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 6231: Deep 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.

---

1. Consider a simple neural network with a single hidden layer of one node. The input is $x \in \mathbb{R}^{1\times 2}$ while the predictions $\hat{y}$, and ground truth y, are scalars. The output of the network can be formulated using the following equation:

$$\hat{y} = w_2 \cdot \max(0, w_1 \cdot x + b_1)$$

Here, w_1 and w_2 are the weights in the hidden and output layers respectively. b_1 is the bias in the hidden layer. The output layer does not have any bias.

The L2 loss, as shown in the following equation, is used to compare the predictions with the ground truth values.

$$L = (\hat{y} - y)^2$$

a) Compute $\frac{\partial L}{\partial w_1}$, $\frac{\partial L}{\partial w_2}$ and $\frac{\partial L}{\partial b_1}$. 6

b) Draw a computation graph depicting the network. Use separate nodes for each of the mathematical operations, including those in the loss function. 5

c) Suppose the weights and bias are initialized as follows: 25

$$w_{1,1,1} = \frac{1}{3}, w_{1,1,2} = \frac{1}{2}, w_{2,1,1} = 2, b_1 = 1$$

Here, $w_{1,1,1}$ and $w_{1,1,2}$ are the weights in the hidden layer while $w_{2,1,1}$ is the weight in the output layer. b_1 is the bias in the hidden layer.

Compute the updated weights and bias of the network if it is trained using the following two training pairs:

- $\{x, y\} = \{[2, 1], 0\}$
- $\{x, y\} = \{[0.5, 1], 1\}$

2. a) Explain the principle of the gradient descent algorithm. Accompany your explanation with a diagram. 7

b) How is data typically divided during the training of a neural network? How can this division help detect overfitting? 7

3. a) Briefly define the following terms: 2×4

i. Supervised Learning
ii. Unsupervised Learning
iii. Overfitting
iv. Underfitting

b) Alice have been tasked with developing a model that can predict whether a patient has high blood pressure or not based on the following features: 7

- Weight
- Height
- Age
- Gender
- Daily active hours

Alice creates a linear regression model which is represented using the following equation:

$$\hat{y} = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + b$$

Here $x_1, \cdots, x_5$ represents the five input features and $w_1, \cdots, w_5$ are their associated weights. The bias is denoted by b.

However, it is unclear from the output of the model which patients have high blood pressure and which do not.

How can this model be modified so that it can clearly predict patients as either having high blood pressure or not having high blood pressure?

What can be done if the new model should now also differentiate between patients that have low blood pressure, patients that have high blood pressure and patients that are healthy?

c) Alice and Bob want to develop a linear regression model that can predict housing prices based on the following features: 10

- Street – road access
- Neighborhood – property location
- Accessibility – transport access
- Year Built – year the house was built in
- Rooms – number of rooms
- Kitchens – number of kitchens
- Fireplaces – number of fireplaces in the house

They have identified rooms, street and neighborhood as the key features of a house that influences its price. Alice believes their regression model should only focus on these key features for accurate predictions while Bob argues that similar importance should be given to the other features as well. They decide to individually build two different models based on their own ideas. The models they build differ only in terms of the regularization term they are using.

What type of regularization did Alice and Bob use? How do their choices help implement their ideas? Accompany your explanation with mathematical representations of their regularization terms.