TreeFlow: Going Beyond Tree-based Parametric Probabilistic Regression

Patryk Wielopolski, Maciej Zięba

Wrocław University of Science and Technology, DataWalk, Tooploox

ML in PL Conference 2023

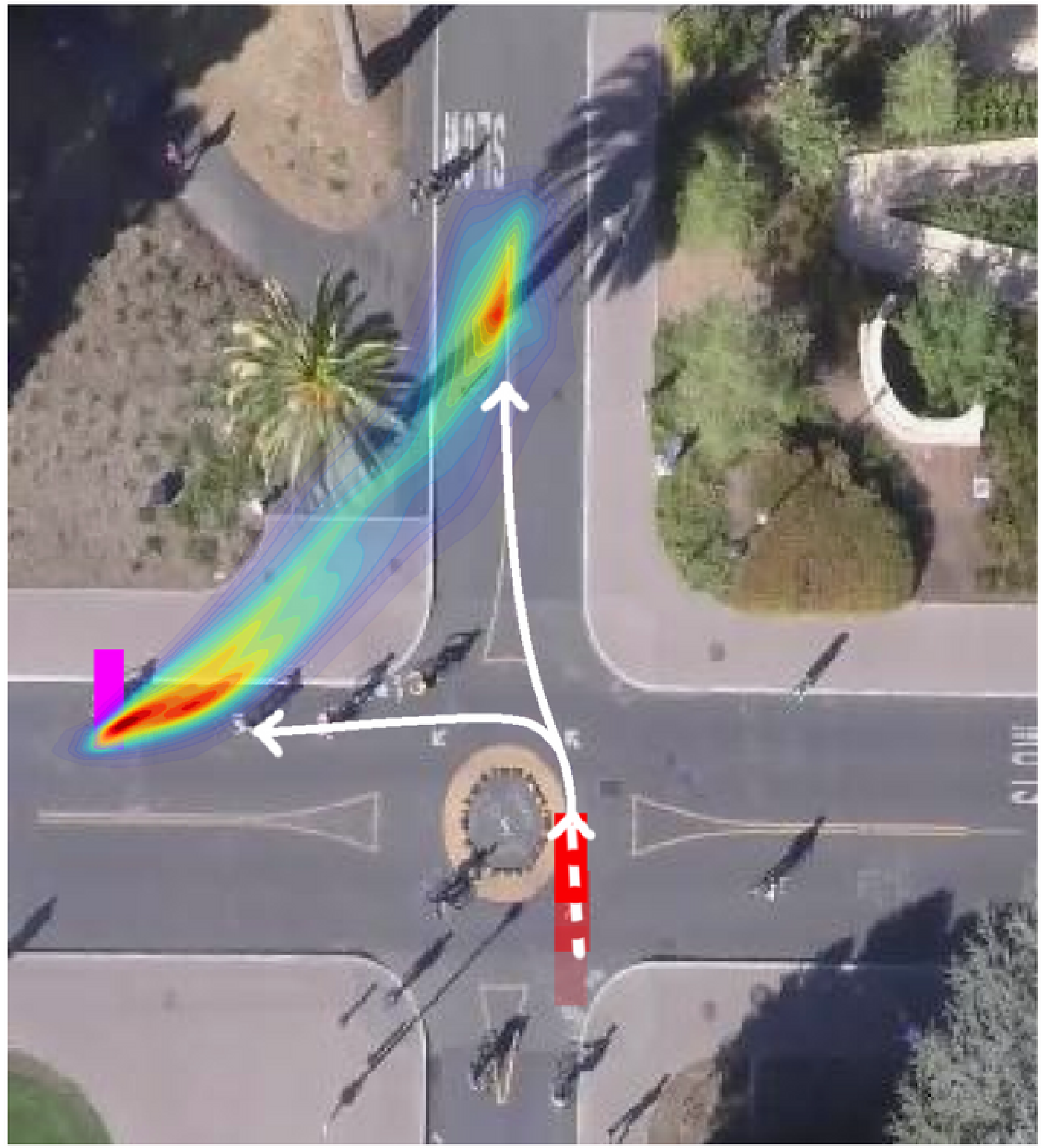
26 - 29 October / Warsaw, Poland

Deterministic vs. Probabilistic Regression

Use Case
Future State Prediction
Car entering a roundabout

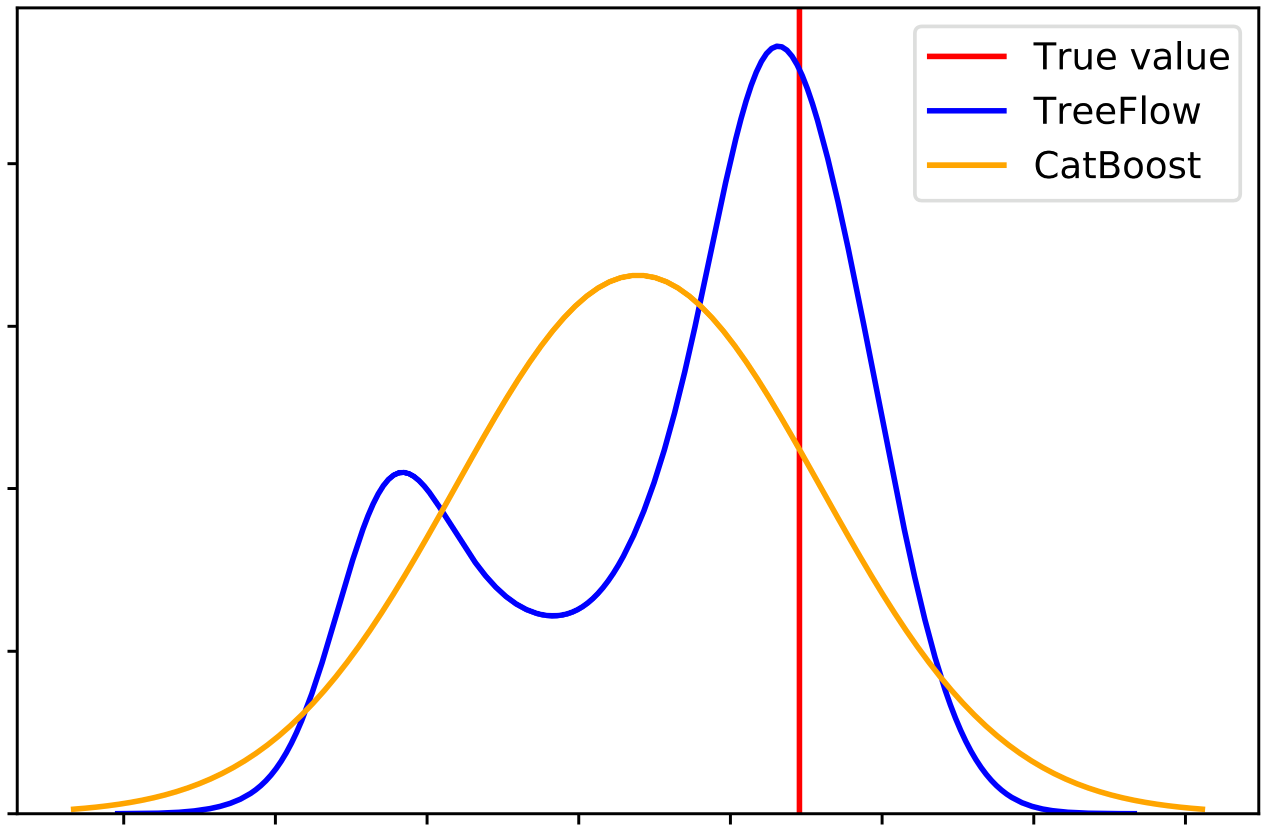
Deterministic Regression
There is exactly only one position.

