NODE-GAMLSS: Interpretable Uncertainty Modelling via Attention-Based Deep Distributional Regression

Ananyapam De Anton Thielmann Benjamin Säfken
TU Clausthal

TU Clausthal
Clausthal University of Technology

NEURAL INFORMATION PROCESSING SYSTEMS

Code & Data & Model:
https://github.com/AnFreTh/NodeGAMLSS

Introduction

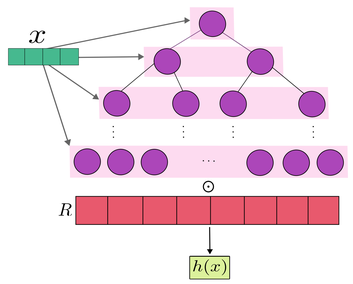
- We propose NODE-GAMLSS, an interpretable attention-based deep learning architecture for distributional regression.
- NODE-GAMLSS models data-dependent location, scale and shape (LSS) parameters, enabling quantile prediction and feature effect interpretation.

Overview

- Given covariates $\mathbf{X}$ and target $\mathbf{y}$, NODE-GAMLSS models the conditional distribution $\mathbf{y} \sim \mathcal{D}(\boldsymbol{\theta}(\mathbf{X}))$ where $\boldsymbol{\theta}(\mathbf{X})$ are the distributional parameters and the k-th one is:

$$\theta^{(k)} = \eta^{(k)}(\beta^{(k)} + \sum_{j=1}^{J} z_j^{(k)}(x_j))$$

- Modeling of the LSS parameters enables quantile estimation uncertainty modelling and interpretability at the feature level



Method

Each feature function z_j consists of L layers with I differentiable Oblivious Decision Trees (ODTs) of depth c. An ODT compares input features to splitting thresholds and returns weighted response:

$$h(\boldsymbol{x}) = R \odot \left(\bigotimes_{c=1}^{C} \begin{bmatrix} \sigma\left(F(\boldsymbol{x}) \leq b^c\right) \\ \sigma\left(F(\boldsymbol{x}) > b^c\right) \end{bmatrix} \right)$$

The feature function for tree i in layer l is:

$$F_{li}(\boldsymbol{x}) = \sum_{j=1}^{d} x_j \cdot G_{ij} + \frac{1}{\sum_{\hat{l}=1}^{l-1}\sum_{\hat{i}=1}^{I} g_{l\hat{i}i}} \sum_{\hat{l}=1}^{l-1}\sum_{\hat{i}=1}^{I} h_{\hat{l}\hat{i}}(\boldsymbol{x}) g_{l\hat{i}i} a_{l\hat{i}i}$$

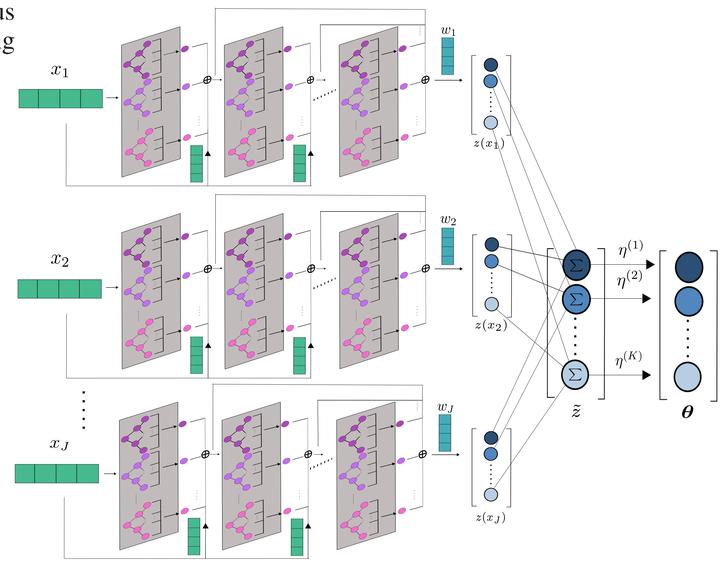
Here, $G_i = \text{entmax}_\alpha\left(F_i/T\right)$ and $g_{li\hat{i}} = G_i G_{\hat{i}}$.

$$a_{l\hat{i}i} = g_{l\hat{i}i}\ \text{entmax}_\alpha\left(\log\left(\boldsymbol{g}_i\right) + \boldsymbol{A}_{li}\right)_{\hat{i}}.$$

$$\boldsymbol{x}^1 = \boldsymbol{x}, \boldsymbol{x}^l = \left[\boldsymbol{x}, h^1(\boldsymbol{x}^1), \ldots, h^{(l-1)}(\boldsymbol{x}^{(l-1)})\right] \text{ for } l > 1$$

The final prediction is given by a weighted average

$$z^{(k)}(\boldsymbol{x}) = \frac{1}{LI}\sum_{l=1}^{L}\sum_{i=1}^{I} h_i^l(\boldsymbol{x}^l) w_{li}$$



Experiments & Results

Comparison with Competing Methods:

