \phi'(v_{f_3})$$

For h - layer,

$$\delta_{h_1} = \phi''(v_{h_1}) \sum_{f \in C} \delta_f w_{fh_1}$$

$$\delta_{h_2} = \phi''(v_{h_2}) \sum_{f \in C} \delta_f w_{fh_2}$$

$$\delta_{h_3} = \phi''(v_{h_3}) \sum_{f \in C} \delta_f w_{fh_3}$$

$$\delta_{h_4} = \phi''(v_{h_4}) \sum_{f \in C} \delta_f w_{fh_4}$$

For p - layer,

$$\delta p_1 = \phi'''(v p_1) \sum_{h \in C} \delta_h w_{hp_1}$$

$$\delta p_2 = \phi'''(v p_2) \sum_{h \in C} \delta_h w_{hp_2}$$

$$\delta p_3 = \phi'''(v p_3) \sum_{h \in C} \delta_h w_{hp_3}$$

$$\delta p_4 = \phi'''(v p_4) \sum_{h \in C} \delta_h w_{hp_4}$$

$$\delta p_5 = \phi'''(v p_5) \sum_{h \in C} \delta_h w_{hp_5}$$

For m - layer,

$$\delta m_1 = \phi''''(v m_1) \sum_{p \in C} \delta_p w_{pm_1}$$

$$\delta m_2 = \phi''''(v m_2) \sum_{p \in C} \delta_p w_{pm_2}$$

$$\delta m_3 = \phi''''(v m_3) \sum_{p \in C} \delta_p w_{pm_3}$$

$$\delta m_4 = \phi''''(v m_4) \sum_{p \in C} \delta_p w_{pm_4}$$

$$\delta m_5 = \phi''''(v m_5) \sum_{p \in C} \delta_p w_{pm_5}$$

$$\delta m_6 = \phi''''(v m_6) \sum_{p \in C} \delta_p w_{pm_6}$$

Question 5

Draw the RNN Neural Network for the following Scenario

In the lockdown you are studying neural network and trying to solve some unsolved problems. You take the course CSE425 and tried to understand the neural networks more closely. The course is comprised with different neural networks. As you are unaware where and how you should start you asked your faculty to point out the heads up and the faculty provides you the guideline which you can use first and which way you can proceed the course.

Study material :

Perceptron (p) < MLP < RNN < CNN < KNN < Transformer < GAN < BERT

As you are a lazy learner you didn't study properly during the course. The final comes up and before the final you are trying to study. Before the finals in the previous day you segmented the hours and clock alarms with you. As you have limited time you decided that you will study the full day in a manner that are easy to understand to study first and then gradually move upwards. Because your mother is anxious about you and when you get sleep by studying any material she visits you and awake you up. When she awakes you up you change your t-shirt and keep studying in the sequence. You have T-shirt colored (black <white< red <yellow) with respective brightness mostly yellow and less black. So decide what will be the input output of the following scenario of studying with colourful t-shirts. You can test your designed RNN with some specific input output. Necessary calculations need to be shown. Can you suggest at last hour which

↥ Add file

Clear form