ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

ORGANISATION OF ISLAMIC COOPERATION (OIC)

Department of Computer Science and Engineering (CSE)

SEMESTER FINAL EXAMINATION **SUMMER SEMESTER, 2021-2022**

DURATION: 3 HOURS **FULL MARKS: 150**

CSE 4835: Pattern Recognition

Programmable calculators are not allowed. Do not write anything on the question paper.

Answer **all 6 (six)** questions. Figures in the right margin indicate full marks of questions whereas corresponding CO and PO are written within parentheses.

---

1. a) Suppose your friend has trained a deep neural network using stochastic gradient descent optimizer and L2 weight regularization with four different learning rates and plotted the corresponding loss curves (as shown in Figure 1). Unfortunately, he has forgotten which curve belongs to which learning rate. However, he remembers that the learning rates are $3 \times 10^{-4}, 4 \times 10^{-1}, 2 \times 10^{-5}$, and, 8×10^{-3}. Assign each of the learning rate values to the probable corresponding curve (A/B/C/D) and explain your choice. 5 (CO2) (PO2)

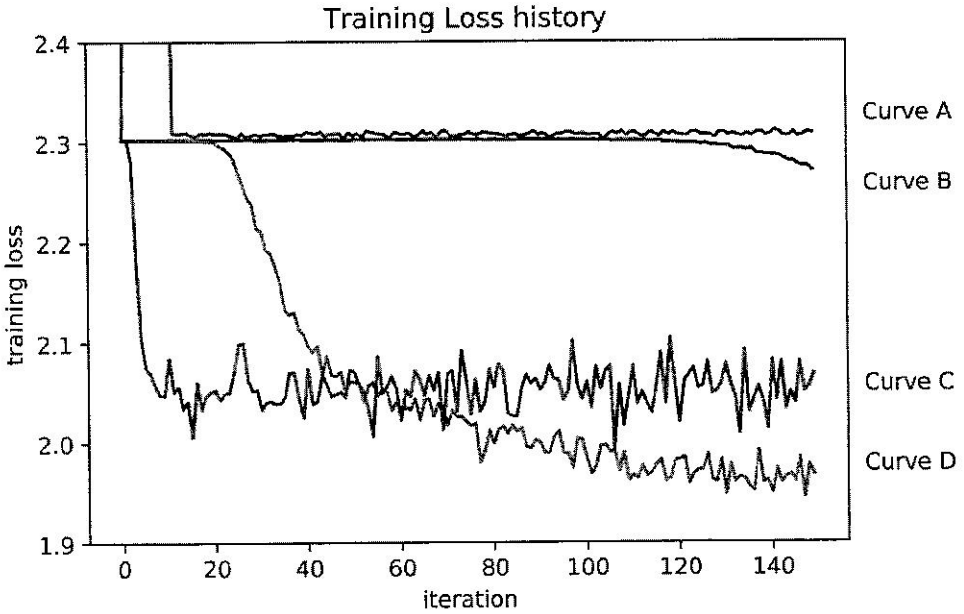

Figure 1: Learning curves for Question 1.a)

b) Consider the following functions and provide comments on the suitability of each of them to be used as an activation function for training a neural network. 6 (CO2) (PO2)

i. $f(x) = -\min(2, x)$

ii. $f(x) = 0.9x + 1$

iii. $f(x) = \begin{cases} \min(x, 0.1x) & x \geq 0 \\ \min(x, 0.1x) & x < 0 \end{cases}$

iv. $f(x) = \begin{cases} \max(x, 0.1x) & x \geq 0 \\ \min(x, 0.1x) & x < 0 \end{cases}$

c) Suppose you want to train a binary image classifier, using mini-batch gradient descent that can distinguish between the images of cats and dogs. You have already split your dataset into train, validation, and test sets. The classes are balanced. You realized that within the training set, images are ordered in such a way that all the dog images come first and all the cat images come after. 4 (CO2) (PO2)

Identify the problem in this scenario and provide a solution. Will full batch gradient descent suffer from the same problem?

