Week 12 - Neural Network

1. Deep-Learning Results

Data size	Hidden layers	Training accuracy	Validation accuracy	Execution time (s)
1 000	1	0.7812	0.7450	2.02
1 000	2	0.7762	0.7450	2.30
10 000	1	0.9965	0.9955	10.37
10 000	2	0.9969	0.9980	11.46
100 000	1	0.9988	0.9986	101.72
100 000	2	0.9985	0.9976	269.82

---

2. Best Deep-Learning Configuration

- **Validation accuracy** is the primary criterion.
- At **1 000** observations both 1- and 2-layer nets achieve 0.7450, but the 1-layer model is slightly faster (2.02 s vs 2.30 s).
- At **10 000**, the 2-layer net edges out in validation (0.9980 vs 0.9955) for a modest time penalty (11.46 s vs 10.37 s).
- At **100 000**, the 1-layer net achieves higher validation accuracy (0.9986 vs 0.9976) in far less time (101.7 s vs 269.8 s).

Conclusion: The **1-hidden-layer, 4-node** network offers the best trade-off

consistently top validation accuracy at large scale and lowest execution time.

---

3. Relevant XGBoost Results (Python, scikit-learn CV)

Dataset size	Accuracy	Time (s)
1 000	0.9480	2.28
10 000	0.9772	3.07

Dataset size	Accuracy	Time (s)
100 000	0.9869	13 119.65

---

4. DL vs. XGBoost Comparison

Data size	Best DL (1-layer) val. acc.	DL time (s)	XGB acc.	XGB time (s)
1 000	0.7450	2.02	0.9480	2.28
10 000	0.9955	10.37	0.9772	3.07
100 000	0.9986	101.72	0.9869	13 119.65

- **At 1 000 rows**: XGBoost substantially outperforms DL in accuracy (0.948 vs 0.745).
- **At 10 000 rows**: DL overtakes XGBoost (0.9955 vs 0.9772), though XGBoost is faster.
- **At 100 000 rows**: DL both out-accuracies (0.9986 vs 0.9869) and completes far more quickly (102 s vs >13 000 s).

Overall Recommendation

- For **small datasets** (~1 000 rows), **XGBoost** is clearly superior (much higher accuracy).
- For **moderate to large datasets** (≥10 000 rows), the **1-layer deep network** is superior—matching or beating XGBoost in accuracy while scaling far better in runtime at high volumes.

Basis: We prioritize validation accuracy first, then execution time. Thus:

- **XGBoost** for quick, small-data problems.
- **1-layer DL** for anything ≥10 000 observations, due to its combination of near-perfect accuracy and practical training times.