(Nature 2025)
Accurate predictions on small data with a tabular foundation model

May 8, 2025

Seoul National University

Related paper

Accurate predictions on small data with a tabular foundation model (2025, *Nature*)

The tabular foundation model TabPFN outperforms specialized time series forecasting models based on simple features (2024, *NeurIPS Workshop*)

Drift-resilient tabPFN: In-context learning temporal distribution shifts on tabular data (2024, *NeurIPS Workshop*)

Bayes' Power for Explaining In-Context Learning Generalizations (2024, *arXiv*)

Large language models for automated data science: Introducing caafe for context-aware automated feature engineering (2023, *NeurIPS*)

Efficient bayesian learning curve extrapolation using prior-data fitted networks (2023, *NeurIPS*)

PFNs4BO: In-Context Learning for Bayesian Optimization (2023, *ICML*)

TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second (2022, *NeurIPS Oral*)

Transformers Can Do Bayesian Inference (2022, *ICLR*)

PFN

Transformers Can Do Bayesian Inference (2022, ICLR)

Algorithm 1: Training a PFN model by Fitting Prior-Data

Input : A prior distribution over datasets $p(\mathcal{D})$, from which samples can be drawn and the number of samples K to draw

Output : A model q_θ that will approximate the PPD

Initialize the neural network q_θ;

for $j \leftarrow 1$ **to** K **do**

  Sample $D \cup \{(x_i, y_i)\}_{i=1}^{m} \sim p(\mathcal{D})$;

  Compute stochastic loss approximation $\bar{\ell}_\theta = \sum_{i=1}^{m}(-\log q_\theta(y_i|x_i, D))$;

  Update parameters θ with stochastic gradient descent on $\nabla_\theta \bar{\ell}_\theta$;

end

Introduction

Prior dataset construction

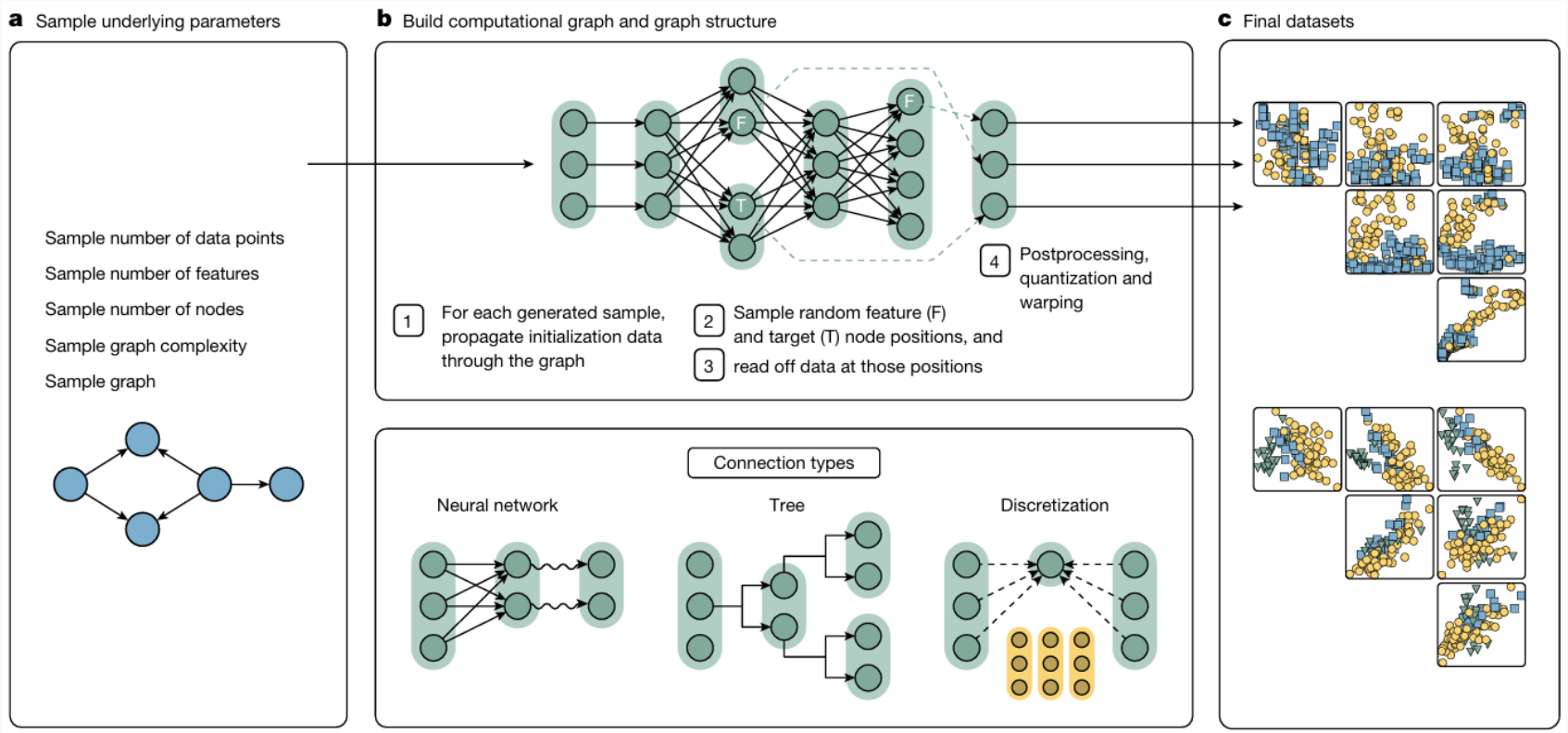
	Dataset			Pre-training		Features		
Version	# Samples	# Features	Structure	# Datasets	Training time	Missing	Categorical	Uninformative
PFN	(?)	60	BNN	500,000 100 steps batch size 100	3:14 hours NVIDIA Tesla V100(?)			
TabPFN	1,024	100	SCM (random MLP edges)	9,216,000 18,000 steps batch size 512	20 hours 8 GPUs (RTX 2080 Ti)	○	○	○
TabPFN v2	2,048	160	SCM (MLP, tree, discretization edges)	128,000,000 2,000,000 steps batch size 64	2 weeks 8 GPUs (RTX 2080 Ti)	○	○	○