d) Suppose you want to evaluate the classifier trained in Question 1.c) Your test set (X_{test}, Y_{test}) is such that the first few images are of dogs, and remaining images are of cats. You evaluated your model on the test set and achieved $a_1\%$ accuracy. After that, you shuffled X_{test} and Y_{test}, and evaluated your model on it and obtained a classification accuracy $a_2\%$. What is the relationship between a_1 and a_2 $(>, <, =, \geq, \leq)$? Explain. 4 (CO2) (PO2)

e) You are training a large feed-forward neural network (100 layers) on a binary classification task, using a *sigmoid* activation in the final layer, and a mixture of *tanh* and *ReLU* activations for all other layers. You notice that, your weights to a subset of your layers stop updating after the first epoch of training, even though your network has not yet converged. Deeper analysis reveals the gradients to these layers completely, or almost completely, go to zero very early on in training. Choose from the following fixes that may improve the scenario. Justify your answer. (Note that your loss is still within a reasonable order of magnitude). 6 (CO2) (PO2)

- Increase the size of your training set
- Switch the *ReLU* activations with leaky *ReLUs* everywhere
- Add Batch Normalization before every activation
- Increase the learning rate

2. a) Suppose a pattern recognition system has been designed to predict whether a given sample is diseased (1/2/3) or healthy. The system has been tested with three unseen samples and has produced scores for each of the four classes. Table 1 contains the scores of different classes for the samples. 12+2 (CO1) (PO1)

Table 1: Scores of different samples for Question 2.a)

	$Sample_1$	$Sample_2$	$Sample_3$
Healthy	1.3	**1.1**	0.3
$Disease_1$	**-0.9**	3.2	-0.9
$Disease_2$	3.7	2.4	**1.35**
$Disease_3$	1.2	0.7	0.9

Based on the provided information, answer the following:

i. Calculate the overall loss of the model using two separate functions: 'Multiclass SVM loss' and 'Cross-Entropy loss'.

ii. Compare the characteristics of these two ways of calculating loss value.

b) Recommend actions on how you can utilize transfer learning with a CNN-based architecture pretrained on the ImageNet dataset in scenarios mentioned in Table 2. 8 (CO2) (PO2)

Table 2: Scenario for Question 2.b)

	Your Dataset similar to ImageNet	**Your Dataset very different from ImageNet**
You have very little data (<1000 samples)	$Scenario_1$	$Scenario_2$
You have quite a lot of data (>=1000)	$Scenario_3$	$Scenario_4$

c) Discuss the role of Dropout and DropConnect techniques in improving the generalization of a network. 8 (CO1) (PO1)

3 a) Given the neural network with fully connected layer and ReLU activations, including two input units (i_1, i_2), four hidden units (h_1, h_2) and (h_3, h_4). The outputs are indicated as (o_1, o_2) and their ground truth targets are indicated as (t_1, t_2). The weights and bias of fully connected layer are called w and b with specific sub-descriptors. The network can be visualized as shown in Figure 2 and the values of the variables are given in Table 3. Consider the Mean Squared Loss, where, $Loss = \frac{1}{n}\sum_i (t_i - o_i)^2$. 8+10 (CO2) (PO2)

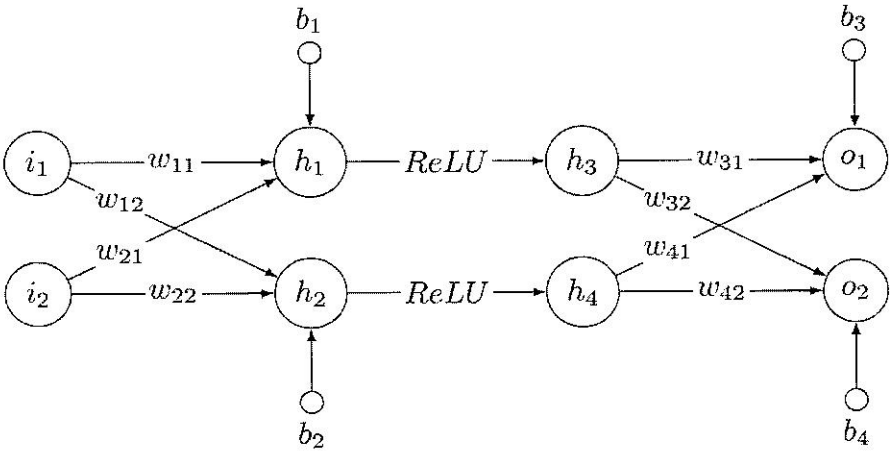

Figure 2: Sample neural network for Question 3.a)

