37

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

ORGANISATION OF ISLAMIC COOPERATION (OIC)

Department of Computer Science and Engineering (CSE)

SEMESTER FINAL EXAMINATION — SUMMER SEMESTER, 2022-2023

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 5 (five)** questions. Figures in the right margin indicate full marks of questions with corresponding COs and POs in parentheses.

---

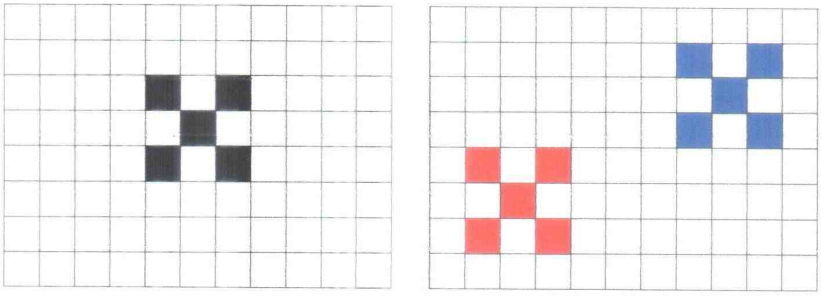
1. a) Suppose a dataset contains 10,000 RGB images of n different classes. A linear classifier was used to classify these samples correctly. To achieve better accuracy, K-fold cross-validation was performed. Each of the K-folds (f_1, f_2, ...,f_k) contained an equal number of images. Firstly f_1 was considered as the test set, f_2 as the validation set and all other $(k-2)$ folds as training set which led to the accuracy: acc_1. In the next iteration, f_2 was considered as the test set, f_3 as the validation set and all other $(k-2)$ folds as the training set leading to another accuracy: acc_2. In this way, the test set and validation set were changed k times leading to k accuracies ($acc_1, acc_2, ..., acc_k$). The final accuracy was claimed to be 95% by averaging all these acc_i values. — 10 (CO2) (PO2)

   Explain the effectiveness of this experimental method. How much can this result be trusted? Write your remarks with possible comments on improving the methodology (if any).

   b) You have access to an algorithm ($\star$) that can efficiently find the position of the maximum value in a matrix. Consider Figure 1 to answer the following questions: — 3 + 4 (CO2) (PO2)

      i. You have to detect the 3×3 cross in Figure 1(a). Design a filter to be convolved over the image. Given the output of this "convolved" operation, the algorithm ($\star$) should give you the position of the cross.

      ii. You have to detect the 3×3 red cross (bottom-left) in Figure 1(b) but not the blue cross (top-right). Design a filter to be convolved over the image. Given the output of this "convolved" operation, the algorithm ($\star$) should give you the position of the red cross.



(a) A black and white 3×3 cross **(b)** Red and blue 3×3 crosses

Figure 1: Image configuration for Question 1.b. 1(a) Image 1 is a binary image. Each pixel has value 1 (white) or 0 (black). There is only a single channel. 1(b) Image 2 is a color image. Each pixel has value 0 (absence of that color) or 1 (presence of that color), but there are 3 color channels (RGB). The cross on the bottom-left is red, and the cross on the top-right is blue.

   c) Keywords: {Score, Weight vector, Gradient Descent, Loss Function, Input data, Backpropagation, Regularization} — 5 + 5 (CO1) (PO1)

   Draw a proper flowchart by arranging the aforementioned keywords according to their roles in solving a classification problem. Discuss their roles and relations with each other in brief.

d) Consider the two-layer fully connected neural network given in Figure 2. All activations are sigmoid and the optimizer is stochastic gradient descent. All the weights and biases are initialized to zero and an input $x \in R^{(n \times 1)}$ is forward propagated within the network. What is the value of output $\hat{y}$? 3 (CO2) (PO2)

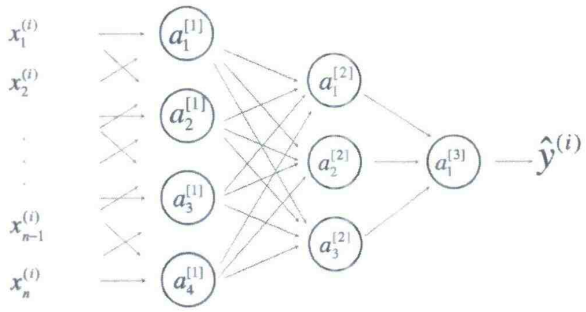


Figure 2: A two-layer Neural Network for Question Question 1.d

2. Suppose you have to classify images of size 256×256 into one of 1000 classes using the CNN architecture given in Figure 3.

