NODE-GAM

Advanced Machine Learning in Big Data Analytics

Gregor Felix Foitzik

Content

1. Why are interpretable models important?
2. GAM *(Generalized Additive Models)*
3. NAM *(Neural Additive Models)*
4. NODE *(Neural Oblivious Decision Ensembles)*
5. NODE-GAM
6. Conclusion

Why are interpretable models important?

Transparency:

Increase trust in modern healthcare

Regulation require transparency in decision-making

Decision Support:

Diagnosis (image diagnosis via MRT or CT)

Bias Detection:

Uncover biases in data

GAM

(Generalized Additive Models)

Intention of GAM

- Extend traditional linear regression and the linear logistic model
- GAM‘s replace linear combination of predictors with a sum of smooth, non-parametric functions
- Allowing to uncover non-linear covariate effects

How GAM work?

- Modeling relationship between target variable and each predictor as an additive

- Smooth functions can be estimated by different techniques
  - Initially a local scoring algorithm using scatterplot smoother

$$\sum_{j=1}^{p} \beta_j X_j \rightarrow \sum_{j=1}^{p} s_j(X_j)$$

Benefits and Drawbacks

Benefits

- Model complex, non-linear relationship without pre-specifying form of relationship

- Predictors effect are modelled separately
  - Easier interpretation of results and predictors effects

Drawbacks

- Estimating smooth functions can be computationally intensive

NAM

(Neural Additive Models)

Intention of NAM

- Combine Neural Networks (NN) with interpretable approaches
- Retain flexibility and scalability of NN to learn non-linear, complex relationships in data with efficient training
- GAM tend to over regularize and miss genuine details in real data
- Enable NN for high-stakes applications (e.g. healthcare)

How NAM work?

- GAM use local score functions $s_j(X_j)$ to predict the contribution of each predictor separately

- Use NN to model each predictors contribution separately

$$\sum_{j=1}^{p} \beta_j X_j \rightarrow \sum_{j=1}^{p} s_j(X_j) \rightarrow \sum_{j=1}^{p} NN_j(X_j)$$

Benefits and Drawbacks

Benefits

- Retain flexibility of NN, capturing non-linear, complex relationships
- Improve performance compared to GAM
- Efficient training with GPUs

Drawbacks

- Increased model complexity due to NN nature
- NN are flexible but tend to overfit

NODE

(Neural Oblivious Decision Ensembles)

Intention of NODE

- Overcome limitations of DNN on tabular data

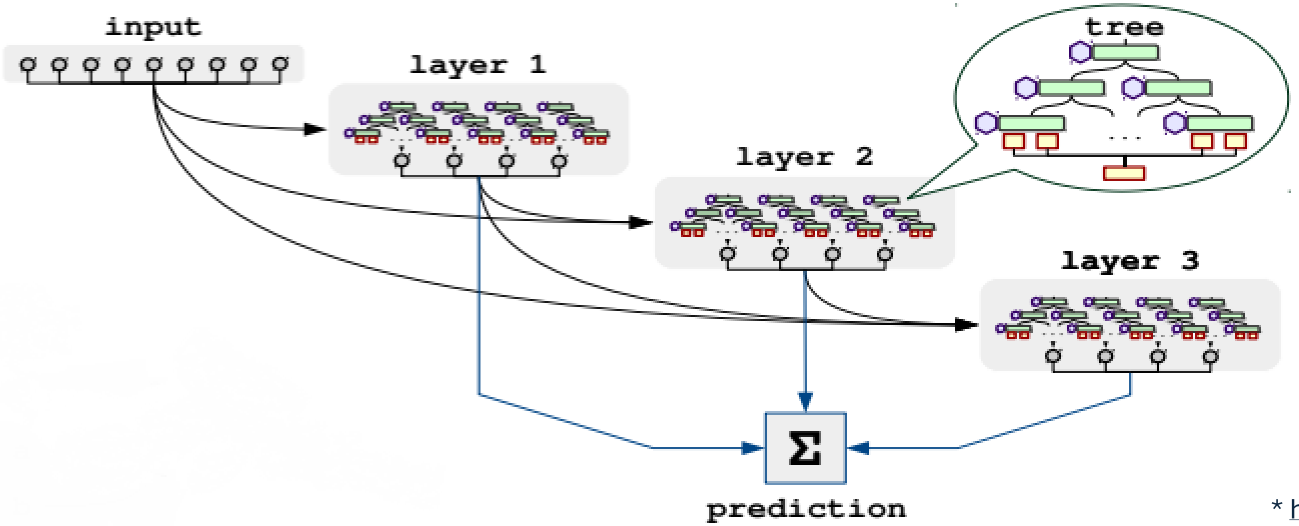
- Gradient-Boosted Decision Trees (GDBT) often perform better on tabular data compared to DNN

