Exam Date & Time: 01-Dec-2023 (02:30 PM - 05:30 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

FIFTH SEMESTER B.TECH END SEMESTER EXAMINATIONS, NOVEMBER-DECEMBER 2023

DEEP LEARNING [DSE 3151]

Marks: 50 **Duration: 180 mins.**

A

Answer all the questions.

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

1) Consider the following training dataset sampled from an EMPLOYEE database. For the given dataset, you are given the task of designing a fully connected neural network to classify test instances based on the class attribute 'Status'. (5)

A)

Instance Number	Department	Status	Age	Salary (Range)
1	Sales	Senior	33	46K-50K
2	Sales	Junior	45	26K-30K
3	Systems	Junior	78	46K-50K
4	Marketing	Senior	96	46K-50K
5	Secretary	Junior	12	26K-30K

i) List the pre-processing steps required to train the Neural Network.

ii) Identify suitable activation function for the hidden layers and output layer.

iii) Demonstrate the working of 1 feed forward considering 2 hidden layers each containing 3 neurons. Assume random initial weights and biases and represent in matrices. Compute error using appropriate loss function after the free forward of Instance Number 3.

B) Explain the working of the Sigmoid activation function. Explain why the usage of the sigmoid activation function in a deep neural network could lead to unstable gradients. (3)

C) What are dropouts in Neural Network? What are the advantages of using dropout in a Neural Network? (2)

2) Assume you are using a Convolutional Neural Network (CNN) for analyzing multi-spectral satellite images. The input is a tensor with 4 channels (each representing a different spectral band) of size 12 x 20 (width x height), which corresponds to a specific (5)

A) area on the Earth's surface. Answer the following questions:

i) Compute the size of the feature map obtained as a result of applying a convolution operation on the input data. To perform the convolution, six filters each of size 3 x 5 is used. The stride length is equal to 1 in the horizontal direction and equal to 2 in the vertical direction, and no padding is applied.

ii) Compute the number of parameters and the total number of connections (including the bias) between the input layer and the convolution layer (whose details are same as described in Q.2A.(i).

iii) Considering the results obtained in part Q.2A.(ii), Articulate the advantages of using a convolution layer instead of a fully connected layer.

B) Explain the following with respect to Deep Convolutional Neural Networks: (3)

i) Inception Block in GoogleNet architecture.

ii) Role of skip connections in ResNet architecture.

Note: Draw block diagrams wherever necessary.

C) You are training a neural network on the ImageNet dataset, and you are thinking of using gradient descent as your optimization function. Analyse if the following statements are true? Justify your answer.

(i) It is possible for Stochastic Gradient Descent to converge faster than Batch Gradient Descent. (2)

(ii) It is possible for Mini Batch Gradient Descent to converge faster than Stochastic Gradient Descent.

(Note: faster here is defined in terms of wall clock time)

3) Consider an RNN with a single hidden layer containing 3 hidden units. The input to the network is a sequence of 5-dimensional vectors, and the output is a sequence of 5-dimensional vectors (eg: i/p at one time-step - [1 2 3 4 5] and o/p at one time step - [0 0 1 0 0]). If there are 3-time steps, answer the following:

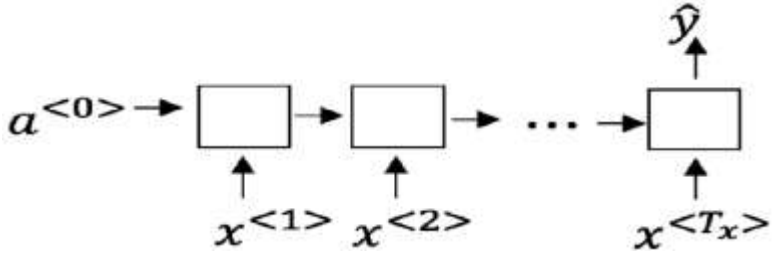
A) i) Compute the total number of parameters in the network, assuming no biases are used.

ii) Assuming random values for Input, Weight and Bias matrices/vectors, show the computations for one feed forward. Use ReLU activation function in the hidden layers, and softmax activation function at the output layer. (5)

iii) Compute the loss for the given network, assuming that you are using cross entropy as the loss function.