- The values directly below the arrows indicate the filter sizes f of the corresponding convolution and pooling operations.
- s stands for Stride. If no stride is specified, a stride of $s = 1$ is used.
- All convolution and pooling layers use a padding of $p = \frac{f-1}{2}$, for corresponding filter size f.

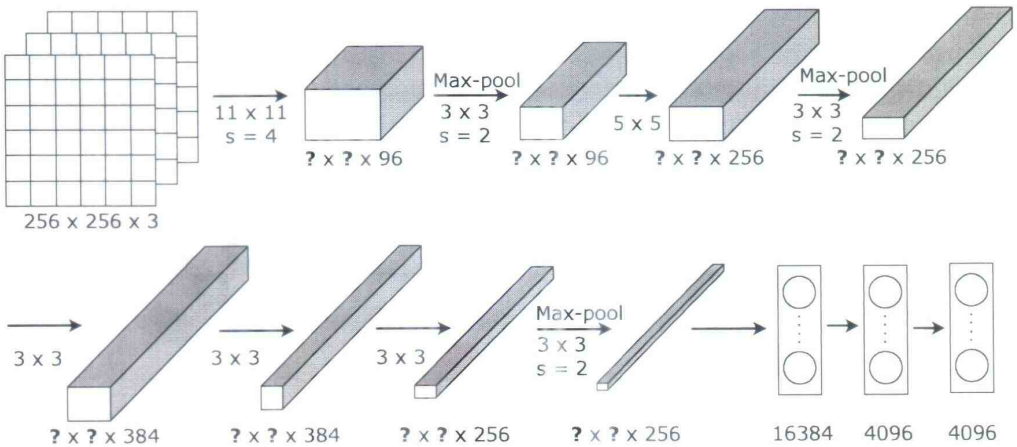


Figure 3: Custom CNN architecture for Question 2

a) Calculate the output feature map, parameter count for each of the layers of the CNN architecture. 6 × 2 (CO1) (PO1)

b) Suppose you have four choices of Activation functions — Sigmoid, ReLU, Tanh, and Softmax. Which one of them would you prefer/ not prefer to use in different layers of the network? Provide justification for each of them. 4 × 2 (CO2) (PO2)

c) Calculate the size of the combined receptive field of the first three layers. This corresponds to the area of pixels in the input image that each neuron after the third layer (before the second MAX-POOL) "sees". 3 × 2 (CO2) (PO2)

d) Criticize the architecture and propose three design improvements. 3 × 3 (CO3) (PO3)

e) Unfortunately, the dataset of the competition is rather small. However, you found a much larger image classification dataset on the internet containing similar images, but the set of classes is different. Explain how you could make use of the larger dataset and the overall training policy of the model. 5 (CO2) (PO2)

3. a) Draw the corresponding Computational Graph representation of the function given in Equation 1 in the most granular fashion. Showing detailed calculations, find the gradients of the function f with respect to the variables (x_1, x_2, y_1, y_2) using the Backpropagation algorithm. Consider $(x_1, y_1) = (2, 3)$ and $(x_2, y_2) = (1, 7)$. 2 + 8 (CO1) (PO1)

$$f(x, y) = \sqrt{(x_1 - x_2)^2 - (y_1 - y_2)^2} \tag{1}$$

b) Discuss the concept of space warping in learning complex decision boundaries with necessary illustrations. 6 (CO2) (PO2)

c) Discuss the role of Dropout as a regularizer. Justify the reason behind dropout layers being commonly used at the end of the network in between the fully connected layers, but not in between the convolutional layers. 5 + 5 (CO2) (PO2)

d) Algorithm 1 mentions the basic idea of the Adam optimizer. 5 × 2 (CO1) (PO1)

i. Explain the role of the $moment1$ and $moment2$ terms in ensuring better convergence.

ii. Identify the limitation of the given algorithm and propose an improved solution.

