Birla Institute of Technology & Science, Pilani
Work Integrated Learning Programmes Division
First Semester 2024-2025

Mid-Semester Test
(EC-2 Makeup)

Course No.   : AIMLCZG511
Course Title   : Deep Neural Network
Nature of Exam   : Closed Book
Weightage   : 30%
Duration   : 2 Hours
Date of Exam   : 13-07-2025 (FN)

No. of Pages	= 03
No. of Questions	= 06

Note to Students:
1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

Q.1 Answer the following questions: **[5 Marks]**
  A. True or False: A perceptron is guaranteed to learn any set of training data given a suitable learning rate. Justify your answer precisely.
  B. Consider a supervised learning problem with only two examples in a 5-dimensional space:
         Example 01: (1, 5, 2, 7, 9), label 1
         Example 02: (–3, 8, 2, 4, 6), label 0
  Give two distinct perceptrons that would correctly classify these two points. For each perceptron, write the weights $W = (w_1, w_2, w_3, w4, w_5)$ and the bias b. Briefly explain how each set of weights separates the two points.
  C. An XNOR function cannot be implemented using a single perceptron because it has a nonlinear decision boundary. Given that the number of inputs n=3, compute the number of perceptrons required to implement the XNOR function using both a shallow network and a deep network approach. Additionally, show the MLP diagrams for both approaches, indicating the input layer, hidden layers, and output layer.
Note: *Answer/Justify/Illustrate in no more than 30-50 words and a precise response w.r.t the given question only. Vague answers will be penalized.*

Q2. You are designing a classifier for an image recognition problem with 5 classes: cat, dog, car, bird, and horse. Your model outputs logits for a given image as follows: **[5 Marks]**
Z = [2.5, 0.3, −1.2, 3.1, 0.7]
Answer the following questions:
  A. Compute the SoftMax probabilities $ŷ$ for each class (round off to two decimal places).
  B. If the true class is "bird" (fourth class), compute the cross-entropy loss 'L'.
  C. Suppose you implement a fully connected neural network for this problem. The network has 20 input features, two hidden layers with 15 and 10 neurons respectively, and an output layer with 5 neurons (one for each class). Biases are not included. Compute the total number of trainable weights in this network.

Q3. Given the following neural network with a fully connected layer and ReLU activations, including two input units (i_1, i_2), four hidden units (h_1, h_2), and (h_3, h_4). The output units are indicated as (o_1, o_2), and their targets are indicated as (t_1, t_2). The weights and biases of a fully connected layer are called w and b, represented below: **[5 Marks]**

The values of the variables are given in the table below:

Variable	i_1	i_2	W_{11}	W_{12}	W_{21}	W_{22}	W_{31}	W_{32}	W_{41}	W_{42}	b_1	b_2	b_3	b_4	t_1	t_2
Value	2	-1	1	-0.5	0.5	-1.0	0.5	-1.0	-0.5	1.0	0.5	-0.5	-1.0	0.5	1.0	0.5

Answer the following questions:
- A. Compute the output (o_1, o_2) with the input (i_1, i_2) and network parameters as specified above. Write down all calculations, including intermediate layer results.
- B. Compute the mean squared error of the output (o_1, o_2) calculated above and the target (t_1, t_2).
- C. Update the weight W_{21} using gradient descent with a learning rate of 0.1, as well as the loss computed previously.

  Hint:

  $$\frac{\partial MSE}{\partial w_{21}} = \frac{\partial \frac{1}{2}(t_1 - o_1)^2}{\partial o_1} \times \frac{\partial o_1}{\partial h_3} \times \frac{\partial h_3}{\partial h_1} \times \frac{\partial h_1}{\partial w_{21}} + \frac{\partial \frac{1}{2}(t_2 - o_2)^2}{\partial o_2} \times \frac{\partial o_2}{\partial h_3} \times \frac{\partial h_3}{\partial h_1} \times \frac{\partial h_1}{\partial w_{21}}$$

Note: *Show all intermediate calculations step by step clearly. Partial answers without steps will not receive full marks.*