Probabilistic Regression
There is a distribution of possible positions.

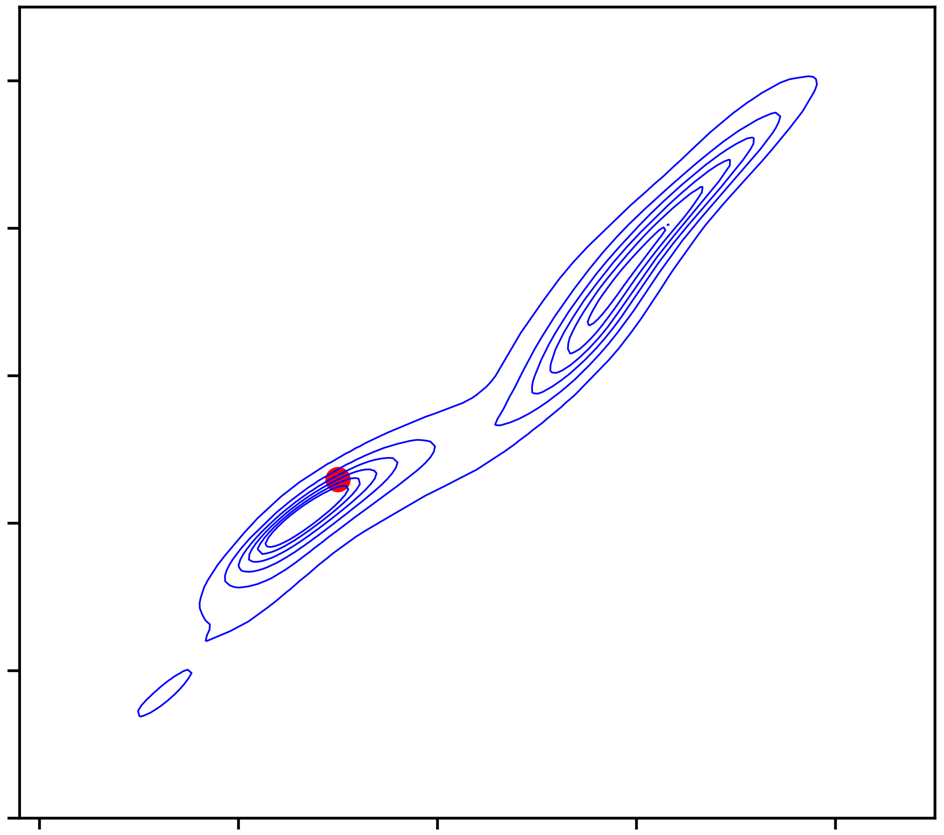


The Quintessence of TreeFlow

Probabilistic Regression for Tabular Data with a Flexible Distribution



Univariate Multimodal Distributions



Multivariate Multimodal Distributions

Fact #1

Tree-based ensembles excel in classification and regression with mixed-type variable tabular data.

