Department of Computer Science & Engineering – SVNIT - Surat.

End- Sem Examination March - 2023

B.Tech. - IV – 8th Semester

U19CS012

Course: Deep Learning (CS436)

Date: 2th May 2023 **Time:** 14:00 hrs to 17:00 hrs **Max Marks:** 50

Q1 Answer the following (Any Three) **[09]**

A Consider an application of classifying the probability of a "Gift-Shop" to be Open or Closed. The dataset consists of the following attributes (into brackets its data types have been mentioned) i. Festive time (Binary) ii. Location of Shop (Categorical) iii. Popularity of shop (Numerical range between 1-10). Identify and explain which optimum activation function/s can be used to classify the same at the last layer of deep learning model? [3]

B Refer to the following table-1 having probability ratios of Co-occurrence Matrix and answer the following questions using the same (Assume all calculations already given): [3]

Probability and Ratio	k=hello	k=hi	k=goodbye
P(k | Good-Morning)	2.5×10^{-7}	5×10^{-6}	4.5×10^{-3}
P(k | Ice-Cream)	4.1×10^{-3}	7.8×10^{-6}	6.7×10^{-3}
P(k | Good-Morning)/ P(k | Ice-Cream)	7.5	1.09×10^{-9}	0.87

Table-1: Co-occurrence Probabilities of target words "Good-Morning" and "Ice-Cream" selected from corpus of 10 million token

i. Comment on the relation of the target word "Good-Morning" with each of the given probe words in Matrix?

ii. To what extent is the target word "Ice-Cream" related with the word "hi"?

C Differentiate Operant learning and Classical learning in terms of natural learning. "An accountant has been rewarded for his performance, by which he has been motivated for working harder in his duties" is an example of which kind of learning? Justify your answer. [3]

D What do you mean by the Generalization Gap of a Deep Learning model? How can this problem be solved? Discuss any one technique in detail. [3]

Q2 Answer the following **[11]**

A Consider an application of a 'Temperature prediction system of an area in a city' having the following attributes in dataset i. Longitude ii. Latitude iii. Humidity iv. Previous 10 year temperature record. Problem: "The numerical attributes are not uniformly distributed due to which training of deep neural network models is not proper". What can be the optimum solution for solving the above issue, discuss it in detail. [3]

B Comment on the truthfulness of the following statement: "Weight Initialization is an inevitable step for training a deep-learning model, regardless of distribution of data in the dataset" and justify your answer. [3]

C Consider an application of 'Identifying Juliet Rose' which is to be made by deep learning models. Following is the given scenario: 'The Juliet Rose is supposed to be identified from the images taken by a high quality camera. However, due to security reasons an adequate number of images can't be taken'. Identify the problem in the given scenario and discuss an optimum solution to it. [3]

D What is the significance of IDF in the Word Embedding? Calculate IDF of word "Attitude" for the following Document of a corpus: [2]
"Attitude is key to success. It's a person's positive or negative attitude which takes him/her to the level that he/she wants. Whatever we are is defined by the attitude we have applied in our whole life. The word 'Attitude' is often misunderstood in many situations ". Assume the following information given:
i. Total number of documents in the corpus is 1000.
ii. The number of documents with the word 'key' are 40.

Q3 Answer the following (Any Three) [24]

A Suppose you have an autoencoder model for reconstructing images, with the following architecture: [8]
Input layer: 28x28 grayscale images
Encoder layer: 128 neurons with ReLU activation
Encoder layer: 64 neurons with ReLU activation
Bottleneck layer: 32 neurons with linear activation
Decoder layer: 64 neurons with ReLU activation
Decoder layer: 128 neurons with ReLU activation
Output layer: 28x28 grayscale images
Assuming we are using mean squared error (MSE) loss function and the optimizer with a learning rate of 0.001,
Answer the following questions:
- How many learnable parameters does this autoencoder have?
- If we have a training set of 10,000 images and a batch size of 128, how many batches are needed for each epoch?
- After training for 20 epochs, the model achieves a training MSE of 0.001 and a validation MSE of 0.002. Is the model performing well? Explain your reasoning.

B Suppose you have a CNN model for image classification with the following architecture: [8]
Input layer: 32x32 RGB image
Convolutional layer 1: 32 filters of size 3x3, stride=1, padding=1, ReLU activation
Max pooling layer 1: pool size=2x2, stride=2
Convolutional layer 2: 64 filters of size 3x3, stride=1, padding=1, ReLU activation
Max pooling layer 2: pool size=2x2, stride=2
Fully connected layer 1: 512 neurons, ReLU activation Output layer: 10 neurons with softmax activation
Assuming we are using the cross-entropy loss function and the an optimizer,
Answer the following questions:
- How many learnable parameters does this CNN have?
- If we have a training set of 50,000 images and a batch size of 128, how many batches are needed for each epoch?
- After training for 20 epochs, the model achieves a training accuracy of 94% and a validation accuracy of 90%. Is the model overfitting? Explain your reasoning.

C Suppose we have an LSTM model for predicting the stock prices of a particular company, with the following architecture: [8]

Input layer: 20 time steps of 5 features each
LSTM layer: 64 memory units, return sequences=True
LSTM layer: 32 memory units
Fully connected layer: 1 neuron with linear activation
Assuming we are using mean squared error (MSE) loss function and the RMSprop optimizer, answer the following questions:

- How many learnable parameters does this LSTM have?
- If we have a training set of 1000 time steps and a batch size of 32, how many batches are needed for each epoch?
- After training for 50 epochs, the model achieves a training MSE of 0.001 and a validation MSE of 0.002. Is the model overfitting? Explain your reasoning.

D Suppose we have an RNN model for predicting the next character in a sequence, with the following architecture: [8]

Input layer: one-hot encoding of size 100
Simple RNN layer: 128 memory units, Tanh activation
Fully connected layer: 100 neurons, softmax activation
Assuming we are using categorical cross-entropy loss function and the Stochastic Gradient Descent (SGD) optimizer,
Answer the following questions:

- How many learnable parameters does this RNN have?
- If we have a training set of 10,000 sequences and a batch size of 64, how many batches are needed for each epoch?
- After training for 50 epochs, the model achieves a training cross-entropy loss of 1.5 and a validation cross-entropy loss of 1.7. Is the model overfitting? Explain your reasoning.

Q4 Answer the following (Any Two) **[6]**

A Justify the advantage of AutoEncoder over Principal Component Analysis for dimensionality reduction [3]

B Compare and contrast LSTM and Gated Recurrent Units. [3]

C Write the Steps in Training a Boltzmann Machine and discuss how Gibbs Sampling and Contrastive Divergence works. [3]

Course Outcomes:

CO1: Understand fundamental principles, theory and approaches for learning with deep neural networks. - Q1,Q2,Q3,Q4
CO2: Learn different types of Neural Network and Deep Neural Networks - Q1,Q3,Q4
CO3: Apply NN and DNN for various learning tasks in different domains - Q1,Q2,Q3
CO4: Evaluate various NN and DNN by performing complex statistical analysis for DL techniques.- Q1,Q2,Q3,Q4
CO5: Design DL algorithms for real-world problems - Q1, Q3

*****ALL THE BEST*****