Table 3: Necessary values for Question 3.a)

Variables	Values
i_1, i_2	$2.0, -1.0$
$w_{11}, w_{12}, w_{21}, w_{22}$	$1.0, -0.5, 0.5, -1.0$
$w_{31}, w_{32}, w_{41}, w_{42}$	$0.5, -1.0, -0.5, 1.0$
b_1, b_2	$0.5, -0.5$
b_3, b_4	$-1.0, 0.5$
t_1, t_2	$1.0, 0.5$

Based on this information, answer the followings:

i. Write the computations of each step in the forward pass, and calculate the Mean Squared Loss.

ii. Update the weight, w_{21} using gradient descent using learning rate 0.1.

b) Address the limitation of random weight initialization in training deep neural networks and discuss how Xavier initialization method tackles this issue. 5 (CO1) (PO1)

4. a) Apply basic convolution operation using the kernel, W with a stride of $s = 2$ and padding of $p = 1$ to the input, X and fill the missing values of the output, Z. 5 (CO1) (PO1)

3	1	6	8	0
2	3	2	7	-6
-8	0	6	-5	-8
1	2	5	-2	-4
8	7	-2	-1	7

X

$1/9$	$1/9$	$1/9$
$1/9$	$1/9$	$1/9$
$1/9$	$1/9$	$1/9$

W

1		1
0		
	1	0

Z

b) Discuss the role of Stem Network, Inception Module, and Global Average Pooling in improving the efficiency of GoogleNet architecture. 12 (CO1) (PO1)

c) Figure 3 shows the basic architecture of a bottleneck residual block. Discuss how the ResNeXt architecture adopts this block with the concept of 'parallel pathways'. 8 (CO1) (PO1)

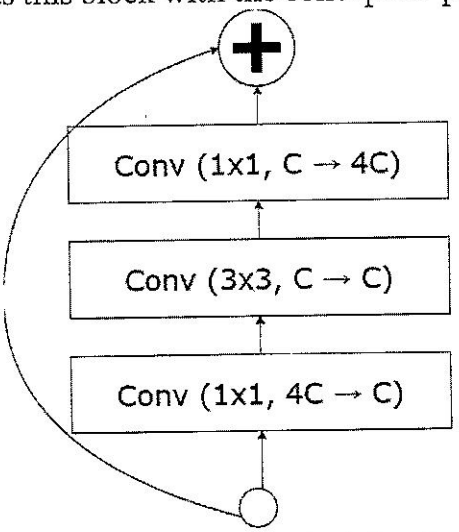

Figure 3: Bottleneck Residual Block for Question 4.c)

5. a) Explain the intuition behind designing the Squeeze-and-Excitation networks with the help of Figure 4. 7 (CO1) (PO1)

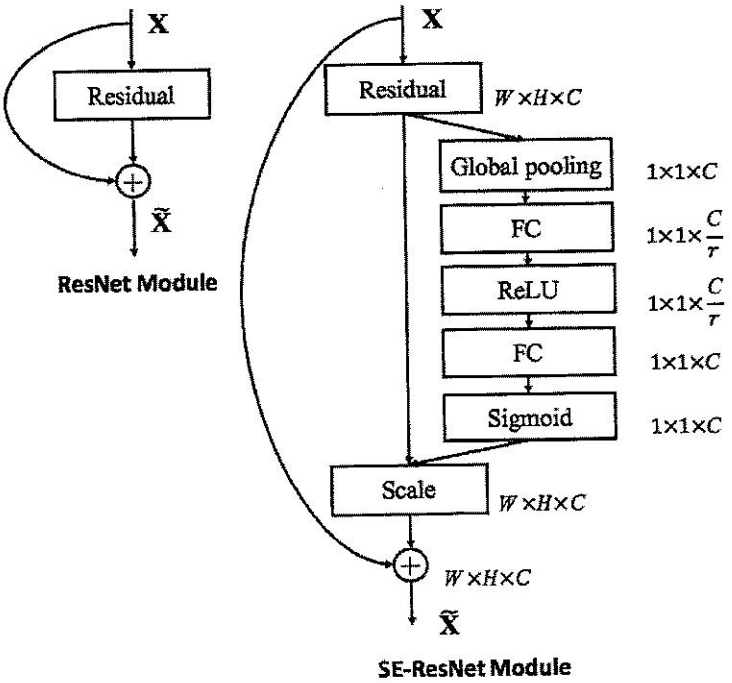

Figure 4: Squeeze-and-Excitation module for Question 5.a)

