ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

ORGANISATION OF ISLAMIC COOPERATION (OIC)

Department of Computer Science and Engineering (CSE)

SEMESTER FINAL EXAMINATION WINTER SEMESTER, 2022-2023

DURATION: 3 HOURS FULL MARKS: 150

CSE 4553: Machine Learning

Programmable calculators are not allowed. Do not write anything on the question paper.

Answer **all 6 (six)** questions. Figures in the right margin indicate full marks of questions whereas corresponding CO and PO are written within parentheses.

---

1. a) Briefly compare the single-unit perceptron with a multi-layer perceptron in terms of their ability to handle complex decision boundaries and non-linearly separable data. 8 (CO2) (PO2)

b) Suppose a neural network for a binary classification has two input units, one hidden layer consisting of two units, and one output unit, as shown in Figure 1. Consider the following:

- The value of the biases are 1
- You have to use linear activation function, $h(z) = cz$ at hidden units
- A sigmoid activation function, $g(z) = \frac{1}{1+e^{-z}}$ at the output unit to learn the function for $P(y = 1|x, w)$ where, $x = [x_1, x_2]$ and $w = [w_1, w_2, \dots, w_9]$

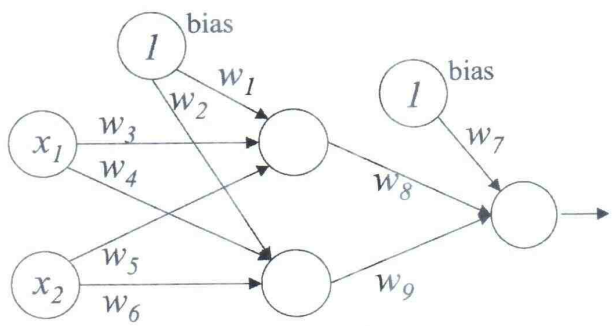

Figure 1: An example feed-forward Neural Network architecture Question 1.b)

Based on the provided information, answer the following:

i. What is the output $P(y = 1|x, w)$ from the neural network? Express in terms of x_i, c, and w_i.. 6 (CO1) (PO1)

ii. What is the final classification boundary? 2 (CO1) (PO1)

iii. Draw a neural network equivalent to the given neural network in Figure 1 with no hidden layer and write weights w of this new neural network in terms of c and w_i. 3 (CO1) (PO1)

c) Explain the performance of an Artificial Neural Network (ANN) for the following cases: 3 × 2 (CO2) (PO2)

i. Working with high-dimensional data

ii. Capturing spatial relationships in data

iii. Weight sharing between neurons in different layers.

2. a) You are tasked with developing a Convolutional Neural Network (CNN) to classify Bengali Sign Digits (0-9). The images in your dataset are of size 28×28 (gray-scale). Consider a CNN architecture is designed as provided in Table 1. Complete the table by calculating the activation volume dimensions and the number of parameters for each layer, assuming a padding of 0 and a stride of 1. 12 (CO1) (PO1)

Table 1: CNN architecture with activation map dimensions and number of parameters Question 2.a)

Layers	Activation Volume Dimensions	Number of parameters
Input	$28 \times 28 \times 1$	0
Conv2D (32 filters, 3×3)		
ReLU		
MaxPooling (2×2)		
Conv2D (64 filters, 3×3)		
ReLU		
MaxPooling (2×2)		
Flatten		
Fully Connected (128 units)		
ReLU		
Fully Connected (10 units)		
Softmax		
Output	$1 \times 1 \times 10$ (probability dist.)	0

b) Consider the ConvNet as shown in Figure 2, where all the variables are scalars. For the input, x_1, x_2, x_3, x_4, x_5; 1D convolution filter or kernel is, $[k_1\ k_2\ k_3]$ and b is the bias in the first layer. The output of the fully connected layers is $\hat{y}$ the weights are w_1, w_2 with bias a. The loss is L.

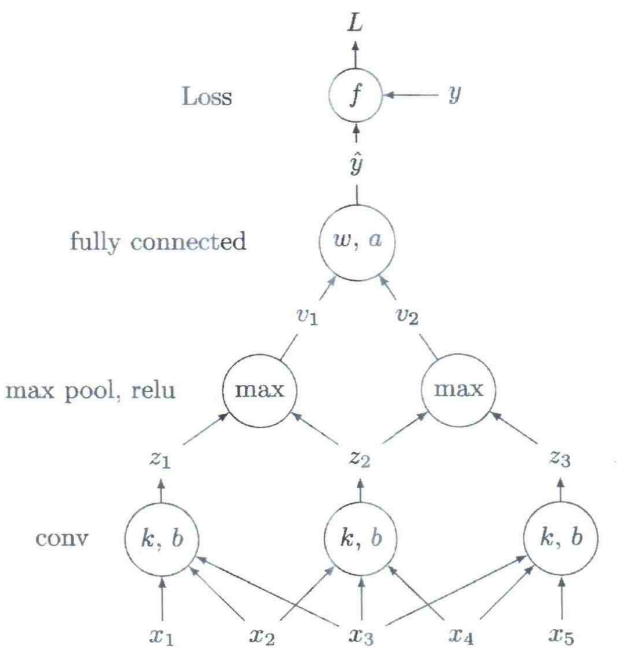

Figure 2: An example CNN architecture Question 2.b)

