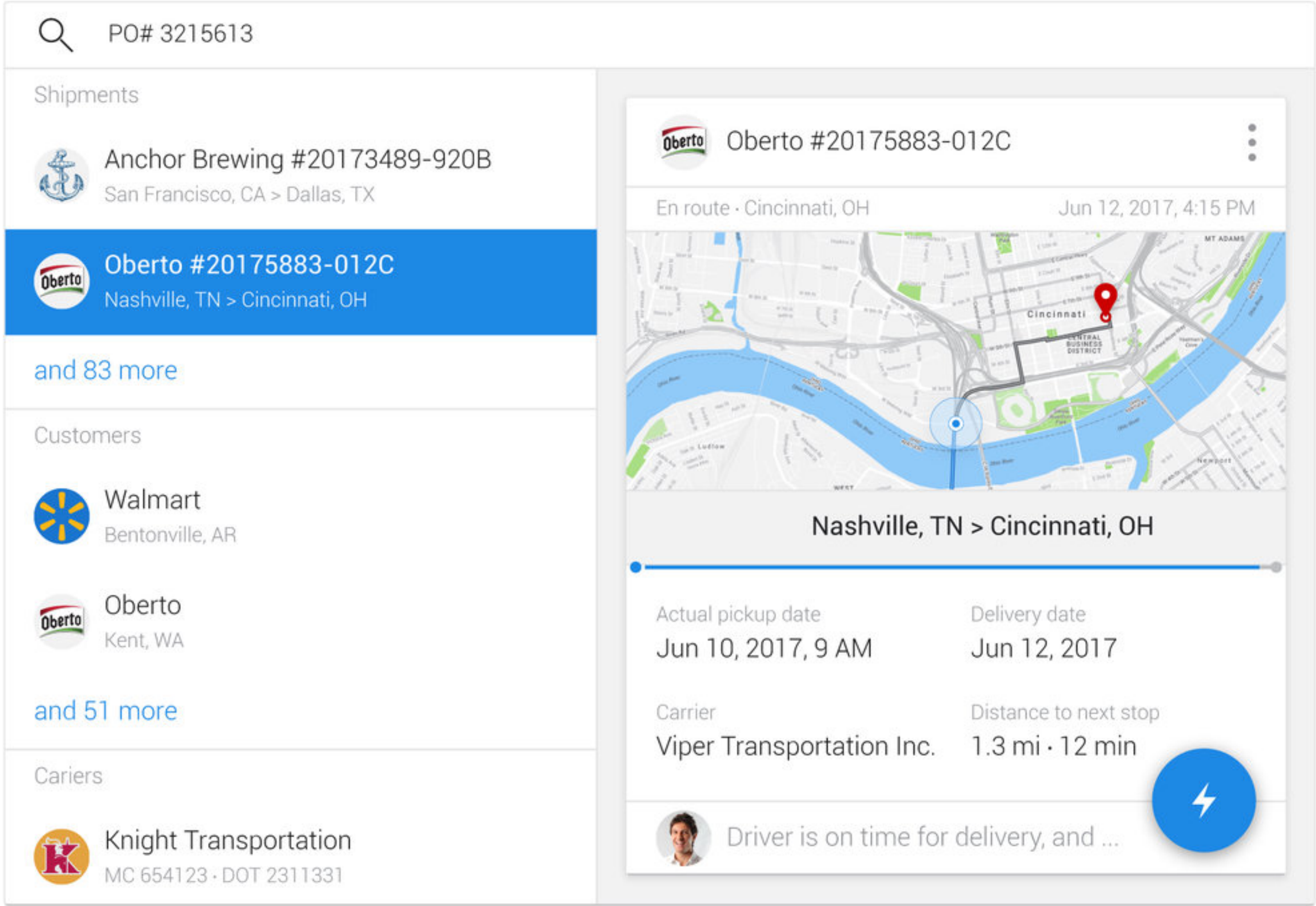
A UNIFIED FRAMEWORK FOR KNOWLEDGE INTENSIVE GRADIENT BOOSTING

Leveraging Human Experts for Noisy Sparse Domains

Harsha Kokel, Phillip Odom, Shuo Yang, Sriraam Natarajan

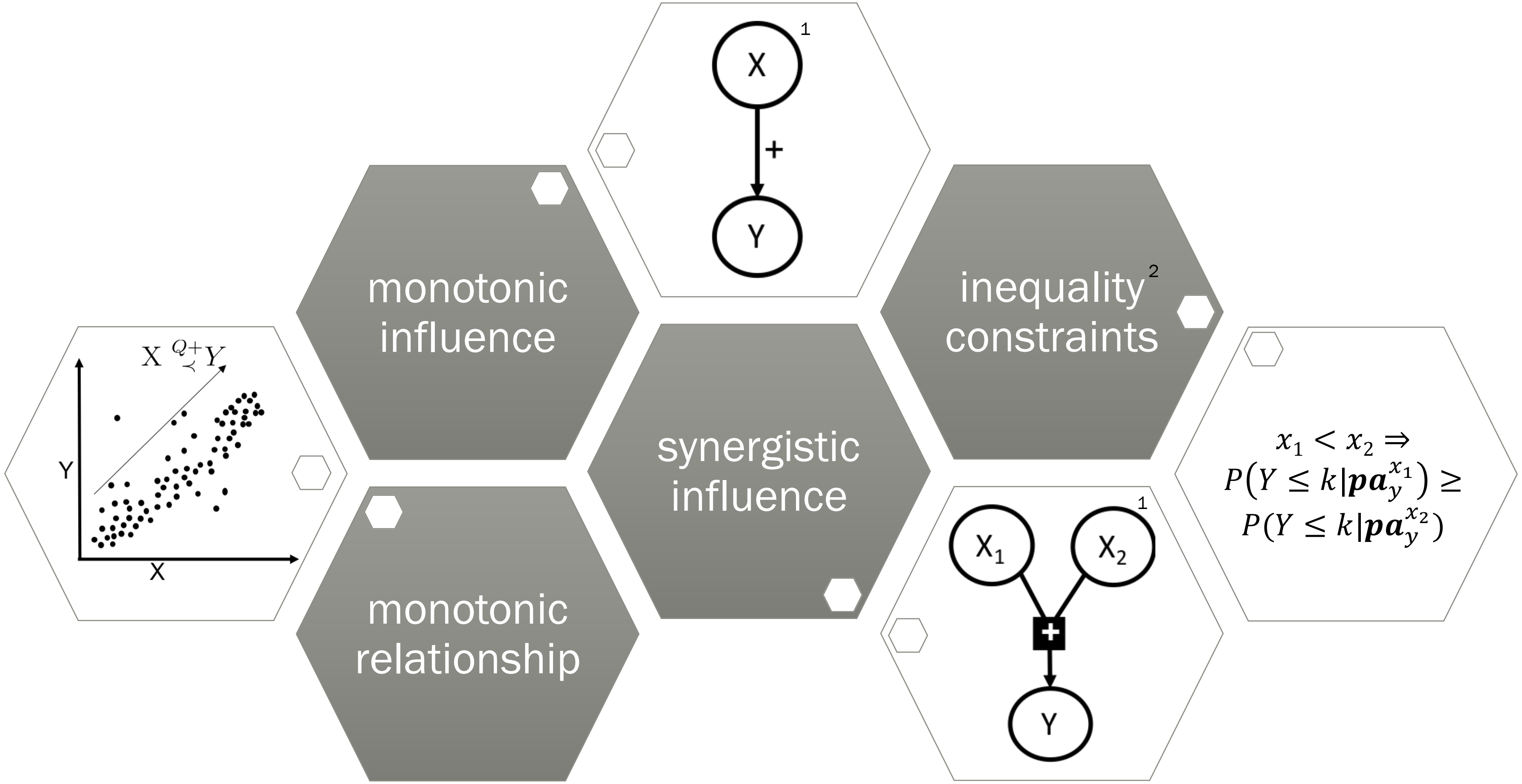
AAAI 2020

Motivation



Turvo

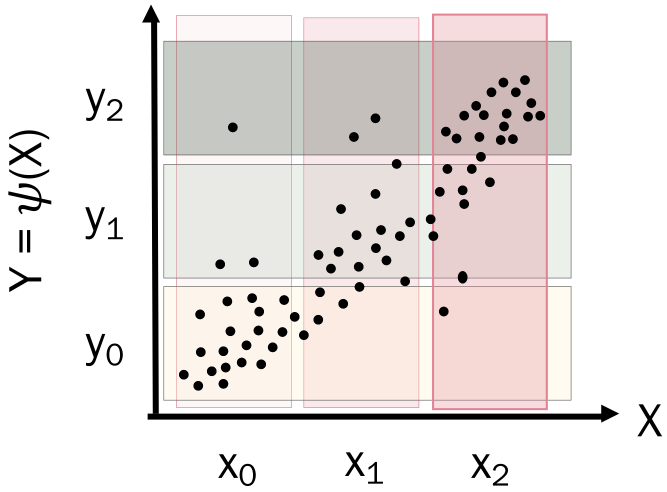
