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

SEMESTER Final EXAMINATION   **SUMMER SEMESTER, 2019-2020**
DURATION: 1 Hour 30 Minutes   **FULL MARKS: 75**

CSE 4621: Machine Learning

Programmable calculators are not allowed. Do not write anything on the question paper.
There are **5 (five)** questions. Answer **4 (four)** of them including Questions **1, 2 & 3**.
Figures in the right margin indicate marks.

Instructions:

i. **Closed Book**: No books (hardcopy or softcopy), notes, slides, cheat-sheets are allowed. No internet browsing.
ii. Use standard loose pages, e.g., Legal or A4 (recommended).
iii. Write your Student ID, Course Code, Course Name, Exam Type, Date, & Semester at the top of the first page.
iv. All other pages should include student ID & page number at the top.
v. Scan all pages into a single PDF file names as **StudentID_CSE4621**.
vi. Make sure images are clear in contrast and intelligible.
vii. Upload your script into the Google classroom as Assignment submission.
viii. You will be given extra 15 minutes to upload at the end of the exam.
ix. In case of difficulties during submission, contact the invigilator, and if approval is given submit through email at **hasanul@iut-dhaka.edu**
x. For other technical difficulties contact the invigilator of the room or the course teacher.
xi. No break is allowed during exam time.

(Mandatory)

1. a) Suppose, all weights of your multi-layer neural network are initialized with the same value. After weight updates, what kind of problem will you be facing? Explain with the help of a three-layer neural network. What is the remedy to that problem? **4+3**

   b) Suppose the following figure represents the forward calculation involved in a feed-forward neural network with a logistic function, σ. **6+1**
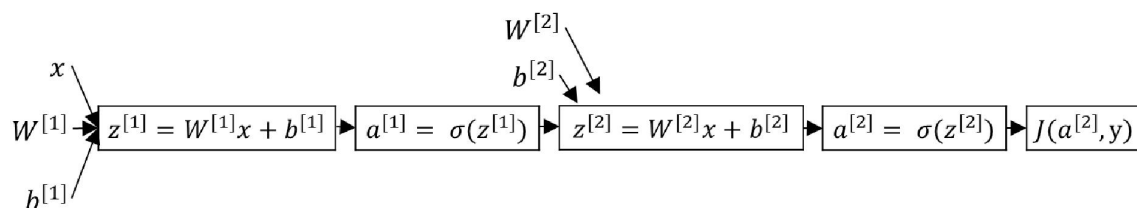

Figure 1.

Here, $J(\hat{y}, y)$ is the log-loss cost function. Derive the mathematical expression of the derivative: $\dfrac{dJ}{dz^{[1]}}$.

Give the weight update equation for $W^{[1]}$.

   c) Why is the cross-validation (CV) dataset required? How can CV data be used for model selection? Explain with the help of a learning curve. Identify the high-bias and high-variance regions too. **2+6**

   d) Suppose your trained model is giving less error for the training dataset, but gives large error for test data. What measures can you take to improve model performance? **3**

(Mandatory)

2. a) Why is convolution operation adopted in neural network? Describe how volume convolution operation is carried out in an $n_H \times n_W \times n_C$ dimensional input data with stride $s=1$ and padding value p.   3+7

b) Deep neural networks are sometime harder to train because of exploding and vanishing gradient problems. What are the reasons behind those aforesaid problems? How can you solve them by changing the network architecture? Draw appropriate network connection.   2+6

c) During decision tree generation for classification, instead of taking a binary split for the numeric attribute, can we use ternary split using two thresholds w_{ma} and w_{mb}? In other words, three potential branches where samples can take j-th branch according to the following conditions: 5+2

$$x_j < w_{ma} \, ; w_{ma} \leq x_j \leq w_{mb} \, ; x_j > w_{mb}$$

Propose a modification of the tree induction method along with impurity measure to learn those two thresholds. What are the advantages of performing ternary split over binary?

(Mandatory)

3. a) Compare between Generative and Discriminative models.   3

b) Derive the decision boundary equation for naïve Bayes classifier. "Naïve Bayes classifier in general is not linear" - Explain why.   2+5

c) What is Kernel trick in Support Vector Machine (SVM) classifier? How does it work?   5

4. a) In Principal Component Analysis, is it possible to have an average squared projection error of zero with k principal components, where k is less than the original feature number n? Describe in which cases this is possible and why.   5

b) "PCA is not a good way to address overfitting"- Do you agree or disagree? Justify your answer.   5

5. a) What is a Cluster? Describe the major considerations for Cluster Analysis.   5

b) "K-medoids clustering algorithm produces nonconvex shaped clusters" - Do you agree or disagree? Justify your answer.   5