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

MID SEMESTER EXAMINATION **SUMMER SEMESTER, 2020-2021**

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.

Answer **all 3 (three)** questions. Marks of each question and corresponding CO and PO are written in the right margin with brackets. Symbols have their usual meaning.

---

1. a) Briefly describe the basic steps in machine learning. Compare among the major types of learning approaches. 7+5 [CO2, PO2]

   b) Suppose you have a friend who is an esteemed medical doctor. Your friend has a dataset containing gene expression data for patients. For each patient, there is a vector containing the expression levels of 10, 000 genes, along with a label indicating whether the patient developed a certain disease. Using this, your goal is to learn a model where, given a new patient (also represented by a similar 10, 000-dimensional vector), the model can output the probability of this patient developing the disease. Your friend also believes a reasonable model will mainly depend on at least 9,000 of the genes. Which machine learning model will you design here? Justify your model with the choice of regularization parameter. 5 [CO3, PO3]

   c) Define the log-loss function $J(\theta)$ for a univariate Logistic Regression problem. Mathematically show that this cost function is convex. 1+7 [CO1, PO1]

2. a) Identify why is feature scaling important before applying the gradient descent technique? Draw necessary illustrations. 7 [CO2, PO2]

   b) Compare between Batch- and Stochastic- gradient descent techniques. 4 [CO2, PO2]

   c) Suppose your linear regression model is suffering from high training error and also high test error. Do you think increasing the number of training data size will help? Explain with the help of learning curve. Choose solutions which can increase the model performance? 4 [CO3, PO3]

   d) Given, the regression cost function $J(\theta)$ with Lasso regularization as below: 8+2 [CO1, PO1]

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(\theta^T x^{(i)} - y^{(i)}\right)^2 + \lambda\sum_{j=1}^{n}\left|\theta_j\right| \quad \text{where } \lambda > 0.$$

   Derive the vectorized closed form expression (normal equation) for the weight vector $\hat{\theta}$ that minimizes the cost function $J(\theta)$. When shall we avoid the normal equation solution to regression problem?

3. a) Suppose the following figure represents the forward calculation involved in a feed-forward neural network with a logistic function, σ. 8+2 [CO1, PO1]

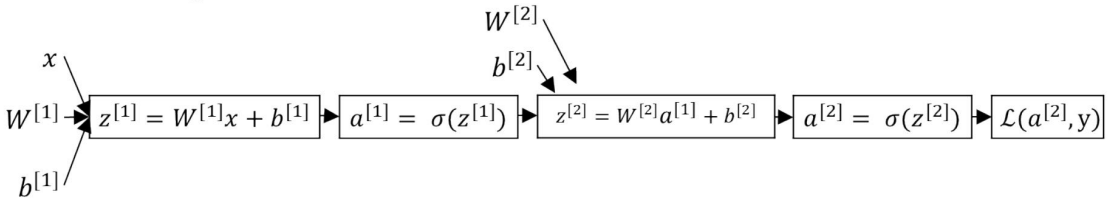

Figure 1.

Here, $L(\hat{y}, y)$ is the log-loss cost function. Derive the mathematical expression of the derivative: $\frac{dL}{dz^{[1]}}$. Show the weight update equation for $W^{[1]}$.

b) Consider a 3-layer neural network with its weight coefficients as specified as in Figure 2. If an input of x=(1,1) is given to this network and the actual output is y=0, then compute the followings:

i. Predicted output, $\hat{y}$ — 3+

ii. Updated weight matrix $W^{[2]}$ — 6+

iii. Updated bias $b^{[2]}$ — 1

[CO1, PO1]

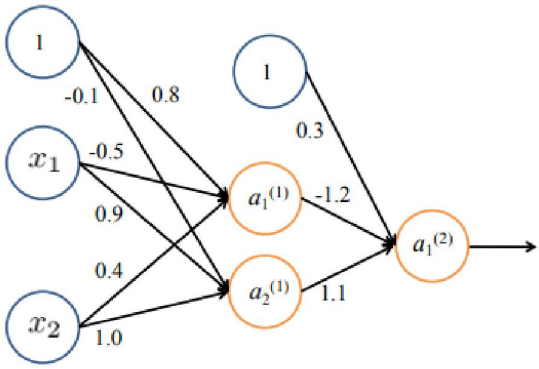

Figure 2.

[Hints: Equations derived in Question 3.(a)]

c) In neural networks, interpret the benefit of the ReLU activation function over a Sigmoid activation function? — 5 [CO2, PO2]