SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)
Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai
Accredited by NAAC-UGC with 'A++' Grade (Cycle III) &
Accredited by NBA (B.E - CSE, EEE, ECE, Mech&B.Tech.IT)
COIMBATORE-641 035, TAMIL NADU

Reg. No: | | | | | | | | | | | | |

B

B.E/B.Tech- Internal Assessment – III
Academic Year 2024-2025 (Even Semester)
Sixth Semester
19CST302-NEURAL NETWORKS AND DEEP LEARNING

Time: $1^{1/2}$ Hours **Maximum Marks: 50**

Answer All Questions

		CO	Blooms	Industry / GATE Ref
1.	Justify why vanishing gradient affects the performance of traditional RNNs and how LSTM overcomes it. Nature of Backpropagation Through Time (BPTT): RNNs use BPTT to update weights. During this process, gradients of the loss function are propagated backward through many time steps.(1) At each time step, the gradient is multiplied by the derivative of the activation function (usually sigmoid or tanh), which has a range between 0 and 1.(1)	CO4	UND	GATE CSE 2024
2.	Compare sparse and denoising auto encoders. **Feature** **Sparse Autoencoder** **Denoising Autoencoder** **Objective** Learn efficient and sparse representations (features) — Learn robust repre[sentations] original input from **Input Data** Original, clean input — Corrupted or noisy input	CO4	UND	-
3.	List few object detection techniques in drone-based surveillance systems. Traditional Computer Vision Methods (Pre-deep learning era): • **Background Subtraction (e.g., MOG2, KNN)**: For detecting moving objects by comparing the current frame with a background model.	CO4	UND	-

- **Haar Cascades**: Lightweight and fast, used for detecting faces, people, or vehicles in relatively controlled environments.
- **HOG + SVM (Histogram of Oriented Gradients with Support Vector Machines)**: Used for pedestrian or vehicle detection.

4.	Demonstrate the role of attention mechanisms in improving the accuracy of image captioning. **Image captioning** is the task of generating natural language descriptions of images. Traditional image captioning models (e.g., CNN + RNN) used a **single, fixed-length vector** to represent the entire image. However, this approach often **misses fine-grained details**, especially in complex images.		CO5	APPLY	-
5.	In the context of Named Entity Recognition, in what ways RNN handle variable-length sequences. In **Named Entity Recognition (NER)**, the input consists of **variable-length sequences** of words (sentences), and one of the key challenges is modeling context over such sequences.		CO5	ANA	GATE 2025

PART – B (2*13=26 Marks)+(1*14=14 Marks)

				CO	Blooms	Industry / GATE Ref
6	a)i)	Analyze the architecture of sparse autoencoders and their use in unsupervised feature learning. 1. **Architecture of Sparse Autoencoders** A **Sparse Autoencoder** is a type of feedforward neural network designed to learn efficient feature representations with a **sparsity constraint** on the hidden units. This is a form of **regularized autoencoder** and is typically used in **unsupervised learning**. Basic Structure: Like any autoencoder, it consists of three parts: • **Input Layer**: x∈ Rnx \in \mathbb{R}^nx∈ Rn • **Hidden Layer** (Encoder): Learns compressed representation hhh h=σ(Wx+b)h = \sigma(Wx +	7	CO4	ANA	-

b)h=σ(Wx+b)

- **Output Layer** (Decoder): Reconstructs the input

x^=σ(W′h+b′)\hat{x} = \sigma(W'h + b')x^=σ(W′h+b′)

The key difference from a basic autoencoder lies in the **sparsity constraint** applied to the hidden layer.

Sparsity Constraint

- **Goal**: Ensure that only a **small number of hidden neurons** are activated (non-zero) for any given input.
- **Why?** Encourages the model to learn **distinct, useful features**, similar to how neurons in the human brain respond selectively.

ii) Analyze a case where object detection fails under low-light conditions and design an attention-enhanced detection model to resolve it. -

Problem Scenario

A drone-based surveillance system fails to detect a **person near a dim streetlamp at night**. The standard object detector (e.g., YOLOv5) either:

- **Misses** the object (false negative), or
- **Detects incorrectly** (false positive: classifies shadows or artifacts as people)

6 CO4 ANA

 Why Detection Fails

Factor	Effect on Model Per
Low Illumination	Reduces object-background co
Visual Noise	Sensor noise mimics texture, c extractors
Blur and Artifacts	Obscure object boundaries
Bias in Training Data	Models trained on well-lit ima poorly
Uniform Focus	CNNs treat all regions equally

irrelevant areas

(OR)

(b)	Analyze how can LSTM networks be used to predict stock price trends when analyzing time-series data in real-world financial market scenarios?				-

Problem Context

Stock prices are inherently **sequential**, influenced by time-dependent patterns like:

- Historical prices
- Trading volume
- News or economic indicators

This makes the task ideal for **time-series modeling** using **LSTM networks** (Long Short-Term Memory), which can **learn long-range dependencies** better than standard RNNs. 13 CO4 ANA

□ Why Use LSTMs?