Q4. Consider the following Python code snippet to train a deep neural network on the MNIST handwritten digit dataset using Keras: **[5 Marks]**

```python
from tensorflow.keras.datasets import mnist
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Flatten
from tensorflow.keras.utils import to_categorical


# Load data
(x_train, y_train), (x_test, y_test) = mnist.load_data()
x_train, x_test = x_train / 255.0, x_test / 255.0
y_train, y_test = to_categorical(y_train, 10), to_categorical(y_test, 10)


# Build model
model = Sequential([
    Flatten(input_shape=(28, 28)),
    Dense(128, activation='relu'),
    Dense(64, activation='relu'),
    Dense(10, activation='softmax')
])


model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
model.fit(x_train, y_train, epochs=5, batch_size=32, validation_split=0.1)
```
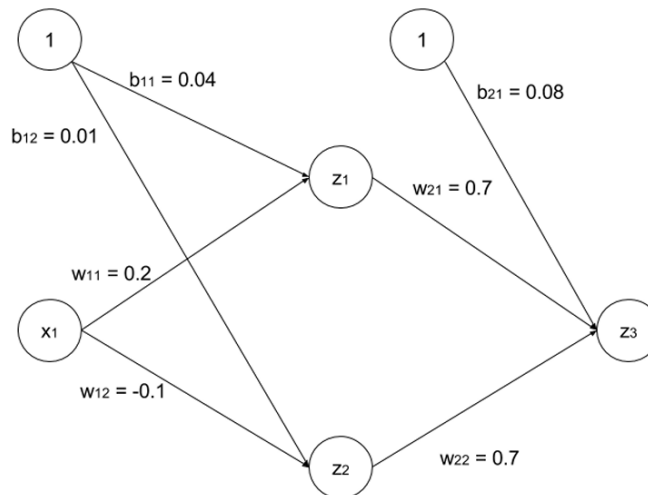
Answer the following questions concisely:
   A. Describe the network architecture: number of layers, depth, number of units per layer, and activation functions used.
   B. Explain why categorical_crossentropy is chosen as the loss function. What is the difference between categorical_crossentropy and sparse_categorical_crossentropy?
   C. Justify the use of softmax in the output layer and ReLU in the hidden layers.
   D. Justify the use of the Adam optimizer. Mention one alternative optimizer and one advantage it provides.
   E. Suggest an alternative network architecture that could also work well on the MNIST dataset and briefly justify your choice.

Note: Answers must be concise within 30–50 words. Justifications are required for full credit.

Q5. Consider the neural network architecture shown below: **[5 Marks]**

Answer the following:

A. Suggest a suitable activation function for the hidden and output layers. Justify your choice for each layer.

B. Name one dynamic learning rate schedule. Explain briefly why using a dynamic learning rate can help during training.

C. Briefly compare GD, SGD, and Mini-batch GD in terms of convergence behavior and stability.

D. Why is it important to initialize weights properly in this network? Name one commonly used weight initialization method and justify its benefit.

E. Briefly explain (in one line) the purpose of adding momentum to SGD.

Note: Answers must be concise within 30–50 words. Justifications are required for full credit.

Q6. Answer the following questions   **[5 Marks]**

    A. You would like to train a dog/cat image classifier using mini-batch gradient descent. You have already split your dataset into train, validation, and test sets. The classes are balanced. You realize that within the training set, the images are ordered in such a way that all the dog images come first and all the cat images come after. A friend tells you: you need to shuffle your training set before the training procedure. Is your friend, right? Explain.

    B. You want to evaluate the classifier you trained in (A). Your test set (X_{test}, Y_{test}) is such that the first m_1 images are of dogs, and the remaining images are of cats. After shuffling X_{test} and Y_{test}, you evaluate your model on it to obtain a classification accuracy a_1%. You also evaluate your model on X_{test} and Y_{test} without shuffling to obtain accuracy a_2%. What is the relationship between a1 and a2 (>, <, =, >=, <=)? Explain.

    C. Data augmentation is often used to increase the amount of data you have. Should you apply data augmentation to the test set? Explain why.

    D. You are training a deep neural network on a small medical image dataset with only 500 labeled samples. During training, you observe that the training accuracy continues to improve, but the validation accuracy starts to drop after a few epochs. Suggest a suitable regularization technique to address this issue and briefly justify your choice.

    E. In one line, explain why deep learning is preferred over traditional machine learning for processing long audio recordings or natural language text.

Note: *Answers must be concise within 30–50 words. Justifications are required for full credit.*

****