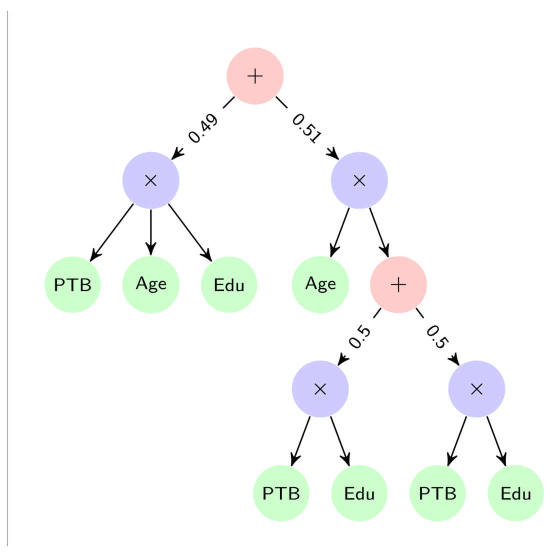
A Unified Framework for Human-Allied Learning of Probabilistic Circuits

Athresh Karanam*, Saurabh Mathur*, Sahil Sidheekh*, Sriraam Natarajan

What?

Background

Probabilistic circuits (PCs) represent joint probability distributions using structured computational graphs; they can efficiently answer probability queries.



"Risk of preterm birth increases with age"

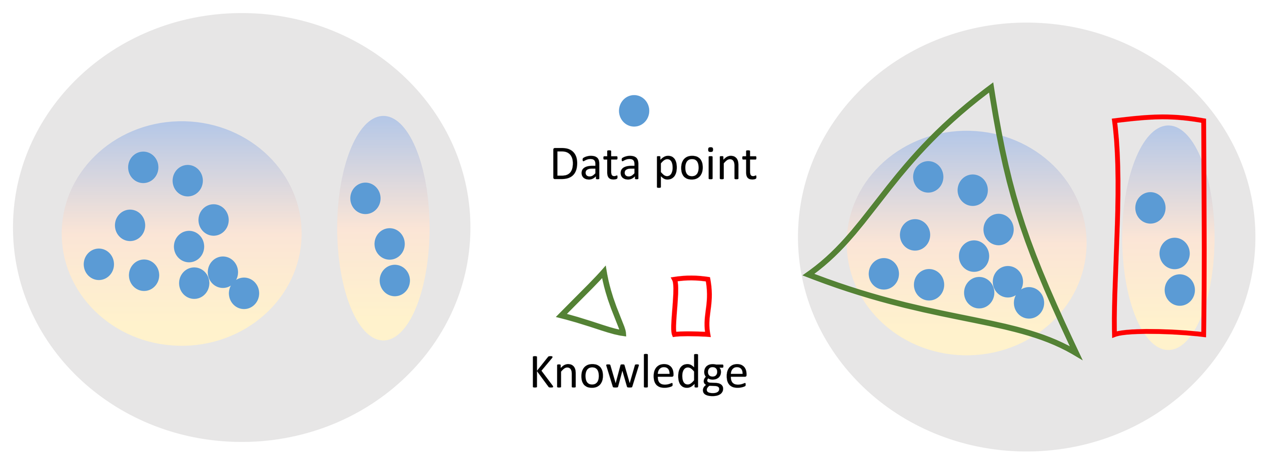
"Risk of preterm birth is conditionally independent of education given low age."

Domain knowledge concisely encodes information about general trends but is insufficient to fully define distributions. e.g., monotonicity, independence.

Motivation

Learning distributions using **Sparse** and **Noisy** data

Exploit Domain Knowledge as inductive bias



Need a principled way to integrate different types knowledge into the learning of PCs

How?

Methodology

Given: Dataset D over variables $\mathbf{X}$ and multiple forms of domain knowledge $\boldsymbol{K}$
To Do: Learn a probabilistic circuit θ that accurately models $P(\boldsymbol{X})$

A unified framework for encoding knowledge: Differentiable functions of probability queries

Equality constraints

$$P(\mathbf{x}) = P(\mathbf{x}'),\ \forall(\mathbf{x},\mathbf{x}') \in \mathcal{D}^2 \text{ s.t. similar}(\mathbf{x},\mathbf{x}')$$

"Similar data points are equally likely"

Inequality constraints

$$P(X_i = 1 \mid X_j = 1) > P(X_i = 1 \mid X_j = 0)$$

"X_i increases with increase in X_j"

Overall Pipeline

Formally state the domain knowledge → Encode the knowledge as equality & inequality constraints on probability queries → Choose a differentiable measure of knowledge violation, e.g., MSE. → Learn PC by solving a series of optimization problems

Solve the following sequence of optimization problems, increasing penalty weight until violation term vanishes

$$\theta_{t+1} = \arg\max_{\theta} \underbrace{\mathcal{L}(\langle\mathcal{G},\theta\rangle,\mathcal{D})}_{\text{Data}} - \lambda_t \underbrace{\zeta(\langle\mathcal{G},\theta\rangle)}_{\text{Knowledge}}$$

- Penalty acts as knowledge-intensive regularization
- Can be computed efficiently & differentiably from PC
- Penalty from multiple forms of knowledge can be added

What we get?

Empirical evaluation

		PC	*PC+Knowledge*
BN	asia	-483.3 ± 4.1	$\mathbf{-313.2 \pm 3.9}$
	sachs	-1097.5 ± 8.8	$\mathbf{-861.2 \pm 8.7}$
	survey	-611.7 ± 7.2	$\mathbf{-476.6 \pm 6.6}$
	earthquake	-272.0 ± 2.4	$\mathbf{-121.8 \pm 2.1}$
UCI	breast-cancer	-2110.8 ± 15.6	$\mathbf{-1271.5 \pm 14.6}$
	diabetes	-7010.3 ± 31.0	$\mathbf{-5070.3 \pm 481.8}$
	thyroid	-351.5 ± 6.1	$\mathbf{-200.5 \pm 23.2}$
	heart-disease	-931.7 ± 15.0	$\mathbf{-739.8 \pm 7.2}$
RW	numom2b-a	-14573.9 ± 69.9	$\mathbf{-7288.2 \pm 1.6}$

	PC	+CSI	+CSI+MIS
earthquake	-272.0 ± 2.4	-137.7 ± 4.7	$\mathbf{-106.1 \pm 1.1}$
survey	-611.7 ± 7.2	-523.5 ± 4.3	$\mathbf{-470.9 \pm 6.6}$
asia	-483.3 ± 4.1	-320.5 ± 9.9	$\mathbf{-284.7 \pm 6.4}$
numom2b-b	-18281.2 ± 218.8	-15122.9 ± 201.7	$\mathbf{-14758.1 \pm 60.3}$

- PCs learned by combining domain knowledge with data outperform purely data-driven ones.
- PCs learned using multiple forms of knowledge outperform those limited to one form of knowledge.

Future Work

- Extending to structured, muti-relational domains
- Learning & refining structure of PCs
- Actively eliciting domain knowledge

Acknowledgements

The authors acknowledge the support by the NIH grant R01HD101246, the AFOSR award FA9550-23-1-0239, the ARO award W911NF2010224, and the DARPA ANSR award HR001122S0039.

https://starling.utdallas.edu/
@STARLing_lab

*Equal contributors

Sidheekh and Natarajan 2024; Choi et al.2020; Poon and Domingos 2013; Odom et al. 2015; Pasunuri et al. 2016; Altendorf et al. 2005; Yang and Natarajan 2013; Fung et al. 2002; Mathur et al. 2023