SOMAIYA VIDYAVIHAR UNIVERSITY

Semester: July 2023 –October 2023		
Maximum Marks: 100 **Examination: ESE Examination**		**Duration:3 Hrs.**
Programme code: 54 **Programme:** Honours in Data Science and Analytics	**Class: LY-Honours**	**Semester: VII (SVU 2020)**
Name of the Constituent College: K. J. Somaiya College of Engineering	**Name of the department: Computer Engineering**	
Course Code: 116h54C701	**Name of the Course: Advanced Machine Learning**	
Instructions: 1)Draw neat diagrams 2) All questions are compulsory 3) Assume suitable data wherever necessary		

Que. No.	Question	Max. Marks
Q1	Solve any **Four**	**20**
i)	What are the advantages of CNN (Convolutional Neural network) over ANN (Artificial neural network).	5
ii)	What is the benefit of using Momentum in Gradient Descent optimization?	5
iii)	Why Seq2Seq model is called as Conditional language model? Justify your answer.	5
iv)	Why Sigmoid or Tanh is not preferred to be used as the activation function in the hidden layer of the neural network?	5
v)	You are designing a deep learning system to detect driver fatigue in cars. It is crucial that your model detects fatigue, to prevent any accidents. Which of the following is the most appropriate evaluation metric: Accuracy, Precision, Recall, and Loss Value? Explain your choice.	5
vi)	Describe in short five phases of NLP	5

Que. No.	Question	Max. Marks
Q2 A	Solve the following	**10**
i)	Consider a neural network with three output neurons for a multiclass classification task. The network produces raw scores [2.0, 1.0, 0.5] as its output for a given input. Calculate the probabilities for each class using the softmax activation function. Explain how these probabilities can be interpreted in the context of classification.	5
ii)	Consider the Objective Function: $J(\theta) = (\theta - 4)^2$, $\theta_0 = 3$ (random initial value) $\alpha = 0.1$ (learning rate). Calculate values of θ_1 & θ_2 using gradient descent method.	5
	OR	
Q2 A	Compare Deep Neural Networks (DNNs) with Shallow Neural Networks based on following parameters: i. Architecture Complexity ii. Feature Representation iii. Training challenges iv. Generalization Capability v. Computational efficiency	10

Q 2 B	Solve any **One**	**10**
i)	Name and explain any Four hyper-parameters used for training a deep neural network?	10
ii)	Explain k fold cross validation technique with example.	10

Que. No.	Question	Max. Marks
Q3	Solve any **Two**	**20**
i)	Explain forward and backward propagation in simple RNN with suitable mathematical equations.	10
ii)	Describe the role of the input gate, forget gate, and output gate in an LSTM network. How do these gates contribute to the network's ability to capture long-term dependencies?	10
iii)	Let us consider a Convolutional Neural Network having three different convolutional layers in its architecture as – Layer-1: Filter Size – 3 X 3, Number of Filters – 10, Stride – 1, Padding – 0 Layer-2: Filter Size – 5 X 5, Number of Filters – 20, Stride – 2, Padding – 0 Layer-3: Filter Size – 5 X5 , Number of Filters – 40, Stride – 2, Padding – 0 If we give the input image to the network of dimension 39 X 39 X 3, then determine the dimension of the vector after passing through a fully connected layer in the architecture.	10

Que. No.	Question	Max. Marks
Q4	Solve any **Two**	**20**
i)	Describe the architecture of a typical Seq2Seq model. Include the roles of the encoder and decoder, and explain how the model processes variable-length sequences.	10
ii)	Discuss one real-world application where Seq2Seq models are commonly employed. How does the architecture of Seq2Seq models make them suitable for this application?	10
iii)	Discuss the training process of the CBOW model in Word2Vec word embedding model. Include the roles of the input and output layers, the softmax function, and how word vectors are updated during training.	10

Que. No.	Question	Max. Marks
Q5	Write a short note on any **Four**	**20**
i)	Text pre-processing techniques (Any Two)	5
ii)	Adam Optimizer	5
iii)	Sentiment analysis	5
iv)	One hot encoding	5
v)	Cross-entropy loss	5
vi)	Different types of pooling in CNN.	5

25-11-2024 (M)