Answer the following:

i. What are the outputs of the activation map, max pool+relu function, and the loss, L. Write the output in linear algebraic form. 5 (CO1) (PO1)

ii. Given the gradients of the loss L concerning the first layer activation z_i, derive the gradient of the loss for the convolution filter k. Assume that, $\frac{\partial L}{\partial z_1} = \delta_1$, $\frac{\partial L}{\partial z_2} = \delta_2$, $\frac{\partial L}{\partial z_3} = \delta_3$. Determine $\frac{\partial L}{\partial k_1}, \frac{\partial L}{\partial k_2}, \frac{\partial L}{\partial k_3}, \frac{\partial L}{\partial b}$ 8 (CO1) (PO1)

3. In the context of developing a Smart Security Shield for an online payment platform, both Bootstraping Ensemble Classification and K-Means Clustering can play vital roles.

The outputs from both the Bootstraping Ensemble Classification and K-Means Clustering are integrated in real-time to provide an adaptive and dynamic security layer. For instance, if an anomaly is detected by the ensemble classifier, the user's behavior is analyzed within their respective cluster. The system can then apply context-aware security measures, such as additional authentication steps, transaction verification, or temporary transaction restrictions, etc. Based on the scenario, answer the following:

a) Draw an architectural diagram representing the integrated solution. Your diagram should illustrate how the system dynamically detects anomalies in individual transactions, categorizes users into behavior-based clusters, and adapts security measures in real-time. 10 (CO3) (PO3)

b) Explain the designed complete machine learning pipeline consisting of data collection and preprocessing, ensemble classification, k-means clustering, integration and real-time adaptation, and accurate security responses. Your answer should cover technical explanations in brief for each stage of the pipeline. 15 (CO3) (PO3)

4. a) Consider the dataset in Table 2 consisting of information about the movie teaser views (in millions) and likes (in thousands) of five movies along with their respective popularity categories (High, Medium, Low).

Table 2: Dataset for Question 4.a)

Movie ID	Teaser Views (millions)	Likes (thousands)	Popularity
1	25	1.5	High
2	18	0.8	Medium
3	35	2.0	High
4	15	1.2	Medium
5	12	0.5	Low

Answer the followings:

i. For a new movie with teaser views of 22 million and likes of 1.1 thousand, using the k-nearest neighbor (k-NN) algorithm with $k = 3$, find the popularity category of the new movie. Show the step-by-step process. 12 (CO1) (PO1)

ii. Discuss how changing the value of k might impact the prediction and the importance of feature scaling in this k-NN application. 6 (CO1) (PO1)

b) Differentiate between training accuracy and testing accuracy. Why is it important to evaluate a model on a separate test set? Discuss scenarios where a high training accuracy may not necessarily translate to good generalization performance on unseen data. 7 (CO1) (PO1)

5. Suppose you have a labeled dataset of movie reviews, where each review is categorized as positive, negative, or neutral. The text reviews are sequential, and the order of words is crucial for understanding the sentiment. The goal is to classify movie reviews based on the sentiments expressed in the text using Recurrent Neural Network (RNN).

   a) Draw an RNN architecture tailored for sentiment analysis on movie reviews. The RNN model should encompass the input, recurrent, and output layers, emphasizing the sequential nature of textual data. — 7 (CO3) (PO3)

   b) Explain how textual data is presented as input emphasizing the preservation of word order. What problem may arise if you use one-hot encoding for text representation? — 6 (CO1) (PO2)

   c) Briefly explain the process of forward propagation, backward error calculation, and the use of a suitable loss function for the RNN. — 12 (CO1) (PO1)

6. Consider a binary classification problem where you have a dataset with two features, $[x_1, x_2]$, and two classes, labeled as *Positive* and *Negative*. Suppose that you want to solve the classification problem using the Support Vector Machine (SVM) algorithm. Answer the following:

   a) Briefly describe the concept of hyperplane and how SVM aims to find the optimal hyperplane for the classification. — 7 (CO1) (PO1)

   b) Provide a simple example to illustrate a scenario where SVM can find a linear decision boundary and another scenario where it may struggle due to lack of linear separability. How can the kernel trick be used to solve non-linearly separable cases? — 10 (CO2) (PO2)

   c) Explain why the SVM algorithm aims to maximize the margin with the help of support vectors and how it contributes to the model's robustness. — 8 (CO2) (PO2)