Total No. of printed pages = 3

CSE 1818 PE 63

16/6/22

Roll No. of candidate ☐☐☐☐☐☐☐☐☐

2022

B.Tech. (CSE) 8th Semester End-Term Examination

CSE

NEURAL NETWORK AND DEEP LEARNING

(New Regulation 2017-18 & New Syllabus 2018-19)

Full Marks – 70   Time – Three hours

The figures in the margin indicate full marks for the questions.

1.   Answer the following questions : $(10 \times 1 = 10)$

   (i)   In CNN, which layer is responsible to perform dimensionality reduction?

   (a)   Convolutional Layer   (b)   Fully Connected Layer

   (c)   Pooling layer   (d)   All of the above

   (ii) Does the size of the feature map reduce upon applying a filter of size 1×1?

   (a)   Yes   (b)   No

   (iii) The larger the Strides the ————— the feature map.

   (a)   Larger   (b)   Smaller

   (c)   Size remains same   (d)   None of the above

   (iv) Which of the following is/are limitations of deep learning?

   (a)   Data labeling   (b)   Obtain huge training datasets

   (c)   Both (a) and (b)   (d)   None of the above

   (v)   In which of the following applications can we use deep learning to solve the problem?

   (a)   Protein structure prediction

   (b)   Prediction of chemical reactions

   (c)   Detection of exotic particles

   (d)   All of the above

[Turn over

(vi) In an MLP, the number of nodes in the input layer is 10 and the hidden layer is 5. The maximum number of connections from the input layer to the hidden layer are

(a) 50

(b) less than 50

(c) more than 50

(d) It is an arbitrary value

(vii) Which of the following neural network training challenge can be solved using batch normalization?

(a) Over fitting

(b) Restrict activations to become too high or low

(c) Training is too slow

(d) Both (b) and (c)

(viii) In a simple MLP model with 8 neurons in the input layer, 5 neurons in the hidden layer and 1 neuron in the output layer. What is the size of the weight matrices between input-hidden layer and hidden-output layer?

(a) $[1 \times 5]$, $[5 \times 8]$

(b) $[8 \times 5]$, $[5 \times 1]$

(c) $[5 \times 8]$, $[5 \times 1]$

(d) $[8 \times 5]$, $[1 \times 5]$

(ix) Which of the following functions can be used as an activation function in the output layer if we wish to predict the probabilities of n classes (p1, p2..pk) such that sum of p over all n equals to I?

(a) Softmax

(b) ReLu

(c) Sigmoid

(d) Tanh

(x) The C_1 curve above denotes tracing accuracy w.r.t. each epoch in a deep learning algorithm. Both the C_2 and C_3 curves denote validation accuracy. Which of these indicate overfitting?
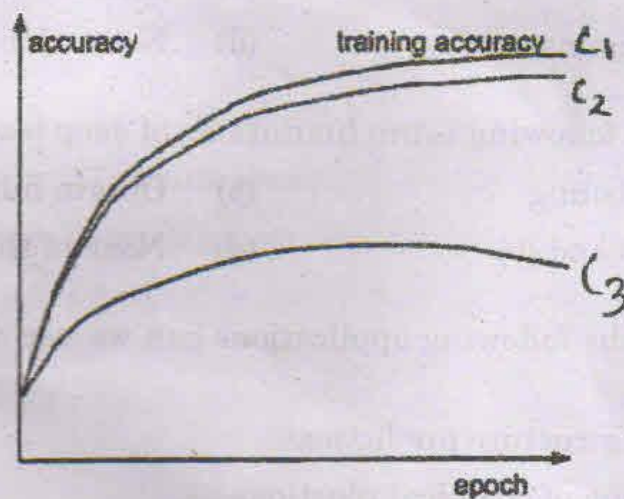
(a) C_2 curve

(b) C_3 curve

Answer the following questions : (Attempt only *four* questions) :

2. (a) Write a simple Multilayer Perceptron program or a pseudocode(without using the MLP library function) having two input nodes, one hidden layer (with 4 nodes in it) and two nodes at the output layer. (12 + 3 = 15)

   (b) What is over-fitting and under-fitting?

3. (a) Explain the gradient descent algorithm. (8 + 7 = 15)

   (b) What is the need of an activation function? Distinguish between sigmoid, rectified linear unit and softmax activation function.

4. (a) Explain in detail, how to perform Principal Component Analysis (PCA). (8 + 7 = 15)

   (b) What are autoencoders used for? How to regularize autoencoders?

5. (a) Below is a diagram of a small convolutional neural network that converts a 13 .× 13 image into 4 output values. The network has the following layers/operations from input to output: convolution with 3 filters, max pooling, ReLu, and finally a fully-connected layer. For this network we will not be using any bias/offset parameters. Please answer the following questions about this network. (8 + 7 = 15)
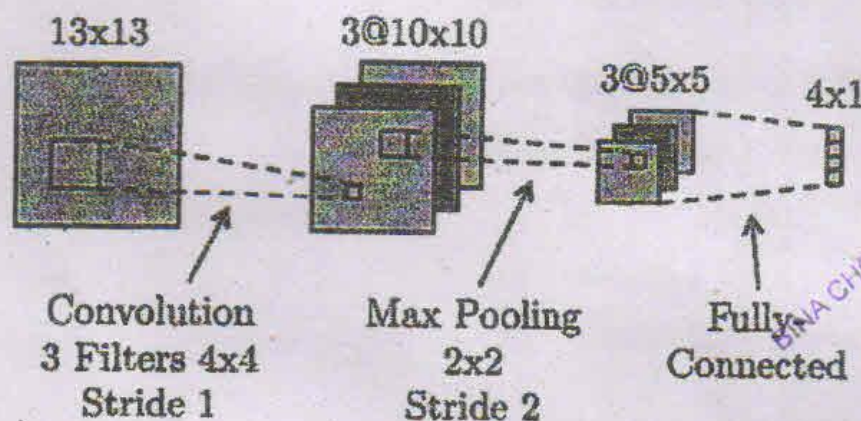
   (i) How many weights in the convolutional layer do we need to learn?

   (ii) How many ReLu operations are performed on the forward pass?

   (iii) How many weights do we need to learn for the entire network?

   (iv) If the input image has been converted into a matrix of size 28 × 28 and a kernel/filter of size 7 × 7 with a stride of 1 is used, what will be the size of the convoluted matrix?

   (b) What is the significance of a Pooling layer? What are the different hyperparameters of a pooling layer? Explain the different types of pooling layer.

CSE 1818 PE 63   3

6. (a) What is the meaning of valid padding and same padding in CNN?

$(4 + 4 + 4 + 3 = 15)$

   (b) What is Dropout and Batch Normalization?

   (c) What are Vanishing and Exploding Gradients?

   (d) What are Hyperparameters?

7. Write short notes on (any *three*) :

$(3 \times 5 = 15)$

   (a) Bias-Variance Tradeoff

   (b) Batch Normalization

   (c) Ensemble methods

   (d) Long Short-Term Memory

   (e) Encoder Decoder Models

   (f) Covariance and Correlation

   (g) Pre-trained network