SOMAIYA VIDYAVIHAR UNIVERSITY

Semester: July 2024 – November 2024		
Maximum Marks: 100 **Examination: ESE Examination**		**Duration:3 Hrs.**
Programme code: 54 **Programme: Honours Programme in Data Science & Analytics**	**Class: LY-Honours**	**Semester: VII (SVU 2020)**
Name of the Constituent College: **K. J. Somaiya School of Engineering**	**Name of the department: Computer Engineering**	
Course Code: 116h54C701	**Name of the Course: Advanced Machine Learning**	
Instructions: 1)Draw neat diagrams 2) All questions are compulsory **3) Assume suitable data wherever necessary**		

Que. No.	Question	Max. Marks
Q1	Solve any **Four**	**20**
i)	Explain the concept of bias and variance in Machine Learning.	5
ii)	In a churn prediction model, the confusion matrix for a test set of 500 customers is as follows: True Positive (TP) = 120, False Positive (FP) = 50, True Negative (TN) = 200, False Negative (FN) = 130. Calculate the accuracy, precision, recall, and F1-score of the model.	5
iii)	List the applications of Deep learning models across various different domains.	5
iv)	How does cross-validation help in model evaluation?	5
v)	In a CNN, if the input image is 32x32 pixels, the filter size is 5x5, and the stride is 1, what will be the size of the output feature map after applying the convolution operation? Assume no padding is used.	5
vi)	Consider the following objective function: $J(\theta) = (\theta^2 - 6\theta + 9)$, Where the initial value of $\theta_0 = 2$, and the learning rate $\alpha = 0.1$. Using the gradient descent update rule, update the value of θ after the first iteration.	5

Que. No.	Question	Max. Marks
Q2 A	Solve the following	**10**
i)	Which activation function is generally preferred for the output layer in a multiclass classification problem, and why?	5
ii)	What is the primary difference between Stochastic Gradient Descent and Stochastic Gradient Descent with Momentum regarding their update rule and how do they affect parameter updates during training?	5
	OR	
Q2 A	Explain the differences between shallow and deep networks in neural networks. Why do deep networks generally perform better than shallow networks?	10
Q 2 B	Solve any **One**	**10**
i)	What is a Convolutional Neural Network (CNN), and how is it different from a traditional fully connected neural network?	10
ii)	Describe the basic architecture of a CNN. What are the key layers in a typical CNN model, and what is the purpose of each?	10

Que. No.	Question	Max. Marks
Q3	Solve any **Two**	**20**
i)	Explain the role of the hidden state in an RNN. How does it help the network capture temporal dependencies in sequential data?	10
ii)	Describe the architecture and functionality of Long Short-Term Memory (LSTM) networks. How do LSTMs solve the vanishing gradient problem?	10
iii)	Consider a simple RNN with the following parameters: Input vector at time step $t = 1$, $x_1 = [1, 2]$ Weight matrix for the input W_{ih}: $\begin{bmatrix} 0.5 & -0.1 \\ 0.3 & 0.7 \end{bmatrix}$ Weight matrix for the hidden state W_{hh}: $\begin{bmatrix} 0.2 & -0.4 \\ 0.1 & 0.5 \end{bmatrix}$ Intial hidden state $h_0 = [0.1, -0.2]$ and Bias vector $b_h = [0.1, -0.1]$ Calculate the hidden state h_1 at time step $t = 1$.	10

Que. No.	Question	Max. Marks
Q4	Solve any **Two**	**20**
i)	What is word embedding and why is it used in Natural Language Processing (NLP)?	10
ii)	Explain the components of the Seq2Seq architecture, including the encoder, decoder, and how they work together in generating an output sequence.	10
iii)	Explain the concept of text classification in deep learning. How do deep learning models, such as CNNs and RNNs, handle text classification tasks?	10

Que. No.	Question	Max. Marks
Q5	Write a short note on any **Four**	**20**
i)	Vector space model	5
ii)	Different types of RNN	5
iii)	TF – IDF (Term Frequency-Inverse Document Frequency)	5
iv)	Basic Pre-Processing steps in NLP	5
v)	Benefits of Dropout technique in deep learning	5
vi)	Role of Pooling layer in CNN	5