Feature	Benefit in S
Memory cell + gates	Remembers trends a gradients
Handles variable sequence lengths	Useful for daily, we data
Captures long-term dependencies	Recognizes recurrin cycles
Works with noisy data	Learns smoothed pa backpropagation

7.	(a)	How can an e-commerce platform like Amazon or Flipkart implement opinion mining using RNNs to analyze customer reviews and enhance product recommendations, customer satisfaction, and overall user experience? **Opinion mining** (or sentiment analysis) refers to the process of using natural language processing (NLP) techniques to extract sentiments, opinions, and emotions from text data, such as **customer reviews**. For an e-commerce platform like **Amazon** or **Flipkart**, implementing **opinion mining** using **Recurrent Neural Networks (RNNs)** can significantly enhance **product recommendations**, **customer satisfaction**, and the **overall user experience**. □ **Why Use RNNs for Opinion Mining?** 1. **Sequence-based nature of language**: Customer reviews are sequential in nature (sentences and words depend on each other), making them well-suited for RNNs, which can capture the **context and flow** of language. 2. **Handling variable-length input**: RNNs are able to handle variable-length text (reviews of different lengths), unlike traditional feedforward networks. 3. **Memory retention**: RNNs, especially **LSTM** and **GRU**, can retain long-term dependencies (sentiment shifts across sentences) and **handle context** effectively. 4. **Scalability**: They work well in processing large-scale textual data from millions of reviews across thousands of products.	13	CO5	ANA	GATE 2023

(OR)

(b)	How can autonomous vehicle companies like Tesla or Waymo use GANs for generating synthetic training data to improve the performance and safety of their self-driving systems in diverse driving conditions?	13	CO5	ANA

Using GANs for Generating Synthetic Training Data in Autonomous Vehicles

Autonomous vehicle companies like **Tesla** and **Waymo** can leverage **Generative Adversarial Networks (GANs)** to **generate synthetic data** that significantly improves the performance and safety of their self-driving systems. GANs can create diverse, realistic, and complex driving scenarios that help train self-driving systems to handle various environmental conditions, edge cases, and sensor interactions.

What are GANs?

A **GAN** consists of two components:

- **Generator**: Produces synthetic data (e.g., images, videos, or sensor readings).
- **Discriminator**: Evaluates the authenticity of the generated data, distinguishing it from real-world data.

Through iterative training, GANs generate data that is indistinguishable from real-world data, making them an ideal tool for enhancing self-driving vehicle training.

Key Benefits of Using GANs for Synthetic Data Generation

Benefit	Explanation
Diverse Data Generation	GANs can create synthetic data for rare or hard-to-capture scenari night driving, and unusual traffic
Improved Edge Case Handling	Synthetic data can simulate rare but critical edge cases like pedest or vehicles stopping in unusual lo
Cost-Effective	Reduces the need for expensive real-world data collection in varied conditions (e.g., snow, rair

Augmented Training	GANs can augment training data for various scenarios, improving generalization and performance in real-world conditions.
Sensor Data Simulation	GANs can simulate **sensor data** (LiDAR, radar, cameras) to test and train sensor fusion systems.

8.	a)i	Analyze a case where LSTM models fail in language modeling and suggest enhancements using attention or transformer-based networks.	7	CO4	ANA	Netflix Research Challenge

Overview:

LSTM (Long Short-Term Memory) models are a class of **Recurrent Neural Networks (RNNs)** that are designed to handle sequential data and are widely used in **language modeling**, **text generation**, **machine translation**, and other natural language processing (NLP) tasks. LSTMs are capable of capturing long-term dependencies by addressing the **vanishing gradient problem** that traditional RNNs face.

However, LSTMs still encounter several limitations in language modeling, particularly when dealing with long-range dependencies, computational inefficiency, and difficulties in parallelization. **Attention mechanisms** and **transformer-based networks**, such as the **Transformer** and **BERT**, have emerged as more effective solutions to these challenges.

Challenges of LSTM in Language Modeling

1. **Limited Long-Range Dependency Modeling**:
   - **Problem**: LSTMs are designed to capture long-term dependencies, but when the sequence length becomes very large (e.g., several paragraphs or documents), LSTMs may still struggle to effectively capture relationships between distant words or phrases.
   - **Example**: In a long document, LSTM may have difficulty connecting the beginning of the text with the end or understanding long-range contextual dependencies, such as resolving anaphora (e.g.,

understanding what "he" refers to in a long sentence).