b) Figure 5 represents the LeNet-5, the 'first architecture' for CNNs (especially when trained on the MNIST dataset, an image dataset for handwritten digit recognition). The architecture is small and easy to understand, yet large enough to provide interesting results. Throughout the network, filters of size 5×5 has been used with stride value of 1 and padding value of 0. Here, each C_i denotes convolution operation, S_i represents pooling operation and F_i represents dense connection. 6×3 (CO3) (PO3)

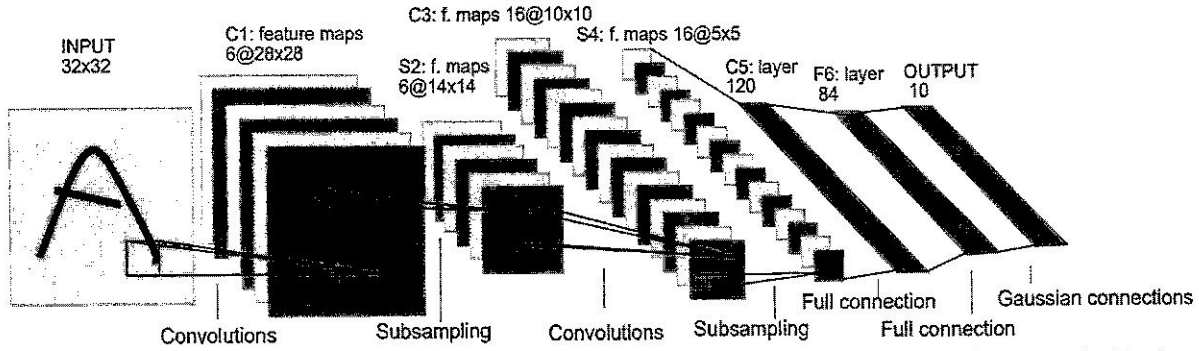

Figure 5: Architecture of LeNet-5 for Question 5.b), a Convolutional Neural Network. Each plane is a feature map, i.e., a set of units whose weights are constrained to be identical

Based on the scenario, answer the following questions:

i. Instead of taking 32×32 images as input, suppose you want to train the network to classify images of size 68×68. List two possible changes in the architecture to support this.

ii. Explain the concept of receptive fields. What is the receptive field of one pixel after the first maxpool operation? What is the receptive field of a neuron in the $C5$ layer.

iii. Instead of digits now you want to classify handwritten alphabetic characters (26 classes). Propose a few guidelines to improve the given architecture to achieve a high performance.

6. a) Two historians have approached you for your deep learning expertise. They want to classify images of historical objects into 3 classes (Antiquity, Middle Ages, and, Modern Era) depending on the time they were created. 3×4 (CO3) (PO3)

i. Over the last few years, they have collected nearly 5000 hand-labelled RGB images. Before training your model, you want to decide the input image resolution to be used. Why is the choice of image resolution important?

ii. Once you have figured out a good image resolution to use, how would you partition the dataset for training the model? Justify your choice.

iii. After visually inspecting the dataset, you realize that the training set only contains pictures taken during the day, whereas the test set only has pictures taken at night. Explain what is the issue and how you would correct it.

iv. As you train your model, you realize that you do not have enough data. Discuss how the 'Cutout' and 'Mixup' techniques can be utilized in regard to overcome the shortage of data.

b) In a Neural Network, each of the nodes in the hidden layer learns different representations of the data, combining which the model produces an appropriate prediction. Despite having no prior direction on what feature to look for, how does the network end up learning the relevant features? How does it know which node(s) should be triggered by what amount while predicting certain classes? 8 (CO2) (PO2)