Qualitative influences



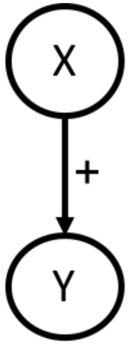
[1]Wellman AI 1990
[2]Robertson, Wright, and Dykstra 1988, Altendorf et al. UAI 2005, Yang and Natarajan ECML-PKDD 2013

Qualitative Constraint

$$X \overset{Q+}{\prec} Y$$



order-restricted constraints



$$x_0 < x_1 \Rightarrow \psi(x_0) \leq \psi(x_1)$$

conditional-probability constraints[2]

$$x_0 < x_1 \Rightarrow P(Y \leq y_1 | x_0) \geq P(Y \leq y_1 | \, x_1)$$

$$x_1 < x_2 \Rightarrow P(Y \leq y_1 | x_1) \geq P(Y \leq y_1 | \, x_2)$$

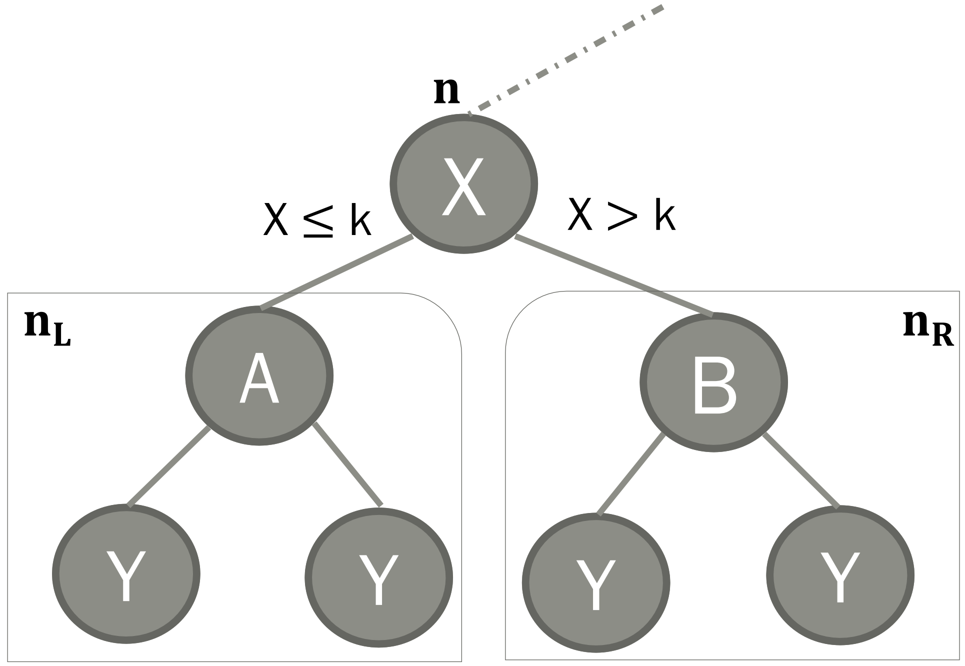
$$x_1 < x_2 \Rightarrow P\left(Y \leq k | \boldsymbol{pa}_y^{x_1}\right) \geq P(Y \leq k | \boldsymbol{pa}_y^{x_2})^*$$