Algorithm 1 An optimization algorithm for Question 3.d

$moment1 = 0$
$moment2 = 0$
while $t \neq numSteps$ **do**
  $dw = compute\_gradient(w)$
  $moment1 = \beta_1 \times moment1 + (1 - \beta_1) \times dw$
  $moment2 = \beta_2 \times moment2 + (1 - \beta_2) \times dw$
  $w -= learningRate \times \frac{moment1}{\sqrt{moment2} + 10^{-7}}$
end while

4. a) Explain the role of the 'Additive shortcut' of the Residual Networks (ResNets) in training very deep CNN models. How does the ResNet-50 architecture offer better performance compared to ResNet-34 despite having less computation cost? 5 + 5 (CO1) (PO1)

b) Explain strategies undertaken by the MobileNet architecture to provide higher performance with lower computational constraints. 8 (CO1) (PO1)

c) Discuss the limitations of vanilla Recurrent Neural Networks (RNNs). How can Long Short-Term Memory (LSTM) architecture improve the scenario? 5 + 5 (CO1) (PO1)

5. a) Team 'Jibananda Das' consisting of Arifa, Rafia, and Syema has decided to compete in the $11\frac{1}{2}^{th}$ National ICT Fest. This time the organizers have launched a 'meme classification' challenge on a highly imbalanced dataset. Syema wishes to train their model 'hoccheJeNet' (a tribute to the famous phrase of Rafia) for different combinations of hyperparameters that they learned in the Pattern Recognition course. Hyperparameters like learning rate, epochs, etc., have so many options to be explored. However, they can't occupy the computer for the entire day as their friend Anik is nagging for a GPU slot to save his thesis, and the president is 5 × 2 (CO2) (PO2)

consistently pushing them to leave the PC to work in the Fest. Hence, they want something like an auto-tuner for the aforementioned hypermeters!

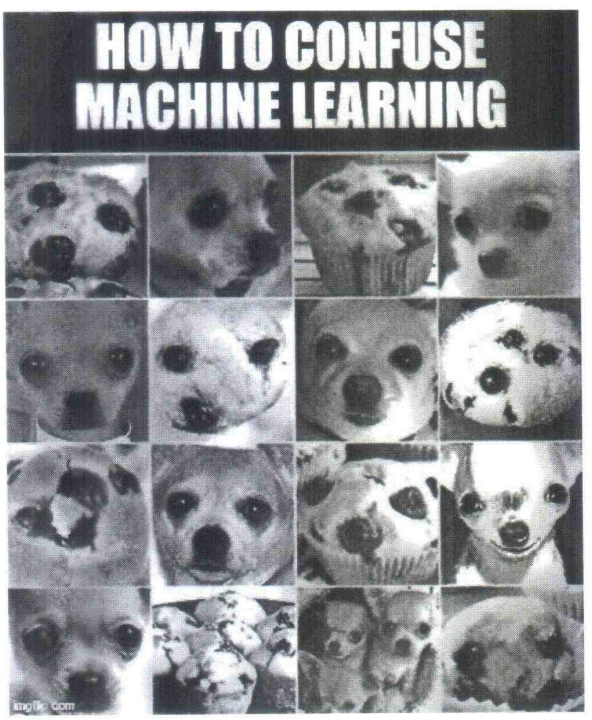
Besides, Arifa is suggesting to augment the entire dataset to enhance the number of samples and then training the model. Whereas, Rafia is asking to separate the test set first, and then augment the dataset for training. Syema is confused as both of these strategies seem similar.

i. What strategy can they adopt to achieve the optimized learning rate and the number of epochs?

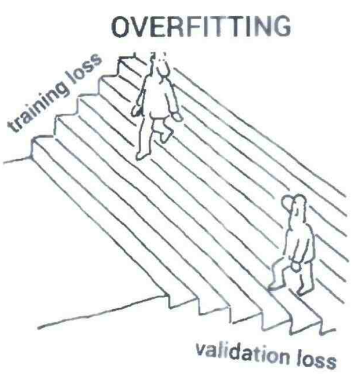
ii. Which augmentation strategy will provide higher performance? Justify your answer.

b) Explain the hidden concepts in the memes provided in Figure 4 using the knowledge of Pattern Recognition. 3×2 (CO1) (PO1)

(a) Meme1



(b) Meme2



(c) Meme3

Figure 4: Samples memes for Question 5.b.