- NODE combines layer architecture of DNN and decision trees aiming to keep differentiability and robustness

How NODE work?

- Oblivious Decision Trees (ODT) split data sharing the predictors and thresholds across all internal nodes of the same depth

- NODE consists of differentiable ODT that are trained end-to-end by backpropagation



* https://doi.org/10.48550/arXiv.1909.06312

Benefits and Drawbacks

Benefits

- Great performance on tabular data
- Differentiable
- End-to-end training

Drawbacks

- Computationally more expensive compared to other state-of-the-art approaches
- Lack of interpretability since interactions between features

NODE-GAM

Intention of NODE-GAM

- Combine interpretability of GAM, differentiability of NN, and robustness of Oblivious Decision Trees (ODT)

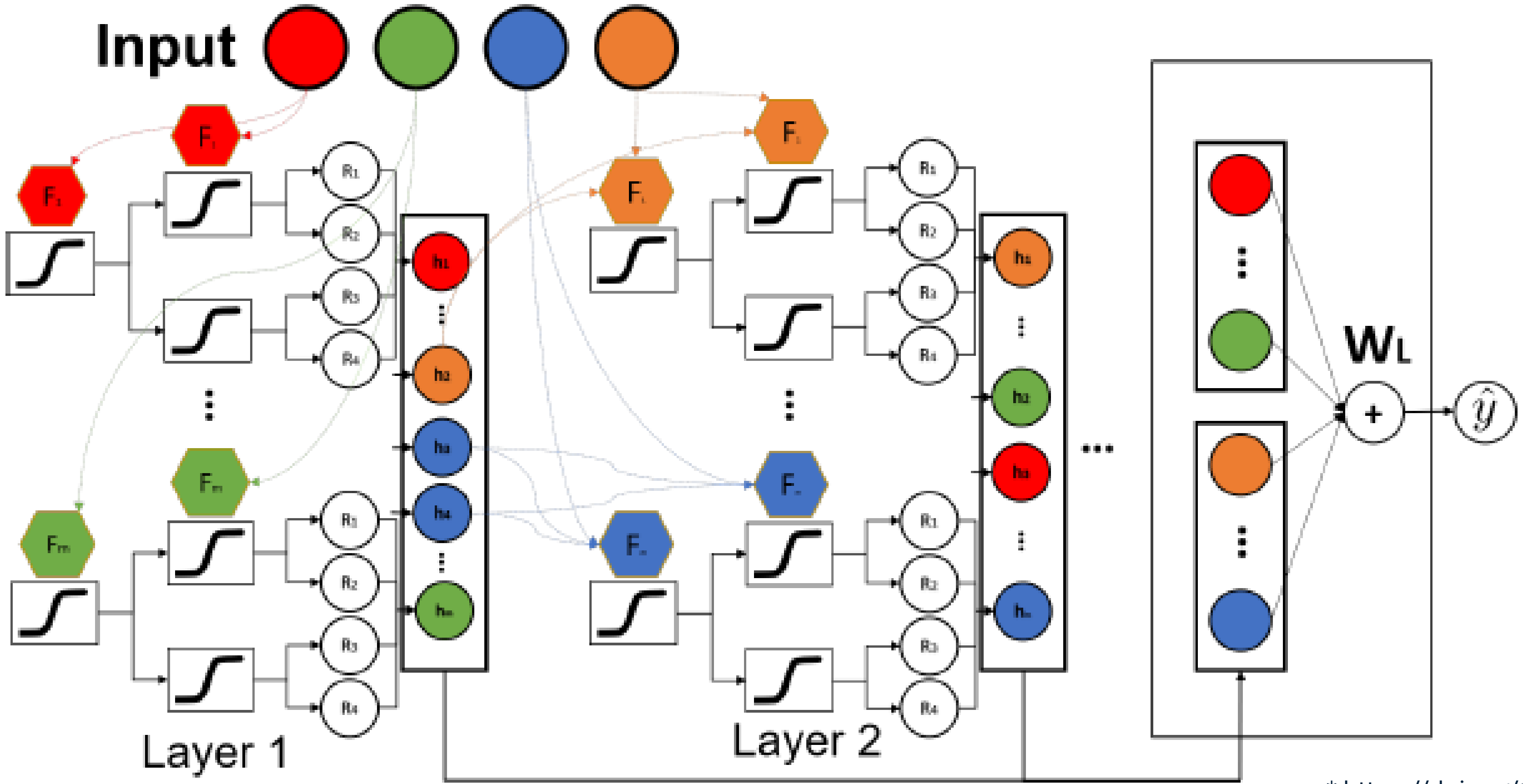
- Enforce no interaction of predictors between tree connections
  - NODE-GA^{2}M is an extension that allows interaction between at most 2 predictors to interact within each tree