[2]Altendorf et al. UAI 2005, Yang and Natarajan ECML-PKDD 2013

***_caeteris paribus_**

Knowledge-intensive Gradient Boosting

$\mathrm{X} \overset{Q+}{\prec} Y$



$$x_1 < x_2 \Rightarrow \mathbb{E}_{\psi}[x_1 | \ldots] \leq \mathbb{E}_{\psi}[x_2 | \ldots]$$

$$\mathbb{E}_{\psi}[\mathbf{n_L}] \leq \mathbb{E}_{\psi}[\mathbf{n_R}] + \varepsilon$$

$$\zeta_{\mathrm{n}} \left\{ \mathbb{E}_{\psi}[\mathbf{n_L}] - \mathbb{E}_{\psi}[\mathbf{n_R}] - \varepsilon < 0 \right.$$

$$\operatorname{argmin} \underbrace{\sum_{i=1}^{N} \left(y_i - \psi_t\left(x_i\right)\right)^2}_{\text{loss function w.r.t data}} + \underbrace{\frac{\lambda}{2} \sum_{\mathbf{n} \in \mathcal{N}(\mathbf{x}_c)} \max\left(\zeta_{\mathbf{n}} \cdot |\zeta_{\mathbf{n}}|, 0\right)}_{\text{loss function w.r.t advice}}$$

KiGB

- Leaf update equation

$$\psi_t^{\ell}(\mathbf{x}) = \underbrace{\frac{1}{|\ell|}\sum_{i=1}^{N} \tilde{y}_i \cdot \mathbb{I}(x_i \in \ell)}_{\text{mean}} + \underbrace{\frac{\lambda}{2}\sum_{\mathbf{n}\in\mathcal{N}(\mathbf{x}_c)} \mathbb{I}(\zeta_{\mathbf{n}} > 0)\zeta_{\mathbf{n}} \cdot \left(\frac{\mathbb{I}(\ell \in \mathbf{n}_{\mathsf{R}})}{|\mathbf{n}_{\mathsf{R}}|} - \frac{\mathbb{I}(\ell \in \mathbf{n}_{\mathsf{L}})}{|\mathbf{n}_{\mathsf{L}}|}\right)}_{\text{penalty for advice violation}}$$

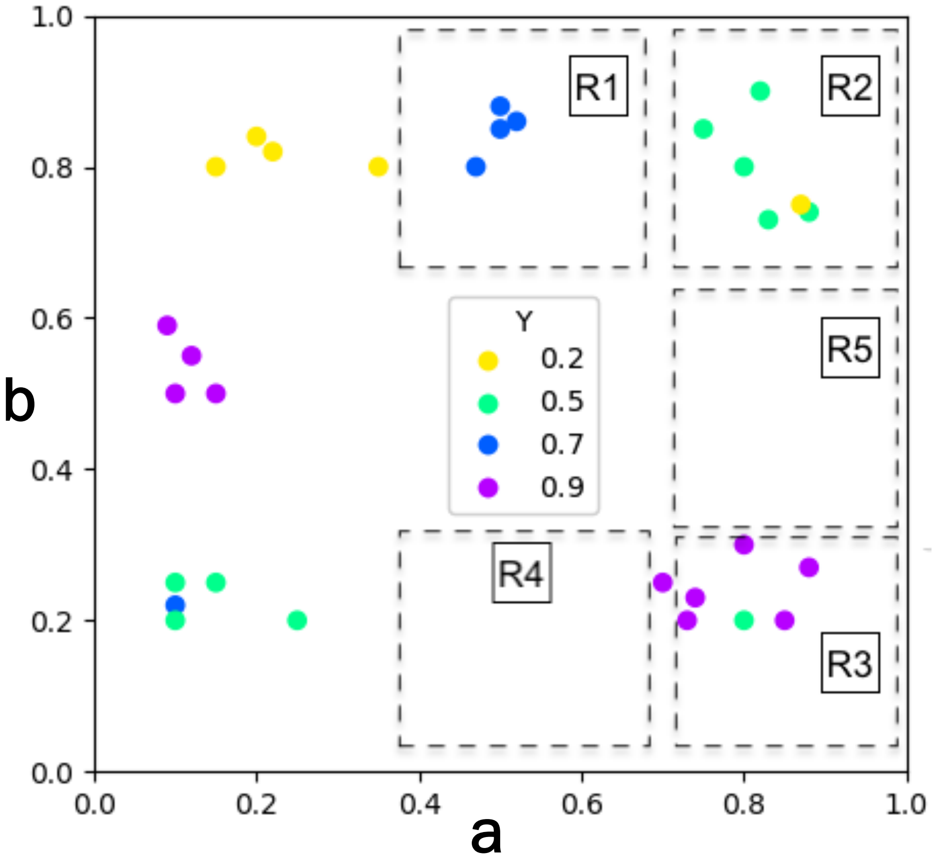
Monotonic Trees Ensemble

- Usually for classification tasks[3]
- Focus on global monotonicity
  - *prune*
  - *preprocessed data by reweighting*
  - *voting mechanism[4]*
  - *restrict split criteria[5]*
- Monoensemble[6] converts trees to rules and recalculate leaf values

