

Name : Edric Jeffrey Sam
Subject : Deep Learning
Test : CIA

Roll no : 23BDS014

Section A

1. C) Neuron
2. B) To introduce non-linearity into the network
3. C) The error between the predicted output and the actual target
4. C) Learning rate
5. C) Presence of multiple hidden layers
6. C) Layer
7. C) Simpler features are learned in early layers and combined...
8. C) Computer Vision
9. C)
10. C) Voice Recognition

Section B

11. A. Deep Learning is a subset of machine learning based on artificial neural networks with representation learning. It involves algorithms that attempt to model high-level abstractions in data by using multiple processing layers.
Key Characteristic: The defining characteristic is depth—the use of multiple hidden layers (usually more than one or two) between the input and output layers to automatically learn hierarchical feature representations.
12. B. A hyperparameter is a configuration variable that is external to the model and whose value cannot be estimated from the data. It must be set before the learning process begins.

Examples:

Learning Rate: Controls how much to change the model in response to the estimated error each time the model weights are updated.

Number of Epochs: The number of times the learning algorithm will work through the entire training dataset.

Section C

- 13 A. Feature hierarchy refers to the composition of features from lower to higher levels of abstraction.

Process:

In a deep network, the initial layers learn low-level features (like edges or lines in an image). As the data propagates through deeper layers, across simple to deeper layers, these simple features are combined to form complex, high-level features (like eyes, noses, or entire faces), allowing the model to understand complex data structures automatically.

- 14 B. Deep Learning models, particularly Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and Transformers, are used to convert spoken language (audio signals) into text. They process temporal sequences of data to identify temporal sequences of audio data to identify phonemes and words, handling variations in accents, noise, and intonation more effectively than traditional statistical models.

- 15 A. Typically consists of an input layer, only one (or very few) hidden layers, and an output layer. It is limited in its ability to learn complex, non-linear patterns.

Deep Neural Network (DNN): Consists of an input layer, multiple (many) hidden layers, and an output layer. The increased depth allows it to model complex relationships and learn hierarchical representations of data.

Section C

- 16 A. Deep Learning is a specialized subset of Machine Learning that uses multi-layered artificial neural networks to solve complex problems. It mimics the human brain's ability to learn by processing data through layers of neurons.

Differences from Traditional Machine Learning:

Feature Extraction: Traditional ML often requires manual feature extraction (hand-engineering features). Deep Learning performs automatic feature extraction, learning relevant features directly from raw data.

Data Dependency: Traditional ML can perform well with smaller datasets. Deep Learning typically requires large volumes of data to perform well and avoid overfitting.

Computational Power: Deep Learning is computationally intensive and relies on GPUs/TPUs, whereas traditional ML can often run on standard CPUs.

Performance: As data volume increases, Deep Learning performance continues to improve, whereas traditional ML algorithms often plateau.

Key Characteristics:

Hierarchical feature learning.

End-to-end learning capabilities.

High capacity for modelling non-linear relationships.

- 17 B. Importance of Loss Functions: The loss function acts as a guide for the training process. It mathematically calculates the difference (error) between the model's prediction and the actual ground truth. By minimizing this loss value (using optimization algorithms like Gradient Descent), the model adjusts its weights to improve accuracy. Without a loss function, the model has no feedback mechanism to learn.

Three Crucial Hyperparameters:

Learning Rate (α):

Explanation: It determines the step size at each iteration while moving toward a minimum of a loss function.

Importance: If too high, the model might overshoot the optimal point; if too low, training will be too slow or get stuck.

Batch Size:

Explanation: The number of training examples utilized in one iteration to update the internal model parameters.

Importance: Smaller batches offer a regularizing effect and lower memory usage but are noisier; larger batches provide accurate gradient estimates but require more memory.

Number of Hidden Layers & Units:

Explanation: This defines the "capacity" or complexity of the model architecture.

Importance: Too few layers/units might lead to underfitting (model is too simple to learn the data); too many can lead to overfitting (model memorizes the noise in the data).