How NODE-GAM work?

- Integrates the architecture of NODE with additive structure of GAM
- Each neural layer consists of multiple differentiable ODT
  - A single ODT takes only one predictor as input
- Output of previous ODT are given to an ODT in the leading neural layer as well as the model input

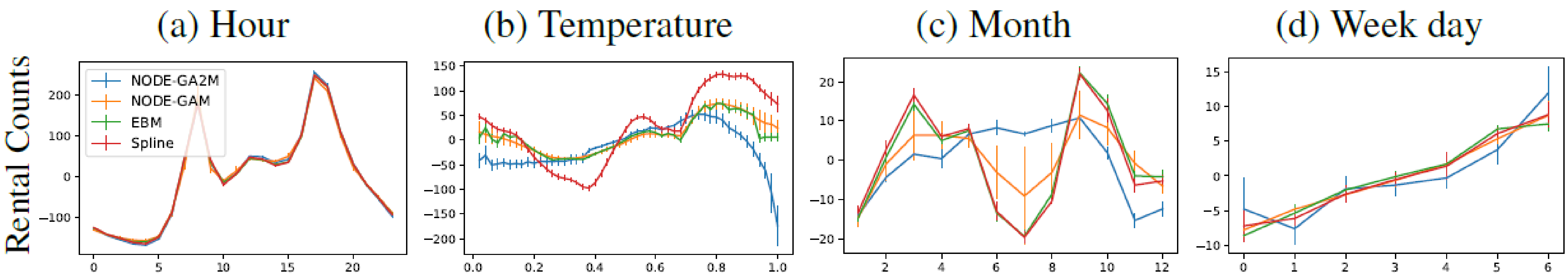
- Output of all layers/ODT are weighted and summed up to final model output

How NODE-GAM work?



* https://doi.org/10.48550/arXiv.2106.01613

What do insights by NODE-GAM look like?



* https://doi.org/10.48550/arXiv.2106.01613

Comparison

Comparison

Classification

- NODE-GAM and NODE-GA^{2}M are competitive, often matching or exceeding performance of other GAMs and Full Complexity models
- Perform better on datasets with less lab
- Improves performance on large datasets

Regression

- NODE-GAM and NODE-GA^{2}M are not as competitive as on classification datasets
- Gets beaten by Random Forests and NODE significantly

Conclusion

Limitations and Benefits

- GAM show association patterns and not causation
- NODE-GAM and NODE-GA^{2}M are not always a good choice but often

- GAM can answer more questions accurately, resulting in higher confidence in explanations
- GAM helps users better to discover patterns and understand importance of predictors compared to Decision Trees
- Great tools to discover biases within data to avoid false conclusions

Thank you for listening!

Sources

- **Introduction:** https://dl.acm.org/doi/pdf/10.1145/3233547.3233667
- **GAM:** https://www.jstor.org/stable/2245459
- **NAM:** https://doi.org/10.48550/arXiv.2004.13912
- **NODE:** https://doi.org/10.48550/arXiv.1909.06312
- **NODE-GAM:** https://doi.org/10.48550/arXiv.2106.01613