NEURAL OBLIVIOUS DECISION ENSEMBLE (NODE) FOR FLIGHT DELAY PREDICTION

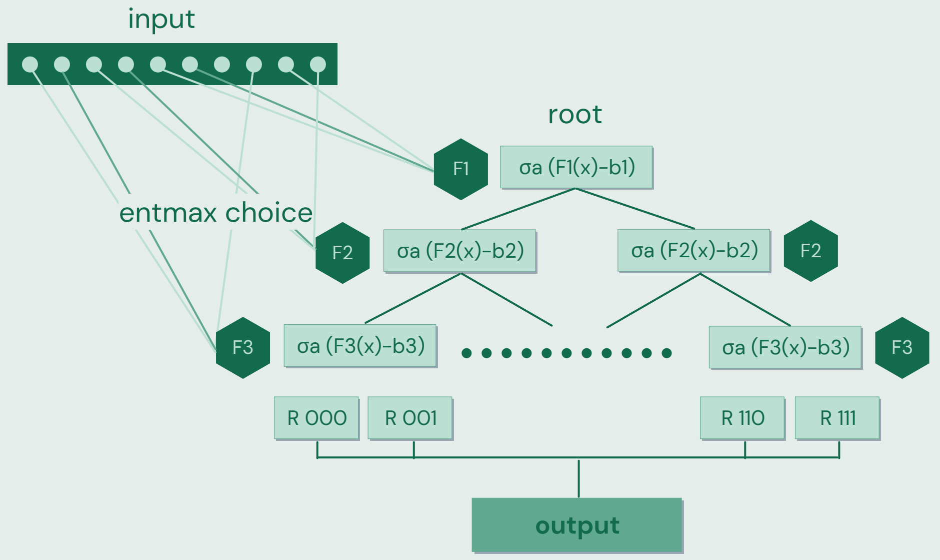
INNOVATION

COMBINATION OF DNN AND DECISION TREES

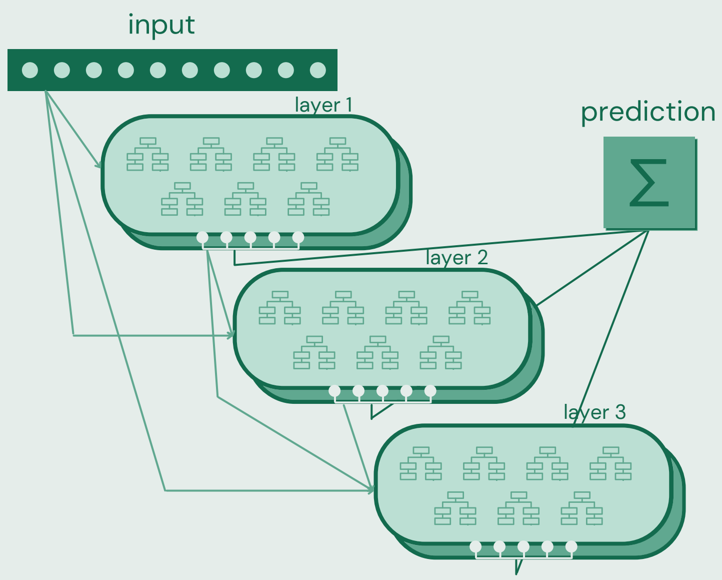
PROBLEM

- For tabular data, **Gradient Boosted Decision Trees** (GBDTs) are state-of-the-art and perform better than **DNNs**.
- **Fusing** GBDTs and DNNs could potentially drastically decrease errors.
- Traditional decision trees are **not differentiable** since they use binary splitting features.
- Differentiable models are necessary for gradient descent optimization and **backpropagation**.
- How do we solve this problem?

OBLIVIOUS DIFFERENTIABLE DECISION TREES



MULTILAYER ARCHITECTURE



NODE'S APPROACH[1]

- The recent ***entmax*** transformation allows for a **"soft"** splitting feature choice.
- The splitting feature choice *(Fi)* and threshold *(bi)* are continuous instead of binary.
- This makes the decision trees differentiable and allows for **end-to-end training** via back-propagation like in "normal" DNNs.
- **Multi-layer architecture**: input for every NODE layer is a concatenation of the original input and previous layers.

EXPERIMENTAL SETUP

DATA PROCESSING AND EXPERIMENT

PRE-PROCESSING

Columns dropped:

- *id, date, wheels off*

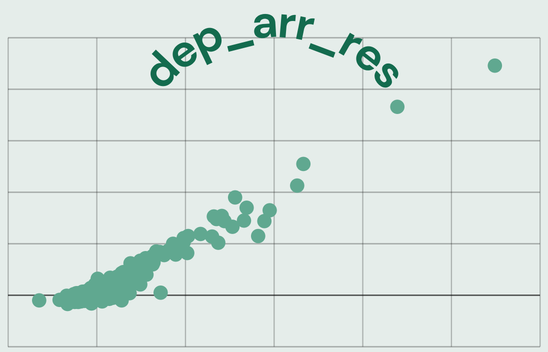
Columns added:

- target encoded variables: *airline, month, day of week and airport*
- 'dep_arr_res': residuals linear regression **departure delay ~ arrival delay** used as target variable which approx. follows a Gaussian distribution
- '*Arrival delay bin*': arrival delay binned into 3 equal bins to categorize delays

LIBRARIES `PyTorch` `xgboost` `qhoptim.pyt` `scikit-learn`

EXPERIMENTS

- Learn input **embedding space**
- **Train NODE** model predicting 'residuals' then transforming back to arrival delay
- **Hyper parameter tuning** with different set of parameters (*num_layers, layer_dim, num_trees, tree_depth*)
- **Comparison** to state-of-the-art XGBoost



- ***Additional things we tried***
- *auto-encoder*
- *one-hot encoding*
- *ensembling the model in random forest-like fashion*

BENCH MARKS

Linear Regression	*166.0*
XGBoost	*150.72*

ANALYSIS

CONCLUSION

The NODE model outperforms both benchmarks. However, our Kaggle MSE score is not competitive, even with variations in depth, layers and trees. Future research should further explore the multi-layer differentiable layer architecture and perhaps look into implementing non-oblivious decision trees.

MSE

88.88

12th place kaggle competition

[1]. *Sergei Popov, Stanislav Morozov, & Artem Babenko. (2019). Neural Oblivious Decision Ensembles for Deep Learning on Tabular Data.*