scikit learn
Random Forest

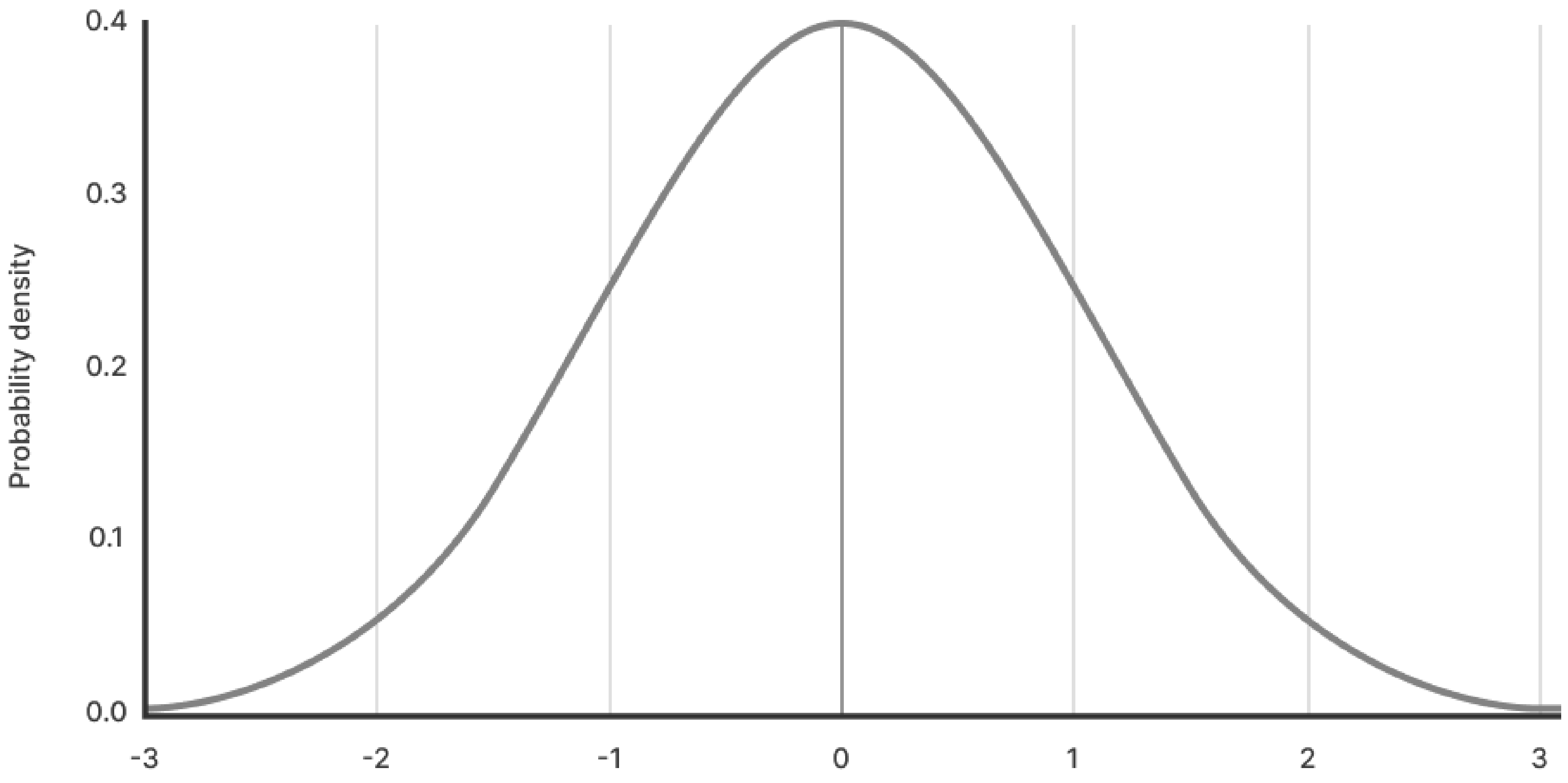
Yandex
CatBoost

dmlc
XGBoost

LightGBM

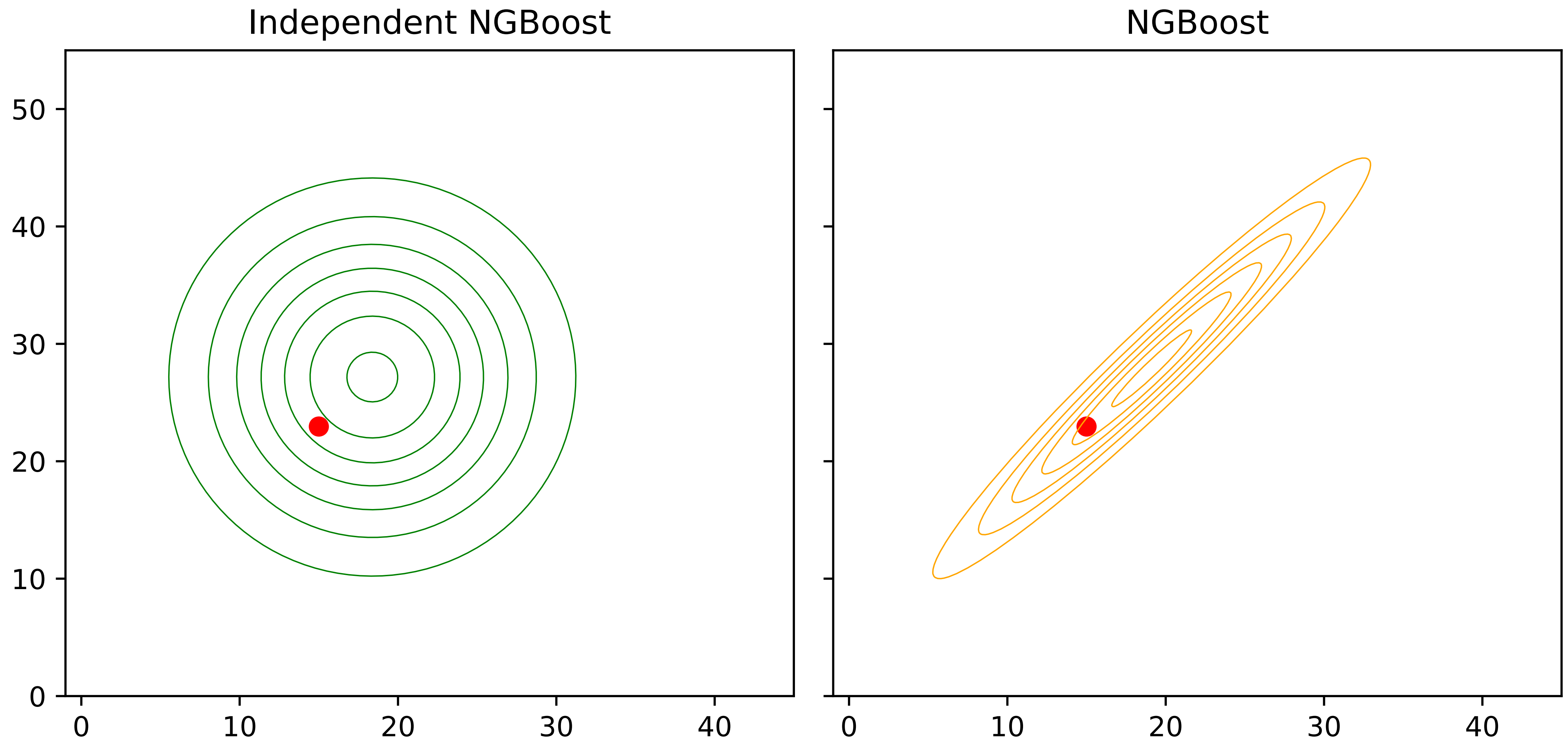
Fact #2

Current approaches use Gaussian or parametric distributions for uncertainty modeling, e.g., NGBoost, CatBoost, PGBM.



Fact #3

Existing methods struggle to handle **multi-modal distributions** and do not support **high-dimensional** probabilistic predictions.

