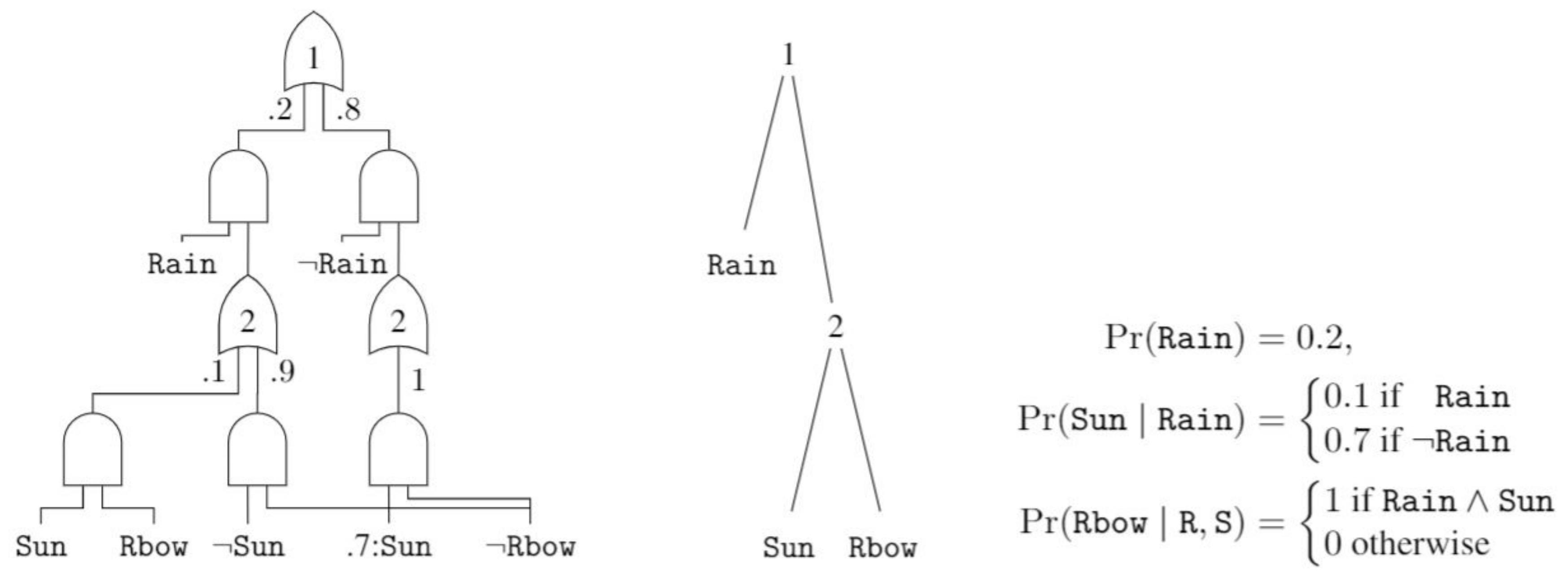
IL-Strudel : Independence Based Learning of Structured-Decomposable Probabilistic Circuit Ensembles

Shreyas Kowshik, Yitao Liang, Guy Van den Broeck
{shreyaskowshik@iitkgp.ac.in}, {yliang, guyvdb}@cs.ucla.edu

INTRODUCTION



(a) PSDD (b) Vtree (c) Equivalent distribution

PSDD : Circuit representation of a Probability Distribution over boolean random variables

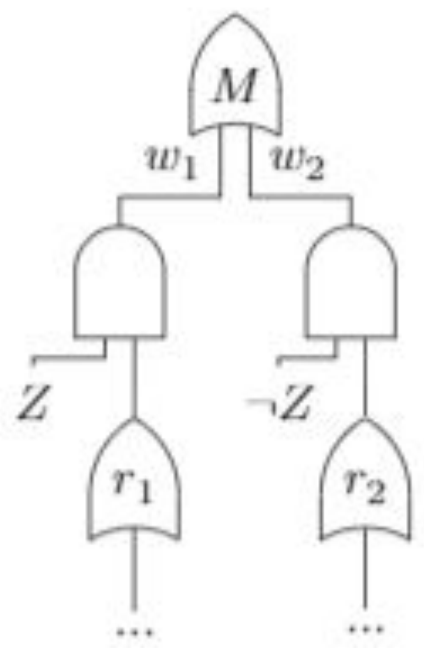
Structured Decomposability : Encoded distribution factorizes according to a given vtree

Context-Specific-Indepenences : PSDDs capture structure in data by inducing different types of CSIs

Proposition 1. (Prime-Sub Independence) *For a PSDD rooted at* r, *for every decision node* n *with prime and sub variables* $\mathbf{X}$ *and* $\mathbf{Y}$ *and their every element* $(p, s)^2$:

