Week 12 Assignment: Deep Learning with Synthetic Data

1. Based on the results, which model do you consider as superior among the deep learning models fit?

Among the deep learning models tested, the 2-hidden-layer model with 4 nodes per layer consistently performed better on smaller datasets (1000 and 10,000 samples) in terms of validation error. However, on the largest dataset (100,000 samples), the 1-hidden-layer model slightly outperformed the 2-layer model, showing a lower validation error (0.485 vs. 0.573) and faster execution time (42.65s vs. 50.80s). Therefore, while the 2-layer model seems more effective at learning from smaller datasets, the 1-layer model appears to scale better with large data in terms of efficiency and accuracy. Overall, the 1-hidden-layer model is slightly superior for high-volume training environments due to its better generalization and faster performance on large-scale data.

2. XGBoost Results Table (from Week 11):

Data Size	Accuracy	Execution Time (s)
1000	0.9500	0.86
10000	0.9777	0.76
100000	0.9872	3.87

3. Comparing Deep Learning vs XGBoost - Which is Superior?

XGBoost clearly outperforms the deep learning models in both accuracy and execution time across all dataset sizes. For example, at 100,000 records, XGBoost achieved 98.72% accuracy in only 3.87 seconds, whereas the best deep learning model had significantly higher validation error (0.485–0.573) and took over 40 seconds to train.

The basis for this judgment is that XGBoost:

- Achieves higher prediction accuracy

- Requires significantly less training time

- Is better optimized for tabular datasets like the one used in this assignment

Therefore, XGBoost is the superior model overall in this scenario due to its efficiency, scalability, and performance.