B.Sc. Engg. CSE 6th Semester 20 February 2023 (Morning)

22

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

MID SEMESTER EXAMINATION **SUMMER SEMESTER, 2021-2022**
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. Figures in the right margin indicate full marks of questions whereas corresponding CO and PO are written within parentheses.

---

1. a) Compare between Human Learning and Machine Learning. 5 (CO2) (PO2)

   b) When is a function called convex? Show that log-loss cost function without any regularization term is convex. 1+7 (CO1) (PO1)

   c) When should you not use feature scaling in gradient descent algorithms? Assess how it affects the convergence speed of the optimization process. 4+4 (CO1) (PO1)

   d) In a multivariate linear regression problem, how can you solve the overfitting problem without any regularization penalty? 4 (CO1) (PO1)

2. a) In logistic regression, prove that the weight vector θ is perpendicular to the decision boundary H, where the surface is linear. 4 (CO1) (PO1)

   b) Consider the dataset in the $\mathbb{R}^2$ feature space as shown in Figure 1. 3×5 (CO2) (PO2)

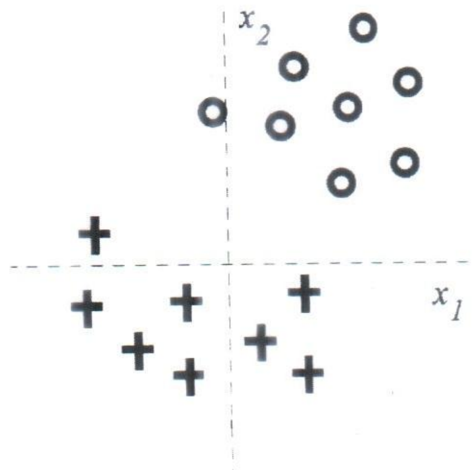

Figure 1: Positive class (y=1) samples are denoted with '+' and Negative class (y=0) samples are denoted with 'o'.

Suppose the regularized cost function of a logistic regression model $h_\theta(x)$ is defined as $J(\theta)$, where we try to minimize the cost for a large value of λ. The regularization term used in the cost function considers only one of the weight coefficients in each case, where $j = \{0,1,2\}$.

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^i \log h_\theta(x^i) + (1-y^i)\log h_\theta(x^i)\right] + \frac{\lambda}{2m}\theta_j^2$$

How does the training error change with the regularization of a single weight coefficient in the following individual cases?

i. Regularizing only θ_0

ii. Regularizing only θ_1

iii. Regularizing only θ_2

Determine whether the training error increases or stays the same (zero) for a large value of λ. Justify your answers.

c) When we change the form of regularization in Question 2.b) to L1-norm (absolute value), and regularize both θ_1 and θ_2 (but not θ_0), we get the following cost function: 6 (CO2) (PO2)

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^i \log h_\theta(x^i) + (1-y^i)\log h_\theta(x^i)\right] + \frac{\lambda}{m}\sum_{j=1}^{2}\left|\theta_j\right|.$$

If we increase the value of λ, which of the following scenarios do you expect to observe?

- First θ_1 will become 0, then θ_2.
- First θ_2 will become 0, then θ_1.
- Both θ_1 and θ_2 will become zero simultaneously.
- None of the weights will become exactly zero, only smaller as λ increases.

Justify your choice.

3. a) Consider a neural network (NN) for a binary classification which has one hidden layer as shown in Figure 2. We use a linear activation function $f(z)=z$ at hidden nodes, and a sigmoid activation function $g(z)$ at the output node. Construct the mathematical expression for the final decision boundary at the output node. 5 (CO1) (PO1)