Introduction

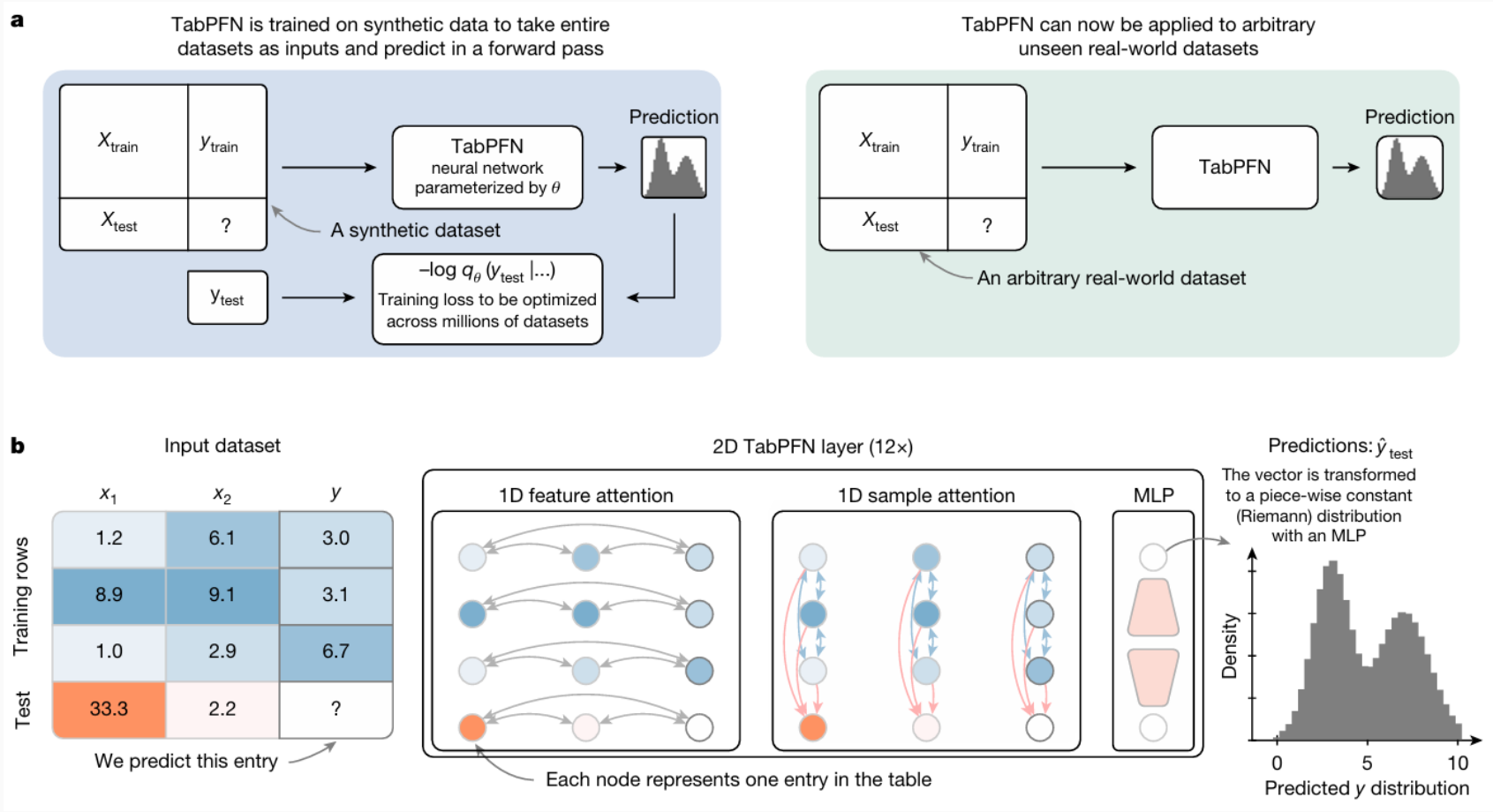
Experiments

	Dataset			Tasks						
Version	Maximum training samples	Maximum features	# Classes	Binary Classification	Multi-class Classification	Regression	Robustness	Density Estimation	Synthetic Data Generation	Representation