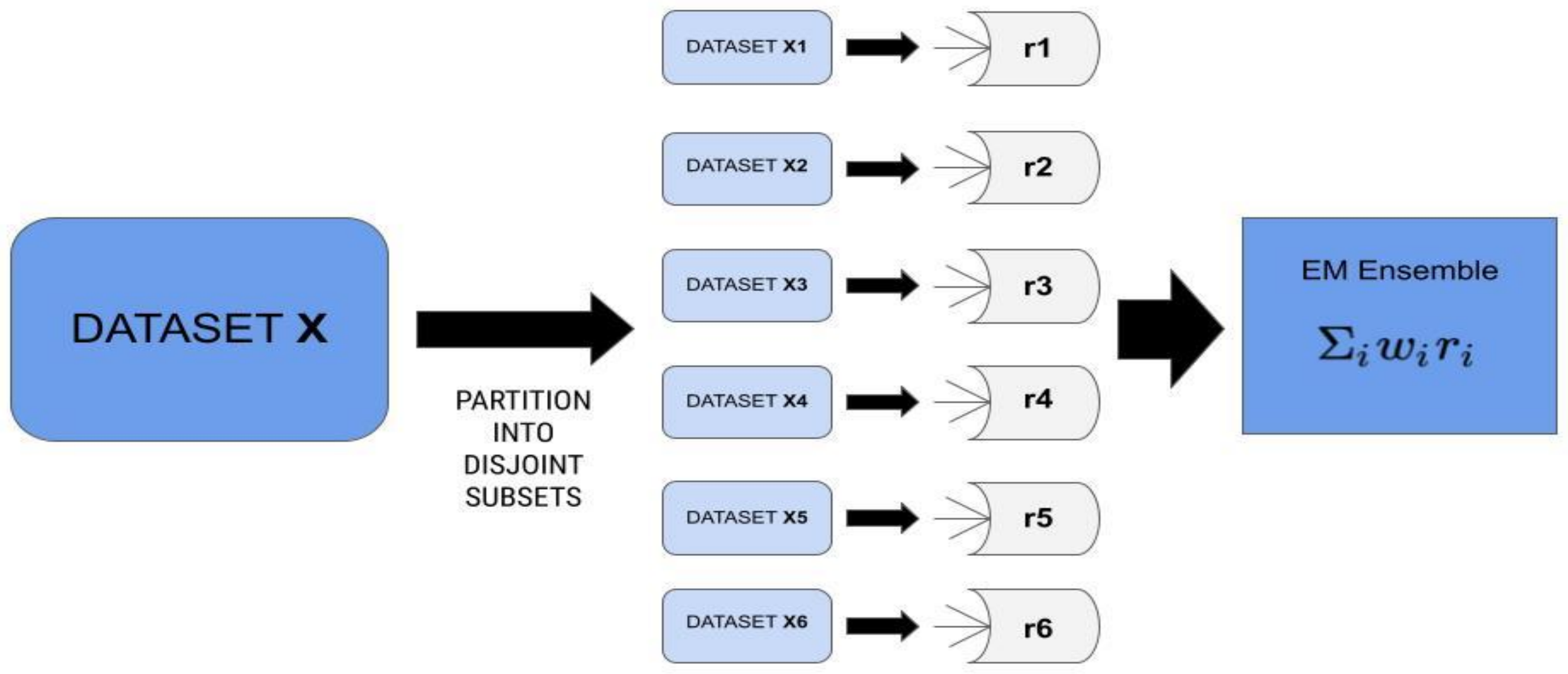
$$
\begin{aligned}
& Pr_r(\mathbf{xy}|[p] \wedge [s] \wedge \gamma_n) \\
= \; & Pr_n(\mathbf{xy}|[p] \wedge [s]) = Pr_n(\mathbf{x}|[p] \wedge [s]) \cdot Pr_n(\mathbf{y}|[p] \wedge [s]) \\
= \; & Pr_r(\mathbf{x}|[p] \wedge [s] \wedge \gamma_n) \cdot Pr_r(\mathbf{y}|[p] \wedge [s] \wedge \gamma_n)
\end{aligned}
$$

ENSEMBLE LEARNING



- **Ensemble of PSDDs :** Represented as a single PSDD with latent variable at root
- For fixed component structures, parameter learning done using **Soft-EM Algorithm**
- Highly sensitive to **Initialization**

WHAT IS A STRONG INITIALIZATION STRATEGY FOR EM ENSEMBLES?



Approach : Partition dataset into disjoint sets. Train a component structure on each subset. Use these as initialization to EM ensemble.

How to partition? : Find maximal subsets such that prime-sub-independence satisfied approximately at component root level (measured by pairwise-mutual-information)

Optimization Problem : Finding maximal partitions can be cast as a discrete optimization problem.

$$
\underset{\mathcal{B} \in [0,1]^N}{\text{Max}} \; \mathbb{1}(\mathcal{B})
$$

$$
\text{subject to} : \mathcal{F}(\mathbf{X}[\mathcal{B}]) \leq \mathcal{T}
$$

RESULTS

Datasets	IL-Strudel	Strudel-EM	EM-LearnPSDD
NLTCS	**-6.03**	-6.07	-6.03
MSNBC	**-6.04**	-6.04	-6.04
KDD	**-2.12**	-2.14	-2.12
Plants	-13.30	**-13.22**	-13.79
Audio	**-40.22**	-41.2	-41.98
Jester	**-52.95**	-54.24	-53.47
Netflix	**-56.99**	-57.93	-58.41
Accidents	-29.86	**-29.05**	-33.64
Retail	-10.84	-10.83	**-10.81**
Pumsb-Star	-25.55	**-24.39**	-33.67
DNA	**-86.93**	-87.15	-92.67
Kosarek	**-10.61**	-10.7	-10.81
MSWeb	-9.78	**-9.74**	-9.97
Book	**-34.12**	-34.49	-34.97
EachMovie	**-51.92**	-53.72	-58.01
WebKB	**-152.79**	-154.83	-161.09
Reuters-52	**-85.60**	-86.35	-89.61
20NewsGrp.	**-152.24**	-153.87	-161.09
BBC	-253.46	-256.53	**-253.19**
AD	**-15.23**	-16.52	-31.78

- IL-Strudel beats Strudel-EM on 14/20 datasets
- Better results on the larger datasets
- More the number of variables, better the potential to capture rich CSIs.

CONCLUSION

Summary : Initial component structures trained on disjoint sub-supports with prime-sub independence holding at component root levels is a better EM initialization strategy

Future work : Explore incorporation of vtree in ensemble learning setup. Use of more robust independence measures than pairwise-mutual-information.