Tabular Data: Deep Learning is Not All You Need

Ravid Shwartz-Ziv and Amitai Armon

IT AI Group, Intel

Tabular Data

- Images or text are not the most frequent data
- Tabular data – a mix of numeric, symbolic, and textual data
- Challenges of tabular data:
  - Mixed features
  - Sparse data
  - Less data
  - Different datasets types
  - Prior knowledge (feature engineering)

Compare Different Models

- Deep models
  - TabNet[1]
  - NODE[2]
  - DNF-Net[3]
  - 1D-CNN
- XGBoost[4]
- Ensemble of models

Datasets and optimization

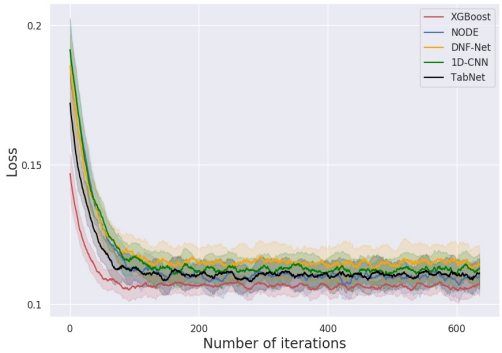
- No common benchmark
- Three datasets from each paper
- Two unrelated datasets
- Bayesian Hyper-parameter search
- 1000 – 10000 runs

Results

Name	Average Relative Performance (%)
XGBoost	3.34
NODE	14.21
DNF-Net	11.96
TabNet	10.51
1D-CNN	7.56
Simple Ensemble	3.15
Deep Ensemble w/o XGBoost	6.91
Deep Ensemble w XGBoost	**2.32**

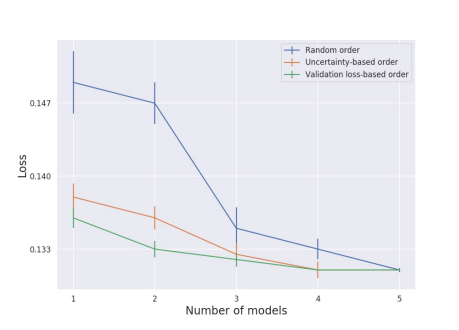
XGBoost had a much better performance than the deep models, their ensemble performed slightly better (lower is better)

Hyper – Parameter Optimization



It is easier to optimize XGBoost

Selecting Subset of Models in the Ensemble



Getting good results with only 3 models

Summary

- On datasets that did not appear in their original papers, deep models were weaker
- XGBoost had better accuracy than the deep models
- Ensemble of deep models with XGBoost performed better
- XGBoost converged more quickly to good performance
- In an ensemble, the order of selecting models was important

References

1. Arik, S. Ö., & Pfister, T. (2021). TabNet: Attentive Interpretable Tabular Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(8), 6679-6687
2. Popov, Sergei, Stanislav Morozov, and Artem Babenko. "Neural oblivious decision ensembles for deep learning on tabular data." *ICLR, 2019*
3. Abutbul, Ami, et al. "DNF-Net: A Neural Architecture for Tabular Data.", ICLR, 2021
4. Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalabletree boosting system." *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.