PFN	100 (train:test=3:7)	10	2	○						
TabPFN	1,000 (randomly split)	100	10	○	○					
TabPFN v2	10,000 (train:test=9:1)	500	10	○	○	○	○	○	○	○

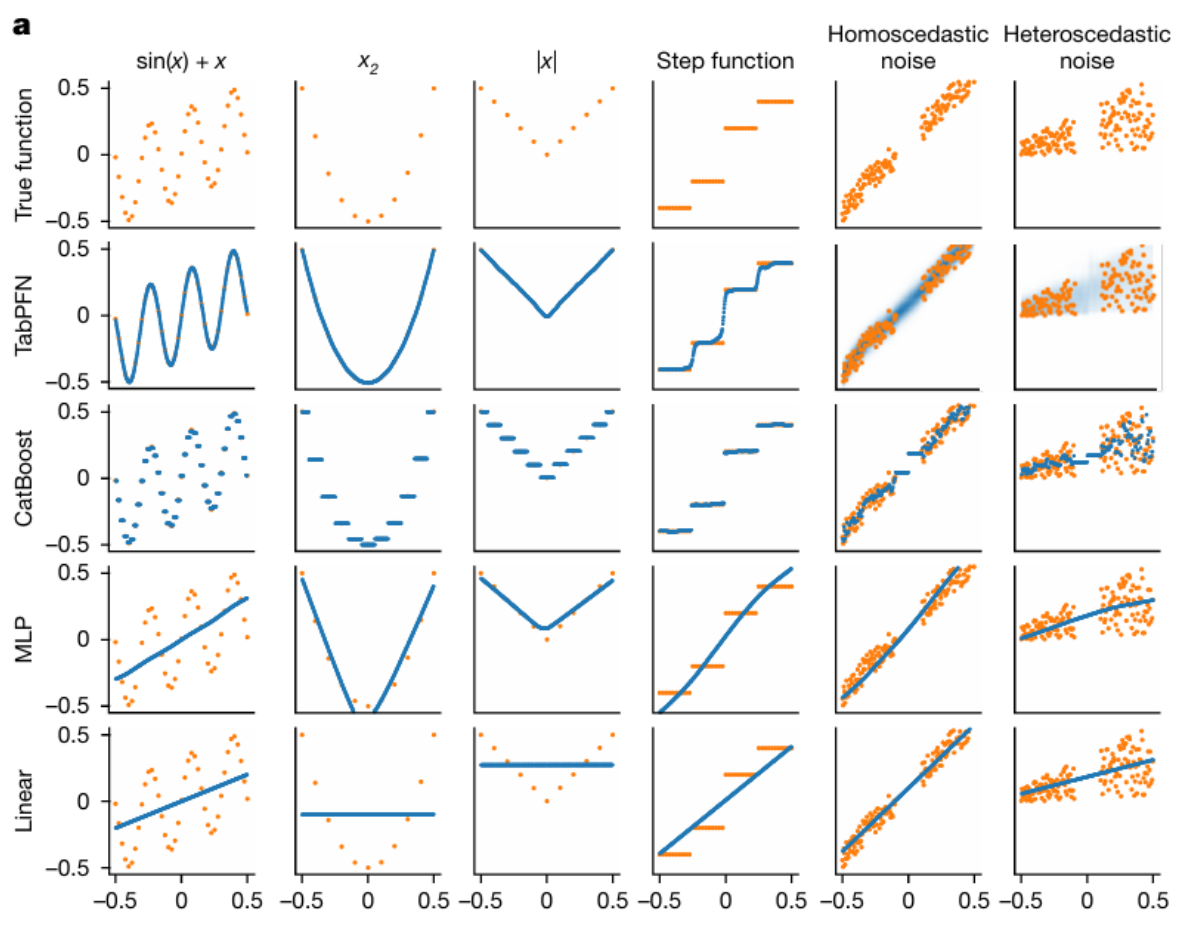
Overview of the TabPFN prior.



