Total No. of Questions:[06]

Total No. of Printed Pages:[03]

PRN. No.	

PAPER CODE	

MAY 2022 - ENDSEM EXAM

FINAL YEAR B. TECH. (COMPUTER ENGINEERING)

(SEMESTER - II)

COURSE NAME: PROFESSIONAL ELECTIVE-IV

[ADVANCED MACHINE LEARNING]

COURSE CODE: CSUA40181B

(PATTERN 2018)

Time: [1 Hr]

[Max. Marks: 30]

Q.1) a) After training a neural network, you observe a large gap between the training accuracy (100%) and the test accuracy (42%). Choose the method to reduce this gap. [4]

Explanation of regularization.

b) Justify the need of Gradient Descent in neural network with detailed explanation of its working [6]

Explanation of Gradient descent working with its types.

OR

Q.2) a) "Mini batch gradient descent is faster than gradient descent." Justify the statement. [4]

Explanation and difference between both.

b) "When should one use L1, L2 regularization instead of dropout layer, given that both serve same purpose of reducing overfitting? Justify your answer. [6]

Explanation of L1, L2 regularization and dropout.

Q.3) a) Explain following terms

i)Deep Dream

ii) Deep Art [4]

Short explanation of both terms

b) Compare various architectures of CNN. [6]
Explanation of LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet.

OR

Q.4) a) "Convolutional Neural Network (CNN) work better with image data." Give justification. [4]

Instead of looking at the entire image, convolution helps to just read the image in parts. Doing this for a 300 x 300 pixel image would mean dividing the matrix into smaller 3 x 3 matrices and dealing with them one by one.

b) Given a Convolutional Neural Network having three different convolutional layers in its architecture as –
in Layer 1 - 10 filters of 3X3, stride 1 and no padding,
in layer 2 -20 filters of 5 X 5 with stride 2 and no padding,
in layer 3- 40 filters of 5 X 5 with stride 2 and no padding,
If 39 X 39 3-D image pass as input to this network, then estimate the dimension of the vector after passing through a fully connected layer in the architecture. [6]

Here we have the input image of dimension 39 X 39 X 3 convolves with 10 filters of size 3 X 3 and takes the Stride as 1 with no padding. After these operations, we will get an output of 37 X 37 X 10.
We then convolve this output further to the next convolution layer as an input and get an output of 7 X 7 X 40. Finally, by taking all these numbers (7 X 7 X 40 = 1960), and then unroll them into a large vector, and pass them to a classifier that will make predictions

Q.5) a) Distinguish BPTT and truncated BPTT [4]

Standard Backpropagation requires saving all the inputs and the activation outputs in forward pass for use in gradient computation in the backward pass. This can be computationally expensive and memory intensive, particularly in character language models (longer sequence of inputs than word based models - so more computation and memory requirement)

In truncated backprop through time (TBPTT), the input is treated as fixed length subsequence. For forward pass, the hidden state of previous subsequence is passed on as input to the next subsequence. However, in gradient computation, the computed gradient values are dropped at the end of every subsequence as we walk back.

b) "RNNs work better with text data". Give justification. [6]

Explanation of suitability of RNN for such problems.

OR

Q.6) a) Compare GRU and LSTM. [4]

Explanation of GRU and LSTM and reason why it is faster than LSTM.

The few differencing points are as follows:

i. The GRU has two gates, LSTM has three gates.
ii. GRU does not possess any internal memory, they don't have an output gate that is present in LSTM.
iii. In LSTM the input gate and target gate are coupled by an update gate and in GRU reset gate is applied directly to the previous hidden state. In LSTM the responsibility of reset gate is taken by the two gates i.e., input and target.

b) Which of the following activation functions can lead to vanishing gradients?
i)ReLU
ii) Tanh
iii) Leaky ReLU
Justify your answer with explanation of its working. [6]

Explanation of Tanh activation function.