B.Sc. Engg. CSE 8th Semester — 04 April 2022

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)
ORGANISATION OF ISLAMIC COOPERATION (OIC)
Department of Computer Science and Engineering (CSE)

SEMESTER FINAL EXAMINATION — **SUMMER SEMESTER, 2020-2021**
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. Marks of each question and corresponding CO and PO are written in the right margin with brackets.

---

1. a) Suppose a dataset contains 10,000 RGB images belonging to n different classes. A linear classifier was used to correctly classify these samples. To achieve better accuracy, K-fold cross-validation was performed. Each of the K-folds ($fold_1, fold_2, \dots\dots, fold_k$) contained an equal number of images.
Firstly $fold_1$ was considered as the test set, $fold_2$ as the validation set and all other $(k-2)$ folds as training set which led to the accuracy: $'acc_1'$. In the next iteration, $fold_2$ was considered as the test set, $fold_3$ as the validation set and all other $(k-2)$ folds as training set leading to another accuracy: $'acc_2'$. In this way, the train, test and validation splits were exchanged k times leading to k accuracies $(acc_1, acc_2, \dots\dots, acc_k)$. The final performance was claimed to be 95% by averaging all these acc_i values.
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). — 12 (CO2) (PO4)

   b) $Keywords$: $\{Score, Weight\ vector, Gradient\ Descent, Loss\ Function, Input\ data, Backpropagation, Regularization\}$
Draw a flow-chart by arranging the *keywords* mentioned above according to their roles in solving a *classification problem*. Mention their roles & relation with each other properly in the chart. — 13 (CO2) (PO3)

2. a) Why does a Convolutional Neural Network (CNN) work exceptionally well on image data compared to traditional Neural Network (NN)? Is there any scenario where the traditional NN can have a similar level performance like CNN? — 3+4 (CO1) (PO1)

   b) What is the role of *filters* in CNNs? How does it know what to look for in an image? Suppose, a Deep CNN model has been successfully trained on a *Face Emotion Recognition* dataset. What do you expect the filters in different layers to learn? — 3+3+5 (CO3) (PO2)

   c) Mention the benefits of introducing Batch Normalization (BN) in a CNN architecture. What is the intuition behind introducing *Learnable parameters* in the equation of BN? — 5+2 (CO1) (PO1)

3. a) i. How can *Max-Pooling* introduce shifting invariance in a CNN architecture? — 5 (CO1) (PO1)

      i. Usually, the filters in pooling operation move in the height and width dimensions; not in the channel dimension. Suppose, a new type of pooling operation is introduced which will move in the channel dimension as well. Discuss on the usefulness on this idea. — 5 (CO3) (PO2)

      ii. With proper justification, mention a scenario where *Average-pooling* is found to be a better choice than *Max-pooling*. — 5 (CO3) (PO2)

   b) Explain drawbacks of Stochastic Gradient Descent (SGD) algorithm. How can the concept of SGD+Momentum improve the scenario? — 6+4 (CO1) (PO1)

4. Figure 1 presents 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.

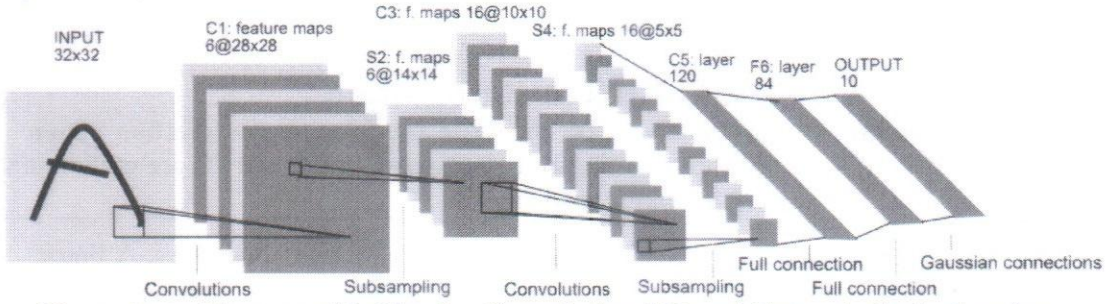

Figure 1: Architecture of LeNet-5, a Convolutional Neural Network. Each plane is a feature map, i.e., a set of units whose weights are constrained to be identical