2. **Slow Training and Inference**:
   - **Problem**: LSTMs, as sequential models, must process one word at a time. This limits the ability to parallelize computations, leading to slow training times, especially with large datasets.
   - **Example**: For training on large corpora (e.g., billions of words), LSTMs require substantial computational resources, and training time increases significantly compared to parallelizable models.
3. **Difficulty with Handling Complex Dependencies**:
   - **Problem**: While LSTMs are effective at capturing some level of temporal dependencies, they may still struggle with more complex patterns, such as capturing interactions between non-adjacent tokens across long sequences.
   - **Example**: Complex syntactic structures or subtleties in meaning (e.g., sarcasm, idioms) may be lost as LSTMs focus more on immediate contexts rather than distant ones.

ii)	Imagine a scenario where a video streaming platform wants to automatically understand viewers' emotional reactions to content. In this context, how could a hybrid architecture combining LSTM and attention mechanisms be implemented to perform effective video sentiment analysis?	7	CO5	ANA	GATE 2025

In a **video streaming platform**, analyzing viewers' emotional reactions to content can provide valuable insights for content recommendations, user engagement, and content optimization. **Sentiment analysis** of videos involves understanding the emotional tone conveyed by the content and the viewer's response to it, which is a combination of **visual**, **audio**, and **contextual information** over time.

To perform effective **video sentiment analysis**, we can design a hybrid architecture that integrates **LSTM** (for capturing sequential dependencies) and **attention mechanisms** (for focusing on important information). This hybrid model can process multimodal data from videos—such as **visual**, **audio**, and **textual** features—while considering long-range dependencies and focusing on the most relevant parts of the content.

Key Steps for Implementing the Hybrid Architecture

---

1. Preprocessing and Feature Extraction
Multimodal Data:

For video sentiment analysis, the input consists of three primary modalities:

- **Visual data**: Frames or sequences of frames extracted from the video.
- **Audio data**: Speech or sound features from the video's soundtrack.
- **Textual data**: Subtitles or transcripts from the video.

Each modality is processed separately to extract meaningful features:

- **Visual Features**: Use **Convolutional Neural Networks (CNNs)** or pretrained models like **ResNet** to extract image features from video frames.
- **Audio Features**: Extract features like **MFCCs** (Mel-frequency cepstral coefficients) or embeddings from speech recognition systems.

(OR)

b)i)	Build a real-time deep learning pipeline using denoising autoencoders for industrial defect detection in manufacturing	7	CO4	APP	Industry Case: Bosch

In an industrial manufacturing environment, detecting defects early in the production process is crucial for maintaining product quality and minimizing waste. **Denoising Autoencoders**

(DAEs) are a powerful tool for unsupervised learning tasks like **anomaly detection** and **defect detection**. By training the autoencoder to reconstruct clean, defect-free images from noisy versions, it can effectively learn the normal patterns in the manufacturing process. When it encounters a new image with a defect, the reconstruction error will be higher, which can be flagged as a defect.

Overview of the Pipeline

1. **Data Collection**: Collect images of the products during the manufacturing process (both with and without defects).
2. **Data Preprocessing**: Clean the images, apply noise (for denoising autoencoders), and augment the dataset.
3. **Model Training**: Train a denoising autoencoder to learn the normal patterns in defect-free products.
4. **Anomaly Detection**: Use the reconstruction error to identify defects during real-time processing.
5. **Real-Time Detection**: Implement a real-time pipeline to capture images, process them, and detect defects during manufacturing.

ii)	Consider a scenario where an e-commerce platform is looking to improve its product recommendation system by analyzing customer purchase behavior. How can sparse or contractive autoencoders be applied to overcome the limitations of standard autoencoders and enhance the feature representation for more accurate and personalized recommendations?	7	CO5	APP	Industry Case: Bosch

In an e-commerce platform, improving the **product recommendation system** is essential to providing a personalized shopping experience for customers. **Sparse** and **Contractive Autoencoders** (CAEs) can be highly effective in overcoming the limitations of standard autoencoders by enhancing the feature representation, leading to better recommendations.

Problem with Standard Autoencoders for E-commerce Recommendation

Standard autoencoders aim to learn a compact representation of input data (e.g., purchase behavior, product interactions, etc.) by training an encoder-decoder architecture. However, they might face the following limitations:

1. **Overfitting**: A regular autoencoder might capture noise in the data and reconstruct overly general features, leading to overfitting on the training set.
2. **Sparse Representation**: The latent space representation might not be sparse enough to highlight the essential features (e.g., specific customer preferences or interests), which can hinder accurate recommendations.
3. **Lack of Robustness**: Standard autoencoders might not be robust to small variations in the input, making the learned representations sensitive to slight changes in customer behavior.

Bloom's Taxonomy:

REM – Remember UND – Understand APP– Apply ANA– AnalyzeEVA - Evaluate

CRT - Create

Faculty in-charge Teaching Coordinator HoD Dean