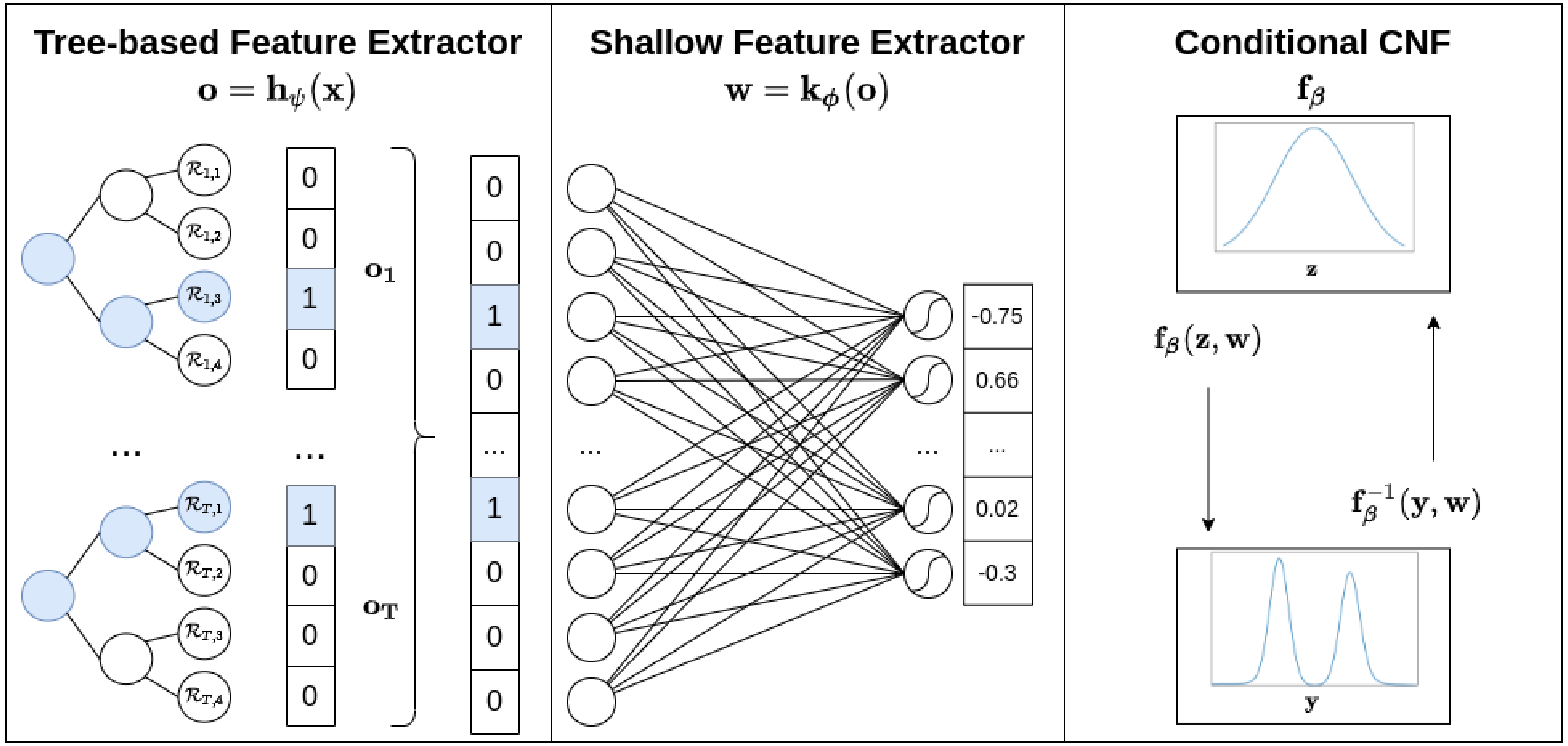


Here comes the TreeFlow!

TreeFlow: Main Characteristics

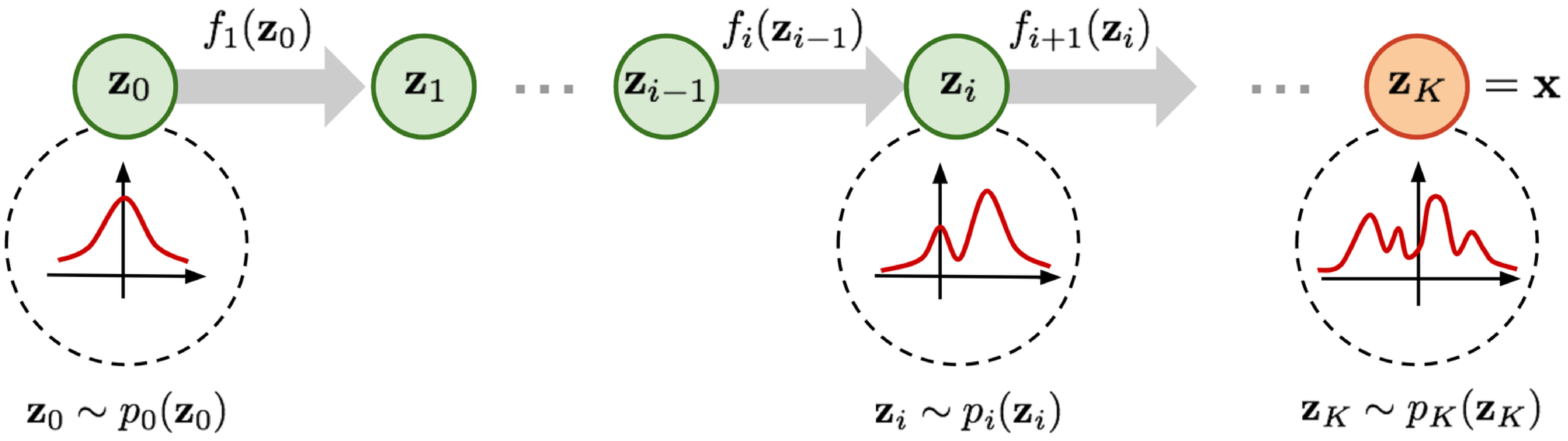
Regression model for **tabular data**
Numerical and **categorical** data
Univariate and **multivariate** targets
Non-Gaussian, **non-parametric** distributions
Probabilistic and **deterministic** predictions