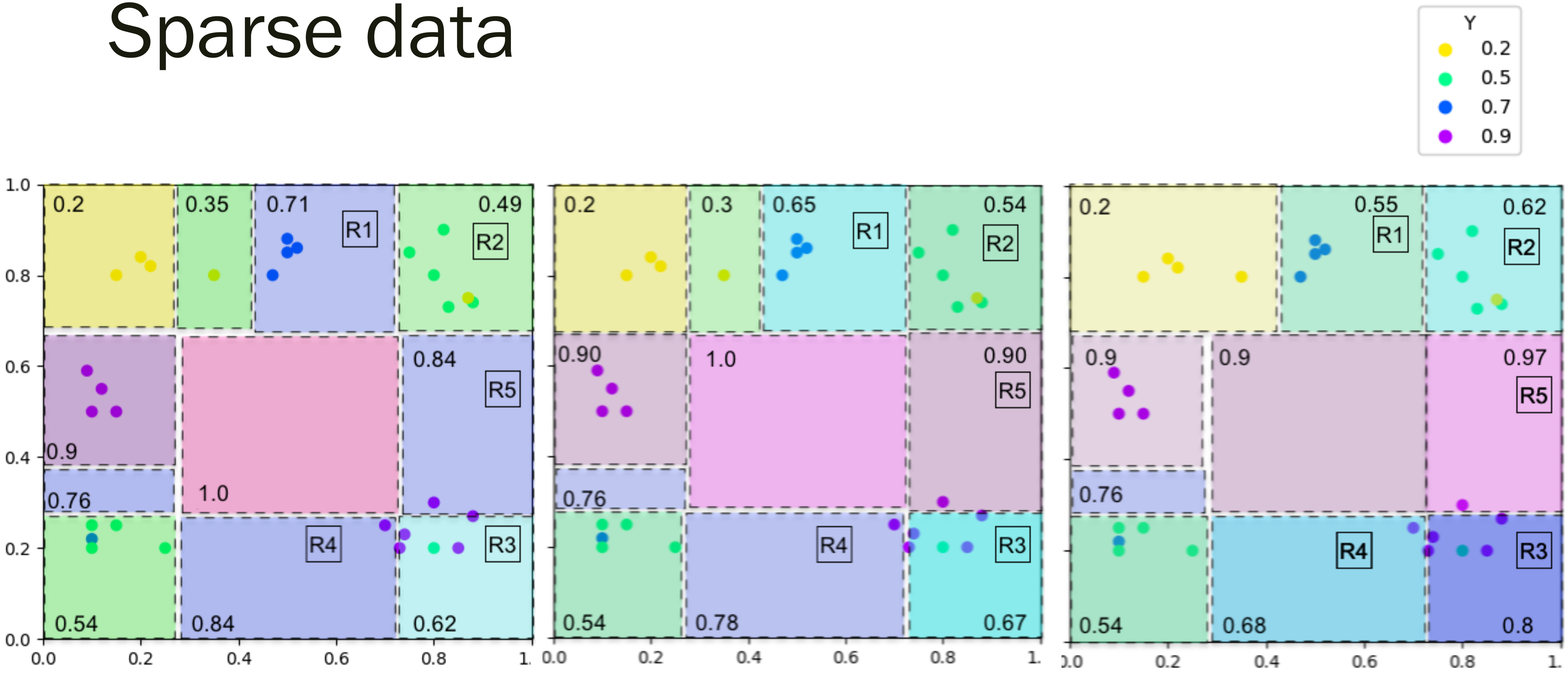
[3]Cano et al. Neurocomputing 2019, [4] Dembczynski et al. 2009
[5]Ke et al. NIPS 2017 (LightGBM), Chen et al. KDD 2016 (XGBoost), [6]Bartley et al. AAAI 2019