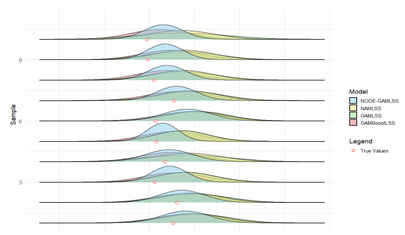
Benchmark comparison on Normal response

Dataset	Model	NLL	MSE	MAE
Abalone	NODE-GAMLSS	**0.951 ± 0.024**	0.492 ± 0.044	**0.494 ± 0.022**
	NAMLSS	1.078 ± 0.079	1.076 ± 0.098	0.787 ± 0.036
	GAMLSS	1.056 ± 0.034	**0.483 ± 0.032**	0.501 ± 0.017
	GAMBoostLSS	0.998 ± 0.031	0.536 ± 0.046	0.732 ± 0.032
California	NODE-GAMLSS	**0.726 ± 0.023**	**0.314 ± 0.011**	**0.392 ± 0.004**
	NAMLSS	0.785 ± 0.047	0.758 ± 0.051	0.654 ± 0.026
	GAMLSS	0.917 ± 0.020	0.366 ± 0.015	0.442 ± 0.004
	GAMBoostLSS	1.025 ± 0.182	0.420 ± 0.011	0.648 ± 0.009
Insurance	NODE-GAMLSS	**0.556 ± 0.147**	**0.151 ± 0.027**	**0.226 ± 0.024**
	NAMLSS	0.653 ± 0.057	0.655 ± 0.069	0.568 ± 0.040
	GAMLSS	0.732 ± 0.048	0.253 ± 0.024	0.503 ± 0.024
	GAMBoostLSS	0.644 ± 0.068	0.269 ± 0.028	0.518 ± 0.028

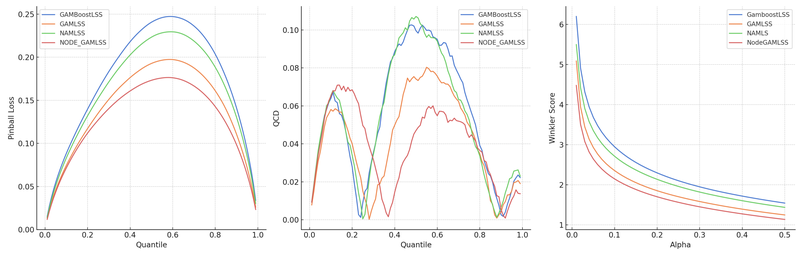
Comparison on using various distributions

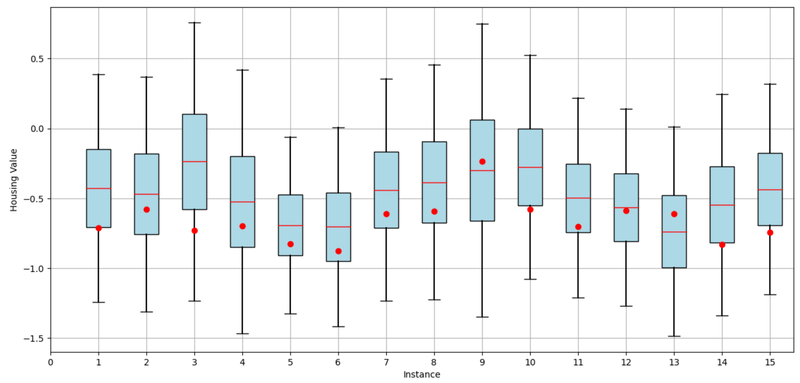
Model	Gamma	Normal	Poisson	Lognormal
NODE-GAMLSS	0.833 ± 0.072	**1.422 ± 0.027**	**1.375 ± 0.024**	1.514 ± 0.033
NAMLSS	**0.827 ± 0.076**	1.442 ± 0.048	1.386 ± 0.035	**1.504 ± 0.033**
GAMLSS	1.064 ± 0.033	1.432 ± 0.039	1.383 ± 0.005	1.535 ± 0.068
GAMBoostLSS	3.132 ± 0.284	1.491 ± 0.011	1.379 ± 0.045	1.524 ± 0.021

Distributional prediction

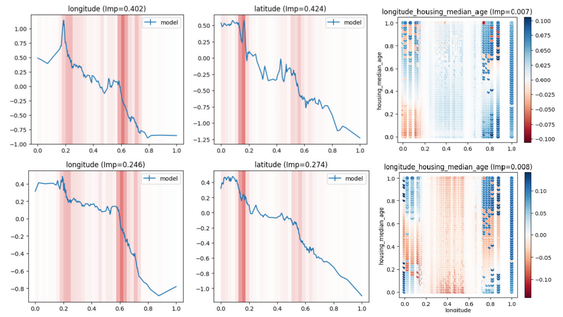


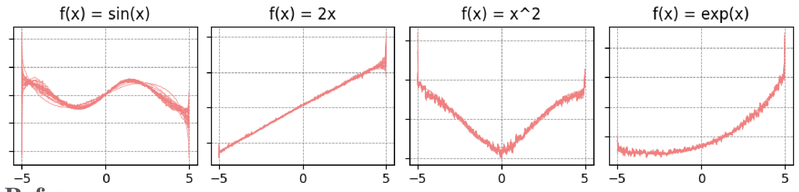
Uncertainty Modelling:





Feature Interpretability:





References:

[1] NAMLSS [Thielmann *et al.*, AISTATS20]
[2] NODE-GAM [Chang *et al.*, ICLR22]
[3] NODE [Popov *et al.*, ICLR2019]
[4] GAMLSS [Rigby and Stasinopoulos]