TreeFlow: Architecture

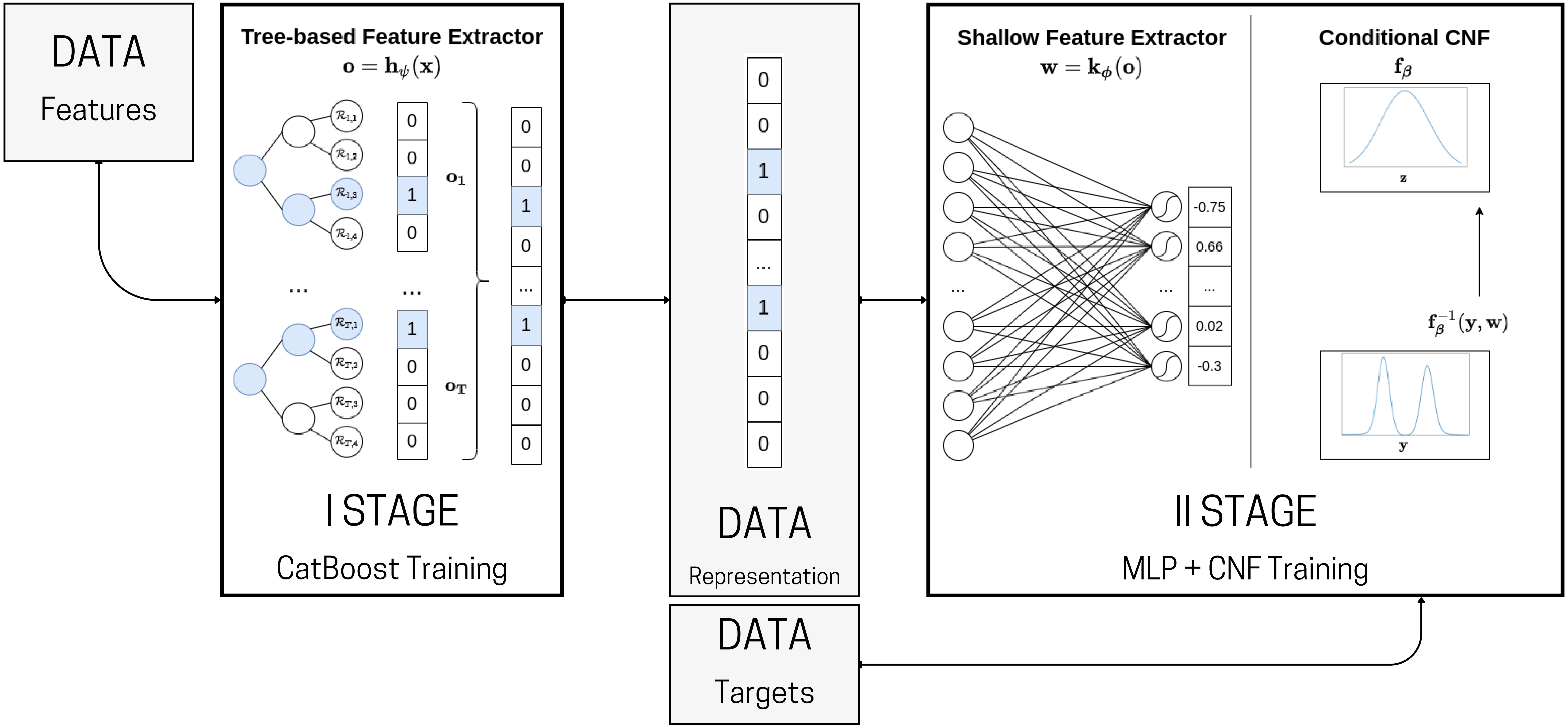


Tree-based Feature Extractor - extract the vector of binary features from the structure of the tree-based ensemble model.
Shallow Feature Extractor - a shallow neural network; maps high-dimensional binary vectors to low-dimensional feature space.
Conditional Continuous Normalizing Flow - takes previous vector as a conditioning factor; models complex probability distribution.