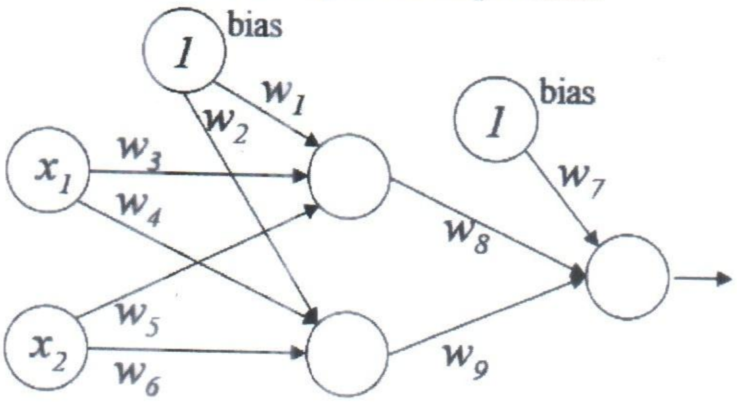

Figure 2: A Multi-layer neural network.

b) How can you simplify the network in Question 3.(a)? Justify why it was possible. 5 (CO2) (PO2)

c) Suppose you have the six training samples for a binary classification problem as shown in Figure 3. Answer the following questions: 10+4 +1 (CO3) (PO3)

i. Design a fully-connected feedforward neural network (FNN) that can classify the samples correctly and contains both decision boundaries (denoted with solid and dashed lines) within it. Intuitively, determine the values of the weights and bias of your network and label your FNN accordingly.

ii. Isolate each of the decision boundary and mark the node where it can be found in your FNN.

iii. Label the region for which the output node predicts as positive class.

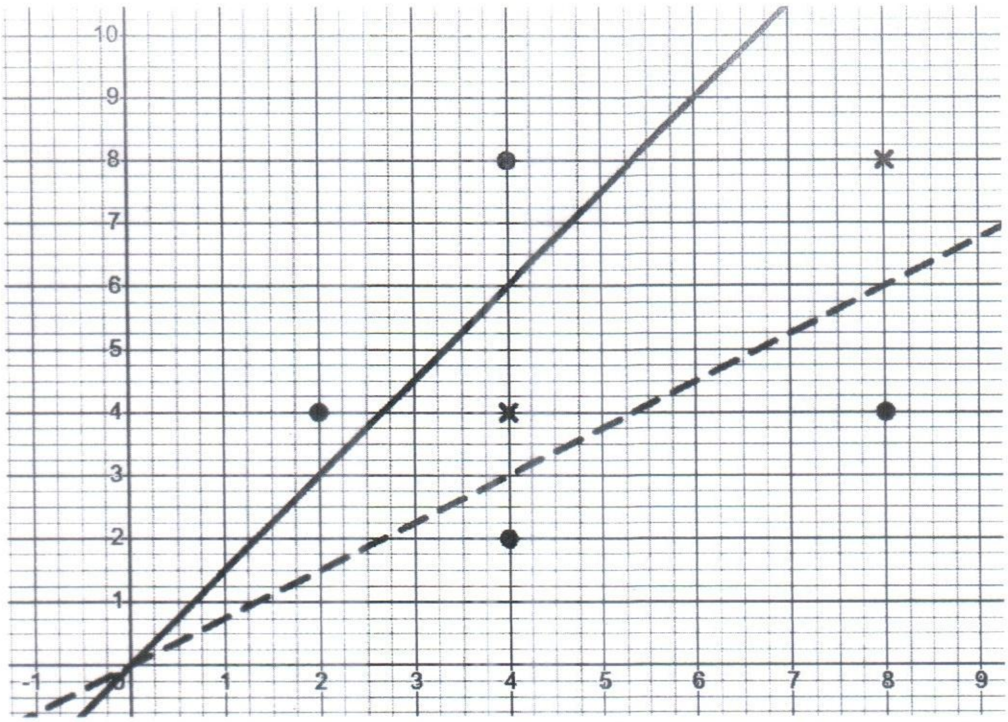

Figure 3: Positive class (y=1) samples are denoted with '×' and Negative class (y=0) samples are denoted with '•'.