B) Summarize the function of the 4 gates in an LSTM cell. Distinguish how the concept of memory is different in a LSTM cell when compared to a GRU cell. (3)

C) Identify any two ML applications that can be modelled using the RNN architecture illustrated in the diagram below.



(2)

4) Consider the task of transliterating sentences from the English script to the Devanagari script, commonly used for languages like Hindi and Marathi. For example, the English word "Namaste" would be transliterated to "नमस्ते" in Devanagari. You are assigned to

A) design an Encoder-Decoder architecture for this task, capable of accepting a word or sentence in English and outputting its phonetic equivalent in Devanagari script. For the given task, explain the following using a neat block diagram and necessary computations: (5)

i) The Design of a suitable Encoder-Decoder model

ii) The Design of an Encoder-Decoder model with attention mechanism.

B) (3)

Distinguish between self-attention and cross-attention in Transformers using a suitable example. Explain in steps how self-attention is performed in Transformer architecture.

C) Compare the working of Neural Machine Translation with traditional machine translation approaches. Explain how round-trip translation helps in the evaluation of Machine Translation. (2)

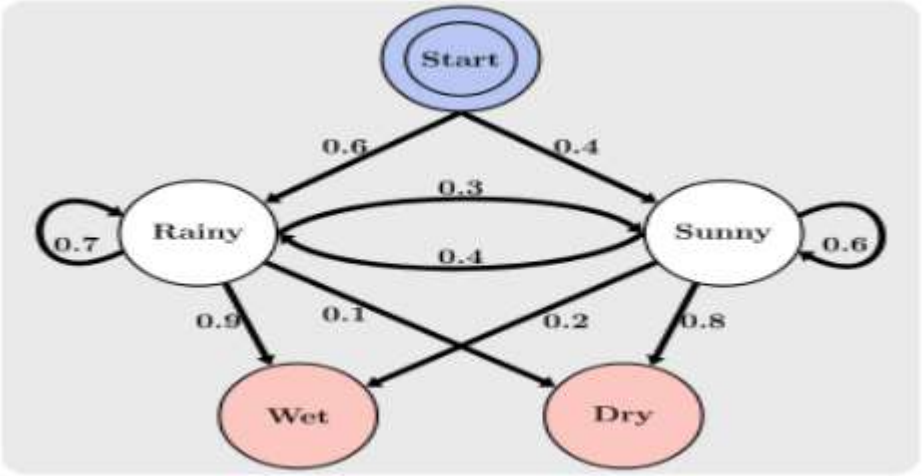
5) You have been tasked with the design of a deep-learning model using autoencoders for anomaly detection in time-series data. Assume you have access to a large dataset of time-series data with both normal and anomalous patterns. Using a neat block diagram and necessary computations, explain the following:

A)

i) Design an Autoencoder Architecture for Time-Series Anomaly Detection. Describe the structure of the autoencoder, including the input layer, hidden layers, bottleneck layer, and output layer. Explain how each component of the autoencoder contributes to learning a representation of normal time-series data. (5)

ii) Illustrate how the trained autoencoder can be used to detect anomalies in new time-series data. Discuss the role of the reconstruction error in identifying anomalies and how you would set a threshold for this error to classify a data point as normal or anomalous.

B) Consider a Hidden Markov Model in the diagram below , the observed variable indicates if the playground is 'Wet' or 'Dry' depending on the weather . (3)



i) Compute the probability of the sequence {Wet,Dry,Dry,Wet, Dry} and represent in a trellis diagram.

ii) Compute the most likely sequence of hidden states for the above-mentioned observed sequence.

C) In an Automatic Speech Recognition system, explain the role of an acoustic model and a lexicon model. (2)

-----End-----