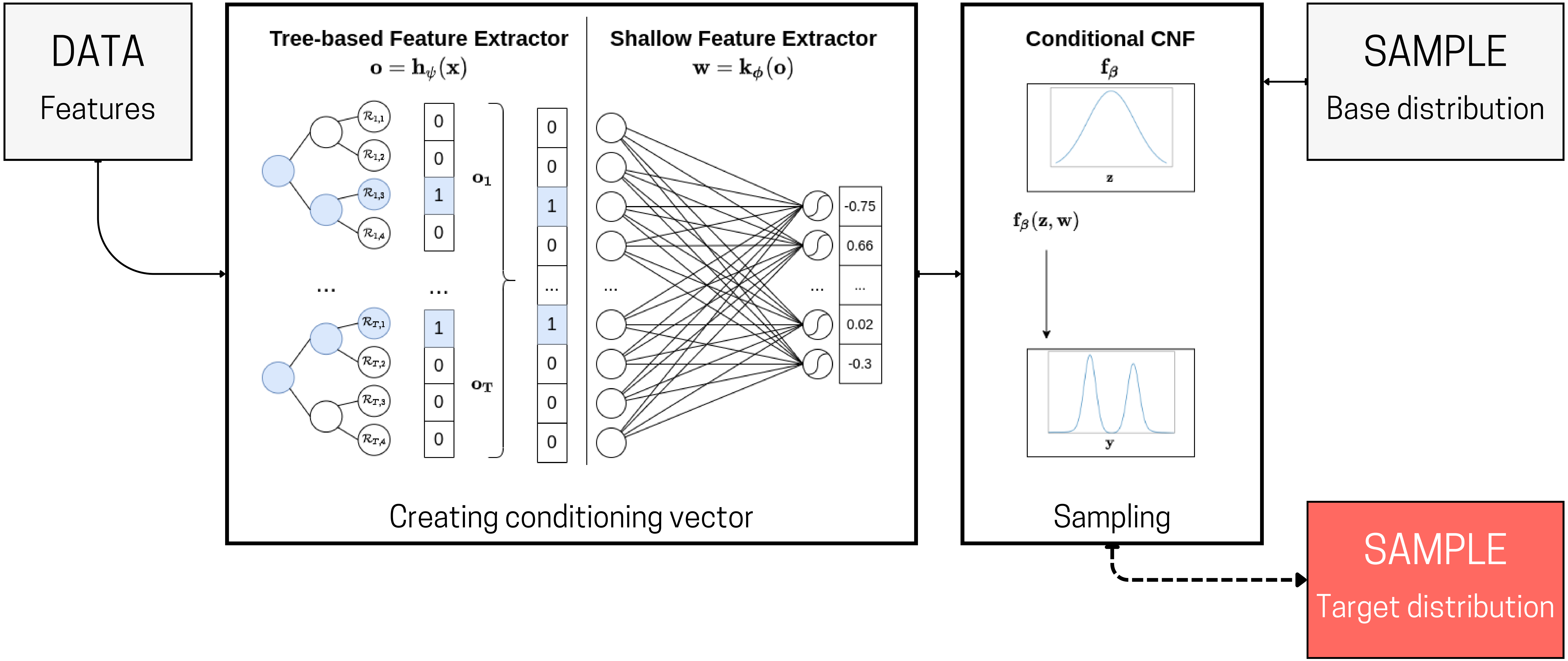
A Word About Normalizing Flows



TreeFlow: Training



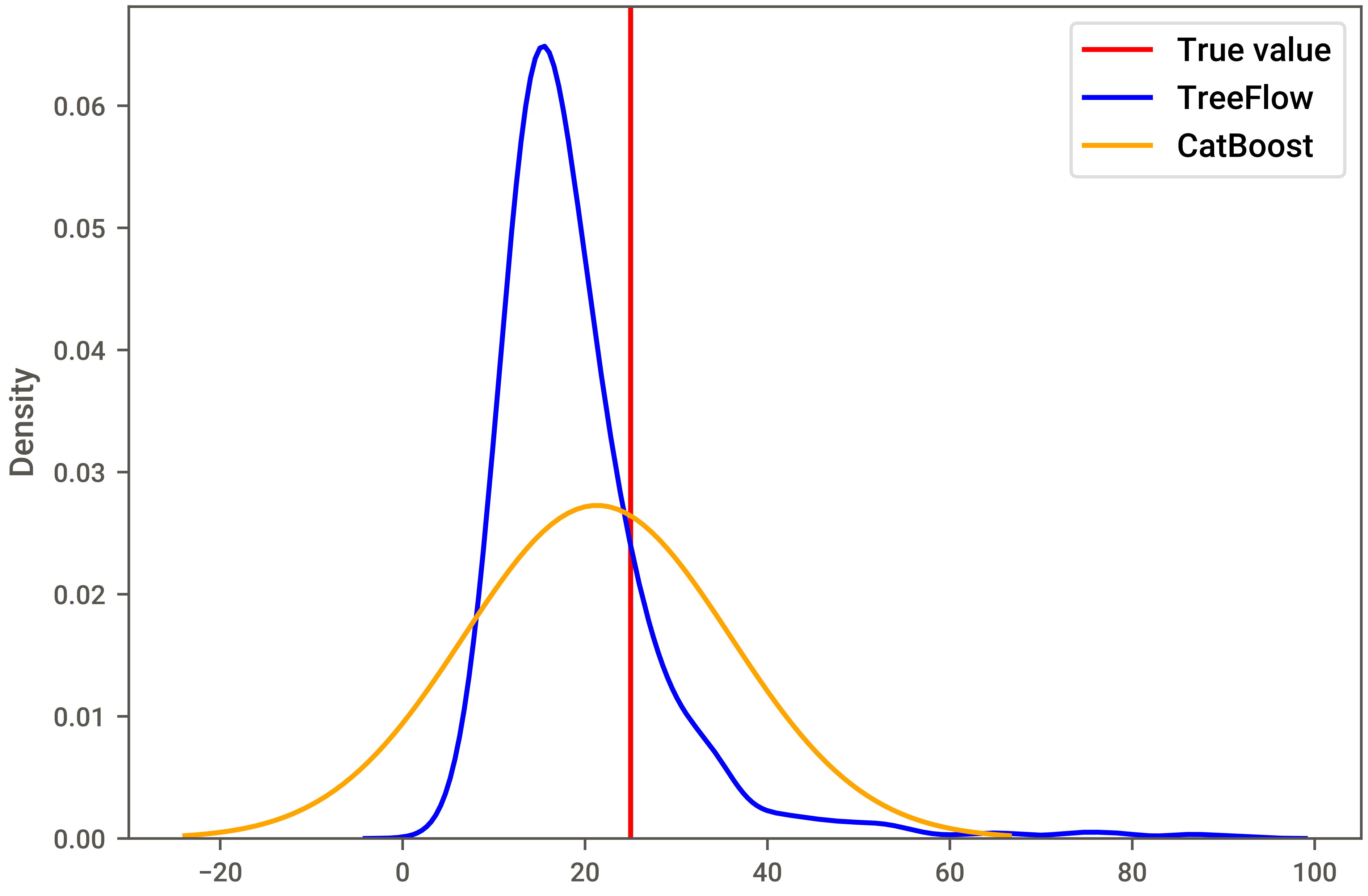
TreeFlow: Sampling



Experiments

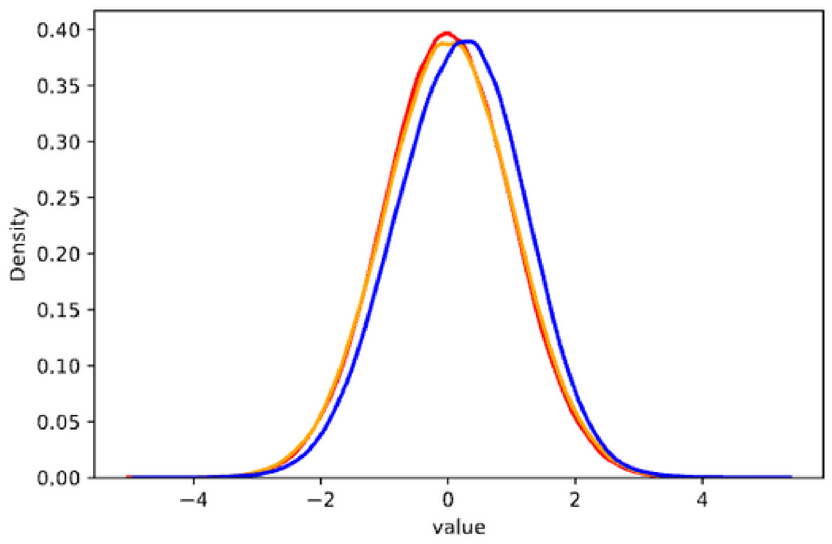
Experiments

Why flexibility of the distribution is important?