Sparse data

$$a^{Q+}_{\prec}\, y$$



Sparse data

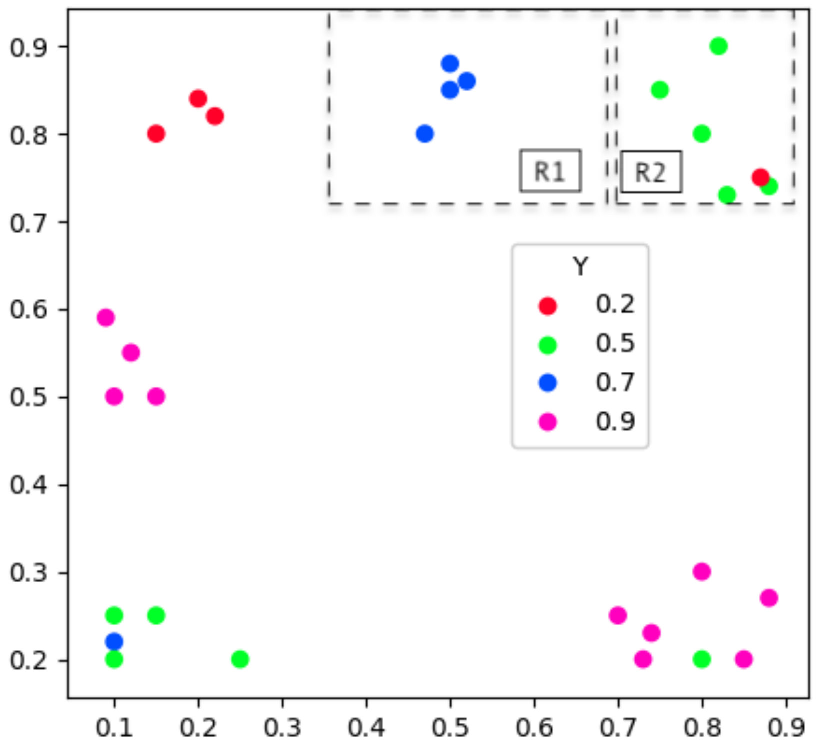


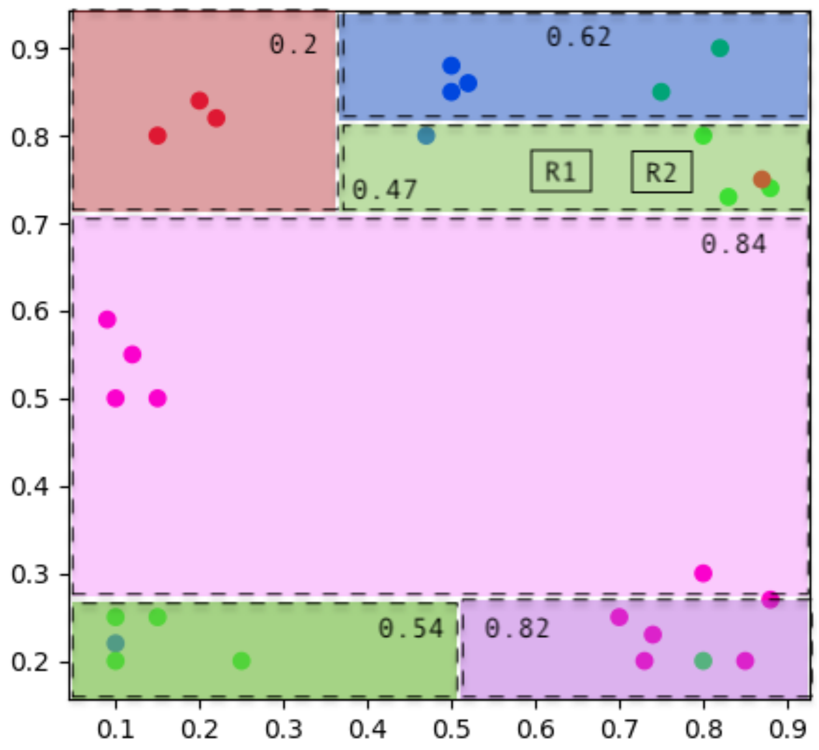
KiGB with $\lambda = 0$
standard boosting

KiGB with $\lambda = 3$

KiGB with $\lambda = 8$

Overfitting by monotonic function





Experiments

Standard baselines

(accuracy)

