Unit - V Deep Learning for Sequential data and Image data

5.1 Sequential Data: Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs)

Sequential Data

Definition:

Sequential data is data where the **order of elements matters**, and each element depends on previous ones.

Examples:

Time series, speech signals, text sentences, stock prices.

Recurrent Neural Networks (RNNs)

Definition:

A **Recurrent Neural Network (RNN)** is a neural network designed to process sequential data by maintaining a **hidden state** that captures information from previous time steps.

Key Features:

- Feedback connections
- Parameter sharing across time
- Processes input one step at a time

Limitation:

- Suffers from **vanishing and exploding gradient problems**
- Poor at learning long-term dependencies

Long Short-Term Memory (LSTM) Networks

Definition:

LSTM is a special type of RNN designed to **learn long-term dependencies** using memory cells and gating mechanisms.

Key Components:

- **Forget gate –** decides what information to discard
- **Input gate –** decides what information to store
- **Output gate –** controls output

Advantages:

- Solves vanishing gradient problem
- Captures long-range dependencies

Applications:

- Speech recognition
- Machine translation
- Time-series forecasting

Gated Recurrent Units (GRUs)

Definition:

GRU is a simplified version of LSTM that uses fewer gates while still capturing long-term dependencies.

Key Components:

- **Update gate**
- **Reset gate**

Advantages:

- Faster training
- Fewer parameters than LSTM
- Comparable performance

5.2 Image Data : Pre-trained Neural Networks, Transfer Learning, Fine Tuning

Pre-trained Neural Networks

Definition:

Pre-trained neural networks are models that have already been trained on **large image datasets** (such as ImageNet) and learned general visual features like edges, textures, and shapes.

Examples:

VGG, ResNet, Inception, MobileNet

Advantages:

- Saves training time
- Requires less data
- High accuracy

Transfer Learning

Definition:

Transfer learning is a technique where **knowledge learned from one task** is reused to solve a **different but related task**.

How it Works:

- Use a pre-trained model
- Replace the final classification layer
- Train the model on a new dataset

Benefits:

- Faster convergence
- Better performance on small datasets
- Reduced computational cost

Fine-Tuning

Definition:

Fine-tuning is the process of **retraining some or all layers** of a pre-trained network to adapt it more closely to a new task.

Approach:

- Freeze early layers
- Retrain deeper layers with a low learning rate

Advantages:

- Improves accuracy

Adapts model to specific features of new data

5.3 Introduction to Transformers, Generative Pre-training Transformer(GPT)

Introduction to Transformers

Definition:

A **Transformer** is a deep learning architecture designed to process **sequential data** using **self-attention** instead of recurrence or convolution.

Key Idea:

Transformers capture relationships between all elements in a sequence **simultaneously**, enabling better modeling of long-range dependencies.

Key Components of a Transformer

1. **Self-Attention Mechanism**
    - Computes relationships between all words/tokens in a sequence.
    - Assigns importance (attention weights) to each token.
2. **Multi-Head Attention**
    - Uses multiple attention heads to learn different relationships in parallel.
3. **Positional Encoding**
    - Adds position information since transformers have no recurrence.
4. **Feed-Forward Neural Network**
    - Applies non-linear transformations to attention outputs.
5. **Layer Normalization and Residual Connections**
    - Improve training stability and speed.

Advantages of Transformers

- Parallel processing (faster than RNNs)
- Handles long-term dependencies well
- Scalable to very large datasets

Generative Pre-trained Transformer (GPT)

Definition:

GPT is a **Transformer-based language model** that is **pre-trained** on large text data and **fine-tuned** for specific tasks.

Key Characteristics of GPT

- Uses **decoder-only Transformer architecture**
- Trained using **self-supervised learning**
- Predicts the **next word/token** in a sequence
- Unidirectional (left-to-right)

Training Phases of GPT

1. **Pre-training**
    - Learns general language patterns from large corpora
2. **Fine-tuning**
    - Adapts model to specific tasks (e.g., Q&A, summarization)

Applications of GPT

- Text generation
- Chatbots and virtual assistants
- Machine translation
- Text summarization
- Code generation