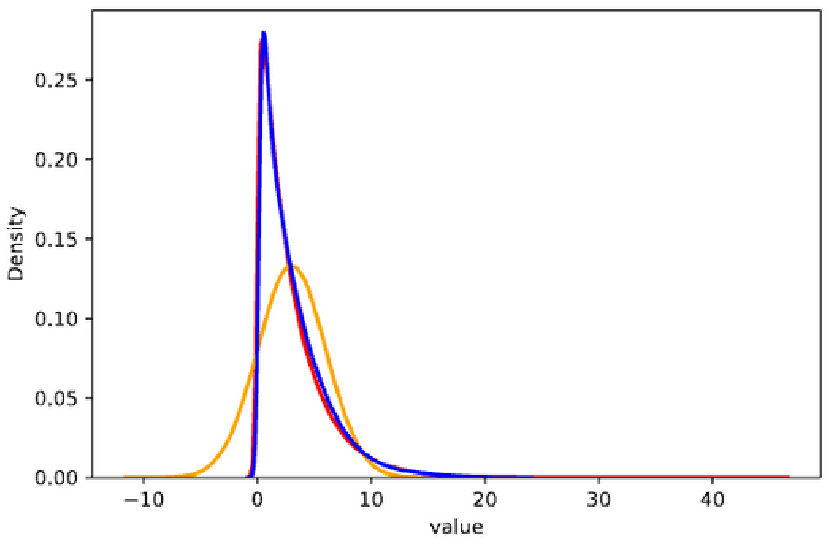


Experiments

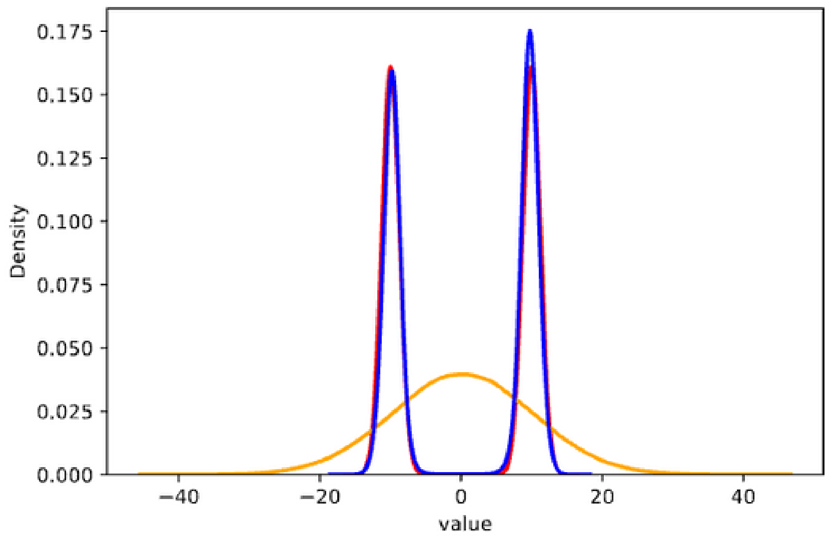
Toy Example



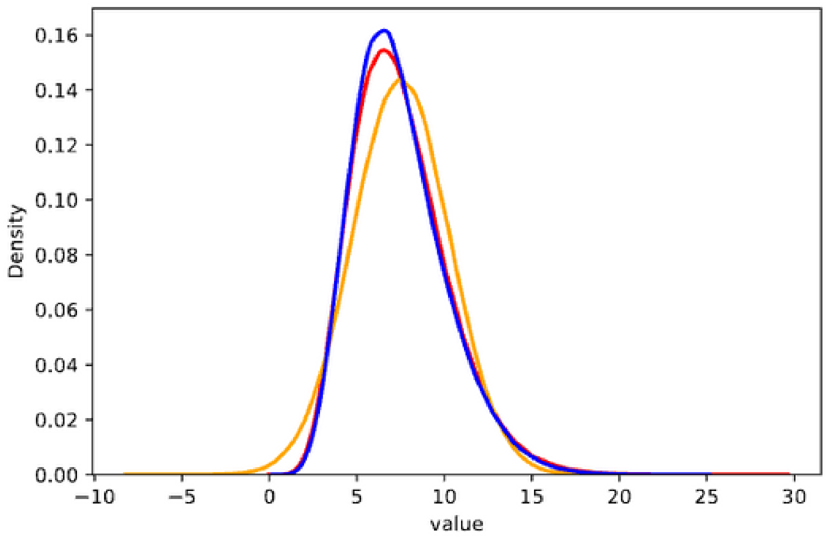
(a) $P(Y|X_1 = 0, X_2 = 0)$



(b) $P(Y|X_1 = 0, X_2 = 1)$



(c) $P(Y|X_1 = 1, X_2 = 0)$

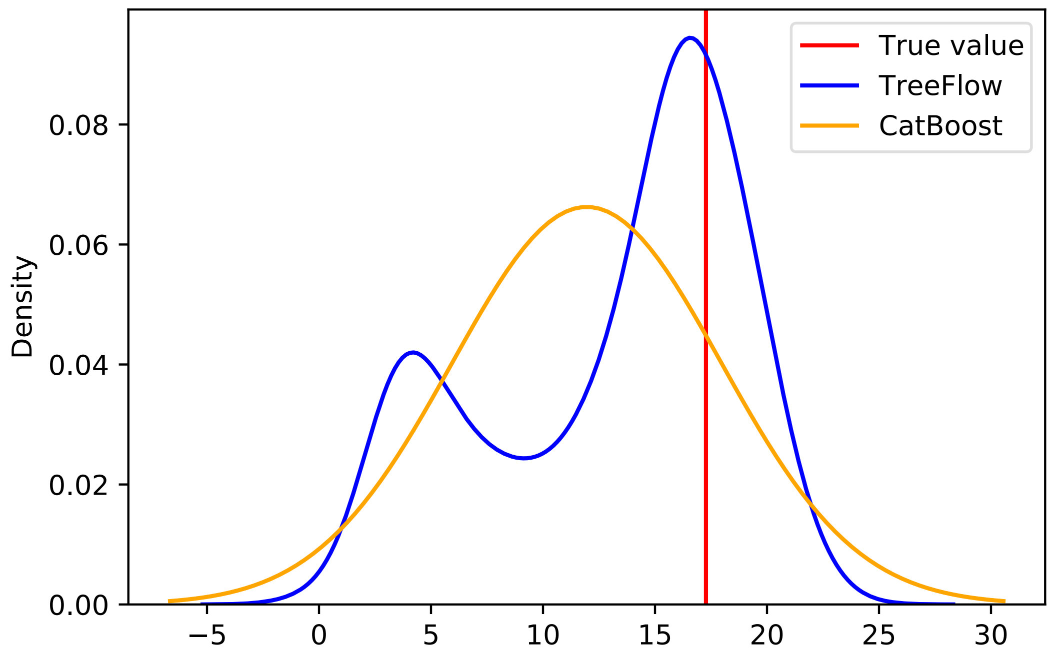


(d) $P(Y|X_1 = 1, X_2 = 1)$

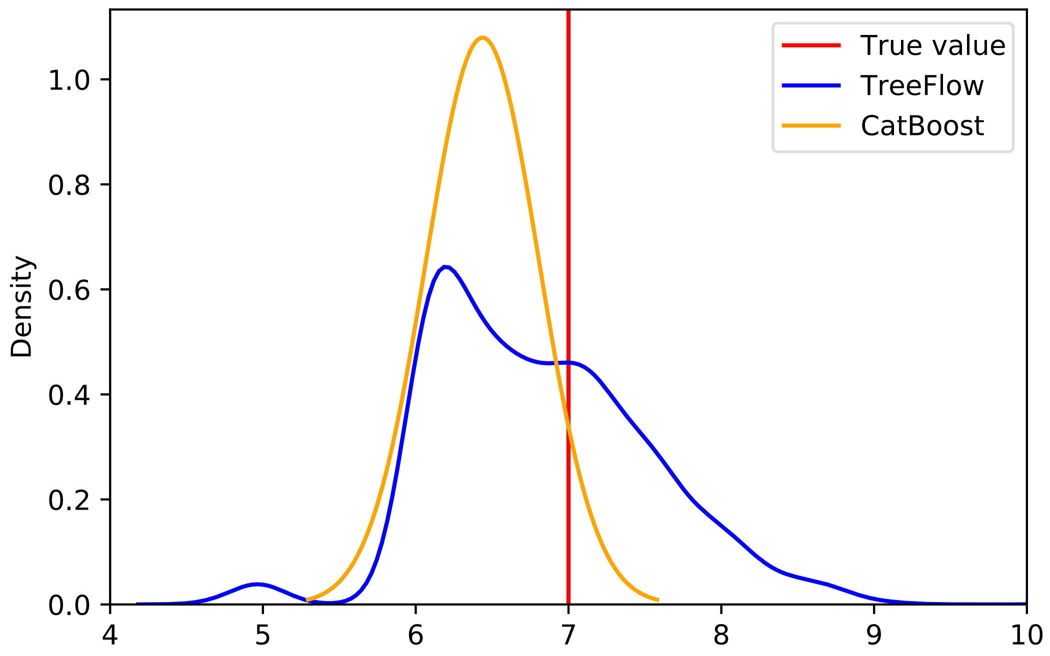
$$
\begin{aligned}
P(Y|X_1 = 0, X_2 = 0) &= \mathcal{N}(Y|\mu = 0, \sigma = 1) \\
P(Y|X_1 = 0, X_2 = 1) &= \mathcal{E}(Y|\lambda = \tfrac{1}{3}) \\
P(Y|X_1 = 1, X_2 = 0) &= \tfrac{1}{2}\mathcal{N}(Y|\mu = -10, \sigma = 1) \\
&\quad + \tfrac{1}{2}\mathcal{N}(Y|\mu = 10, \sigma = 1) \\
P(Y|X_1 = 1, X_2 = 1) &= \Gamma(Y|k = 7.5, \theta = 1.0)
\end{aligned}
$$