	Dataset	SKiGB	SGB	Dataset	SKiGB	SGB
classification	Adult	**0.855**	0.853	Cleveland	**0.737**	0.677
	Australian	**0.855**	0.83	Ljubljana	**0.696**	0.621
	Car	0.984	0.982			
regression	Abalone	**5.377**	5.491	CPU	**0.185**	0.204
	Autompg	**9.793**	13.623	Crime	2.211	2.296
	Autoprice	8.866	8.945	Redwine	**0.381**	0.419
	Boston	24.065	21.493	Whitewine	**0.426**	0.439
	California	47.159	47.468	Windsor	**3.9**	4.626

(mean-squared error)

KiGB: **ours with S/L**
SGB: Scikit-learn gradient boosting
LGBM: LightGBM
LMC: LightGBM with monotonic constraints
MONO: Monoensemble

Monotonic baselines

Dataset	SKiGB	MONO	LKiGB	LMC
Adult	0.855	0.857	**0.865**	0.863
Australian	0.855	**0.884**	**0.878**	0.867
Car	**0.984**	0.765	**0.972**	0.959
Cleveland	0.737	0.74	**0.757**	0.73
Ljubljana	**0.696**	0.611	0.721	0.718

classification task (accuracy)

Dataset	LKiGB	LMC	Dataset	LKiGB	LMC
Abalone	4.786	4.797	CPU	**0.206**	0.208
Autompg	**8.047**	8.33	Crime	1.834	1.847
Autoprice	**14.953**	15.614	Redwine	**0.382**	0.397
Boston	**15.496**	16.292	Whitewine	**0.45**	0.467
California	**48.517**	50.94	Windsor	**2.524**	2.634

regression task (mean-squared error)

Experiments

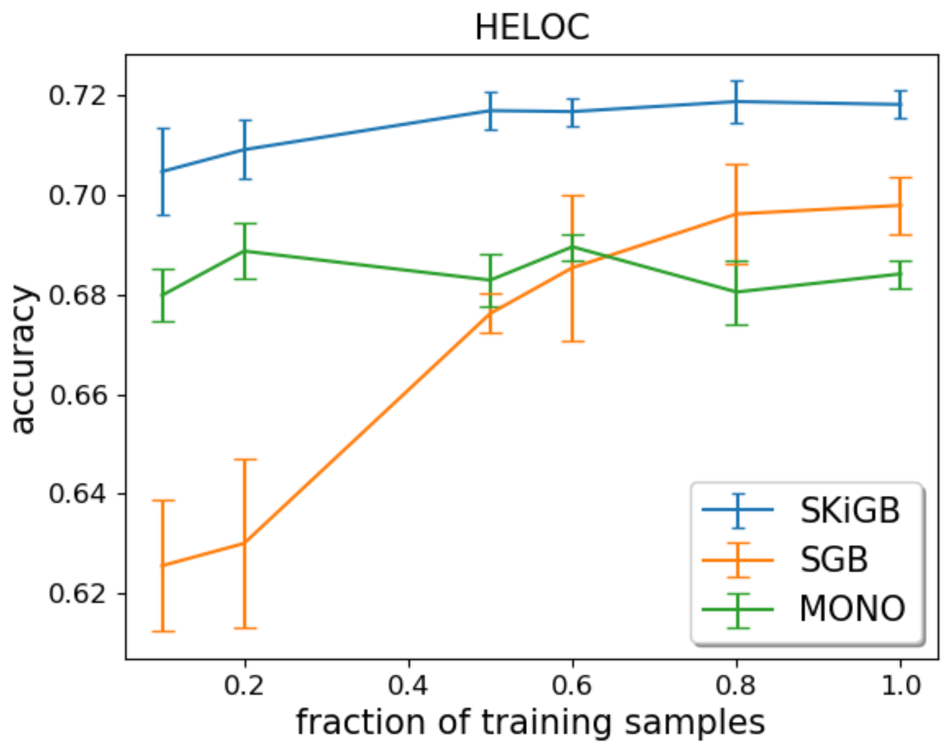
Real datasets

Dataset	LKiGB	LGBM	LMC
Logistics (mse)	**1.851**	1.898	1.889

Dataset	SKiGB	SGB	MONO
HELOC (accuracy)	**0.717**	0.7	0.688

Logistics : Turvo
HELOC : FICO xML challenge

Learning curve



THANKS