Deep Learning vs XGBoost
Final Comparative Report

Prepared by: Asritha

Course: Deep Learning - Week 12 Assignment

Date: May 04, 2025

Table of Contents

Introduction

The objective of this report is to compare Deep Learning models with XGBoost models on synthetic datasets of different sizes. The evaluation focuses on training error, validation error, and execution time to identify the most efficient and accurate model, especially for large datasets.

Deep Learning Results

Dataset	Configuration	Training Error	Validation Error	Time (s)
1000 rows	1 hidden layer (4 nodes)	0.1423	0.1387	4.61
1000 rows	2 hidden layers (4 nodes each)	0.0681	0.0724	5.14
10000 rows	1 hidden layer (4 nodes)	0.0432	0.0402	16.11
10000 rows	2 hidden layers (4 nodes each)	0.0137	0.0135	15.08
100000 rows	1 hidden layer (4 nodes)	0.0026	0.0026	146.00
100000 rows	2 hidden layers (4 nodes each)	0.0020	0.0020	139.29

Observation: Deep Learning models reduce validation errors significantly as the dataset size increases, showing excellent learning capacity. However, training time rises sharply, making them less practical for very large datasets.

XGBoost Results

Method	Dataset Size	Accuracy	Time (s)
Python XGBoost	100	0.8500	0.88
Python XGBoost	1,000	0.9380	1.88
Python XGBoost	10,000	0.9737	7.99
Python XGBoost	100,000	0.9885	4.02
Python XGBoost	1,000,000	0.9918	10.46

Python XGBoost	10,000,000	0.9932	94.28

Observation: XGBoost consistently achieves high accuracy across all dataset sizes. Additionally, it maintains faster execution times compared to deep learning, especially on large datasets.

Detailed Analysis

When dataset size increases, Deep Learning reduces validation errors better but takes significantly more time. XGBoost on the other hand balances speed and accuracy efficiently. At dataset sizes of 100,000 and above, XGBoost not only completes faster but also achieves nearly perfect accuracy.

Model Comparison

Deep Learning:

- Best suited for highly complex data patterns
- Reduces errors effectively with large datasets
- Requires very high training time on large datasets

XGBoost:

- Consistently high accuracy across all dataset sizes
- Faster training and better optimized for large datasets
- Preferred choice for practical and large-scale data problems

Conclusion and Recommendation

XGBoost is superior overall for large datasets due to its combination of high accuracy and fast computation. Deep Learning is still valuable for complex pattern recognition but is not efficient for large datasets due to long training times. Thus, XGBoost is recommended for most practical applications requiring quick and accurate results.