Overview of the proposed method

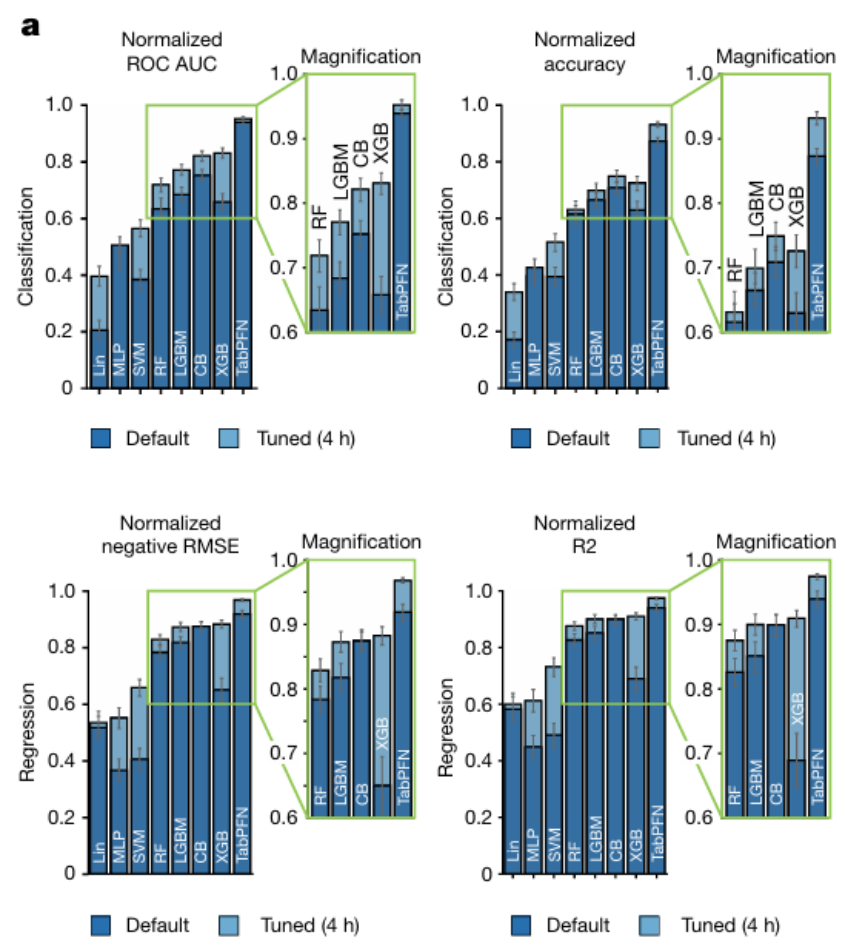


Experiment Result



Orange for the ground truth and blue for model predictions.

Experiment Result



Thank you!