Experiments

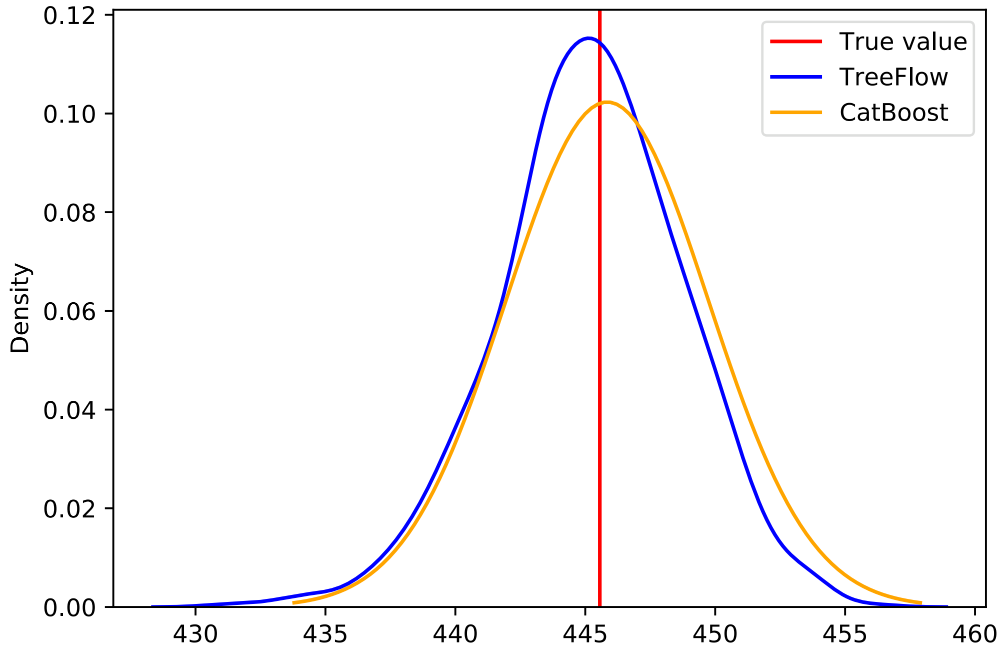
Univariate Flexible Probabilistic Regression



Multimodal Distributions



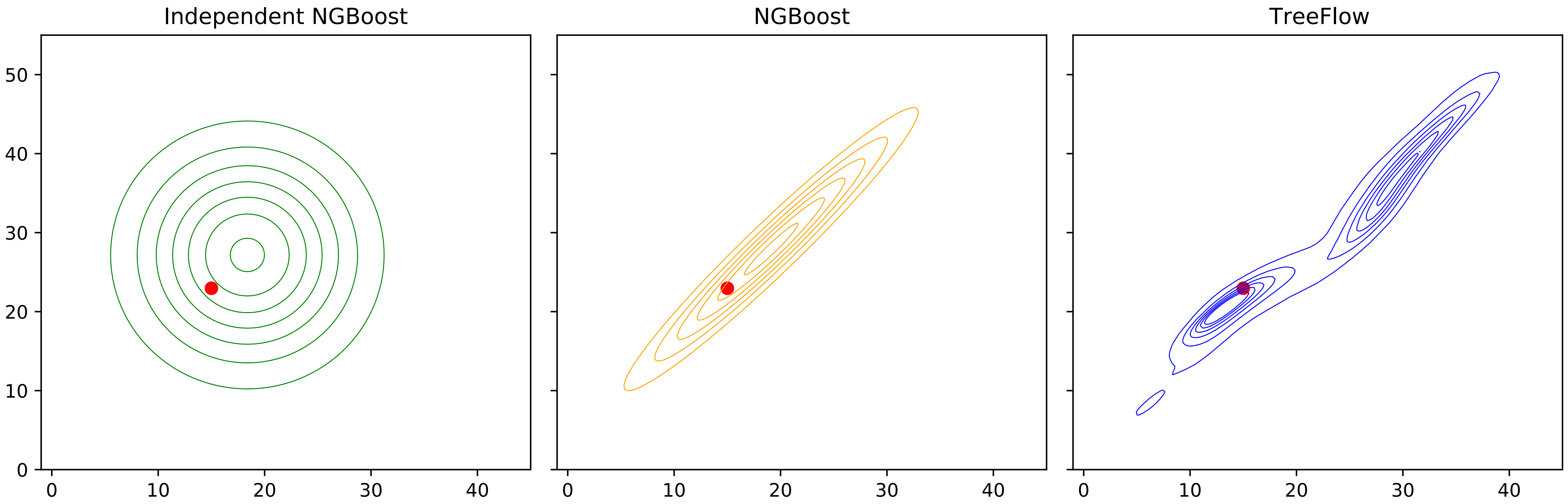
Multimodal, heavy-tailed Dist.



Gaussian Distributions

Experiments

Multivariate Flexible Probabilistic Regression



DATASET	IND NGBOOST	NGBOOST	TREEFLOW
PARKINSONS	6.86	5.85	**5.26**
SCM20D	94.40	94.81	**93.41**
WINDTURBINE	-0.65	-0.67	**-2.57**
ENERGY	**166.90**	175.80	180.00
USFLIGHT	9.56	8.57	**7.49**
OCEANOGRAPHIC	7.74±0.02	**7.73±0.02**	7.84±0.01

Summary

- First-time usage of tree-based models to flexibly model probabilistic regression.
- A novel approach for combining tree-based models with conditional normalizing flows.
- SOTA results for both probabilistic (NLL, CRPS) and deterministic (RMSE) on a variety of datasets.

Thank you for your attention!

Questions?

Paper

Code

Connect