Based on the scenario, answer the following questions:

a) Showing detailed steps, compute the total number of trainable parameters, memory requirements and Flops count involved in training the LeNet-5 architecture. (Ignore the bias terms) [Note: $'C_5'$ is the result of applying filters of size 5×5 on the activation map provided by $'S_4'$.] — 15 (CO1) (PO1)

b) If this architecture is needed to the applied on a complex dataset, LeNet-5 often fails to provide high performance. Propose a few guidelines to modify the vanilla architecture that might lead to a better performance in such scenario. — 10 (CO2) (PO4)

5. a) AlexNet is a CNN architecture which won the 2012 ImageNet competition with a top-5 error rate of 15.3%, compared to the second-place top-5 error rate of 26.2%. It is the first convolutional network which used GPU to boost performance. The architecture consists of 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 Softmax layer. The authors used different filter sizes like 3x3, 5×5, 11×11 in the convolution layers and 3×3 in the pooling layers. — 5×3 (CO3) (PO2)

On the other hand, the Visual Geometry Group (VGG) Networks won the competition in 2014 with a top-5 error rate of 7.9%. The authors followed three basic design principles throughout the networks which are:

i. Instead of using different filter sizes like AlexNet, all convolution operations use filters of size 3×3, stride and padding value of 1.
ii. All max pooling operations use 2×2 filters with stride value of 2.
iii. After each pooling operation, the number of channels should be double.

Justify the intuitions behind each of the design principles of VGG networks and discuss on how they contributed to the overall improvement in performance.

b) How does 1×1 convolution work? Explain the usefulness of this operation in the different positions of the Inception Module mentioned in Figure 2. 2+8 (CO3) (PO2)

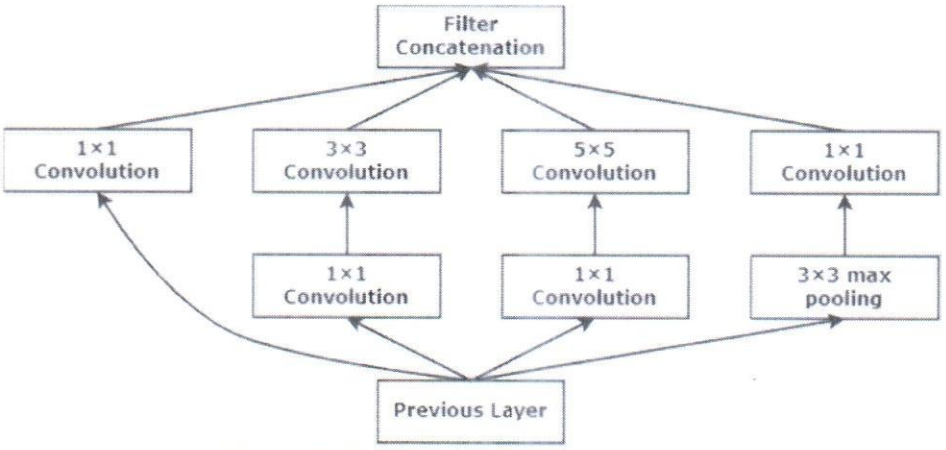

Figure 2: GoogLeNet - Inception Module

6. a) Explain the role of 'Additive shortcut' of the Residual Networks (ResNets) in training very deep CNN models. How does a ResNet-50 architecture offer better performance compared to ResNet-34 despite having less computation cost? 5+5 (CO2) (PO2)

b) Discuss on how the concept of *Depthwise Separable Convolution* can be applied on the Network mentioned in Figure 3. (Assume that the height and width of the image is denoted by 'H' and 'W'). 8 (CO2) (PO2)

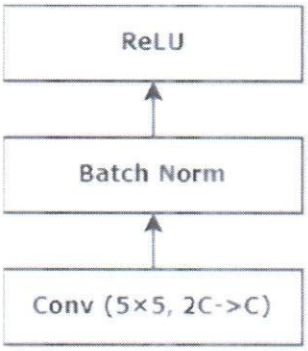

Figure 3: Standard Convolution Block

c) Draw the *Computational Graph* for the following equation and derive the necessary equations to calculate the gradients of 'x' and 'y'. 3+4 (CO1) (PO1)

$$f(x, y) = \frac{x + \sigma(x)}{\sigma(x) + (x + y)^2}$$