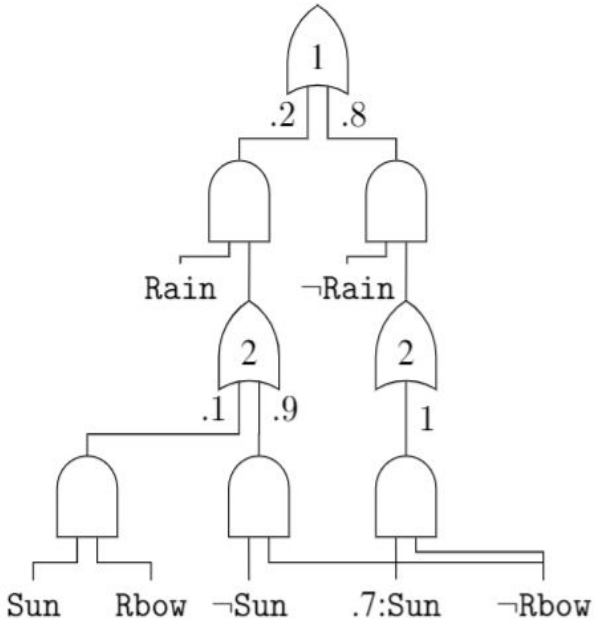
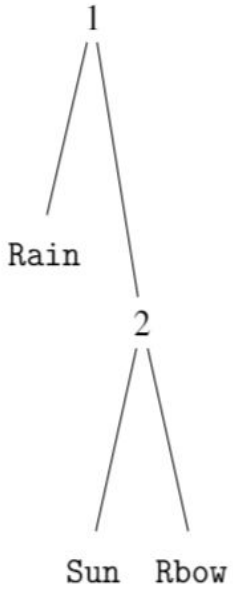
IL-Strudel : Independence Based Learning of Structured-Decomposable Probabilistic Circuit Ensembles

Shreyas Kowshik, Yitao Liang, Guy Van den Broeck

Probabilistic Sentential Decision Diagrams



(a) PSDD



(b) Vtree

$$\Pr(\texttt{Rain}) = 0.2,$$

$$\Pr(\texttt{Sun} \mid \texttt{Rain}) = \begin{cases} 0.1 \text{ if } \texttt{Rain} \\ 0.7 \text{ if } \neg\texttt{Rain} \end{cases}$$

$$\Pr(\texttt{Rbow} \mid \texttt{R}, \texttt{S}) = \begin{cases} 1 \text{ if } \texttt{Rain} \wedge \texttt{Sun} \\ 0 \text{ otherwise} \end{cases}$$

(c) Equivalent distribution

Probabilistic Sentential Decision Diagrams

Context-Specific-Independences (CSI)

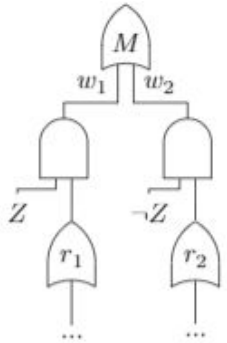
$$\Pr(\mathbf{X}, \mathbf{Y} \mid \gamma) = \Pr(\mathbf{X} \mid \gamma) \Pr(\mathbf{Y} \mid \gamma)$$

Theorem

A PSDD is optimal (in log-likelihood terms) when context-specific-independences between left-right variables of PSDD nodes hold empirically

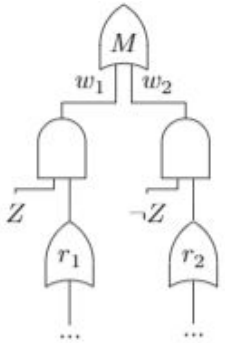
CSI Based Perspective for Structure Evaluation

Ensemble Learning



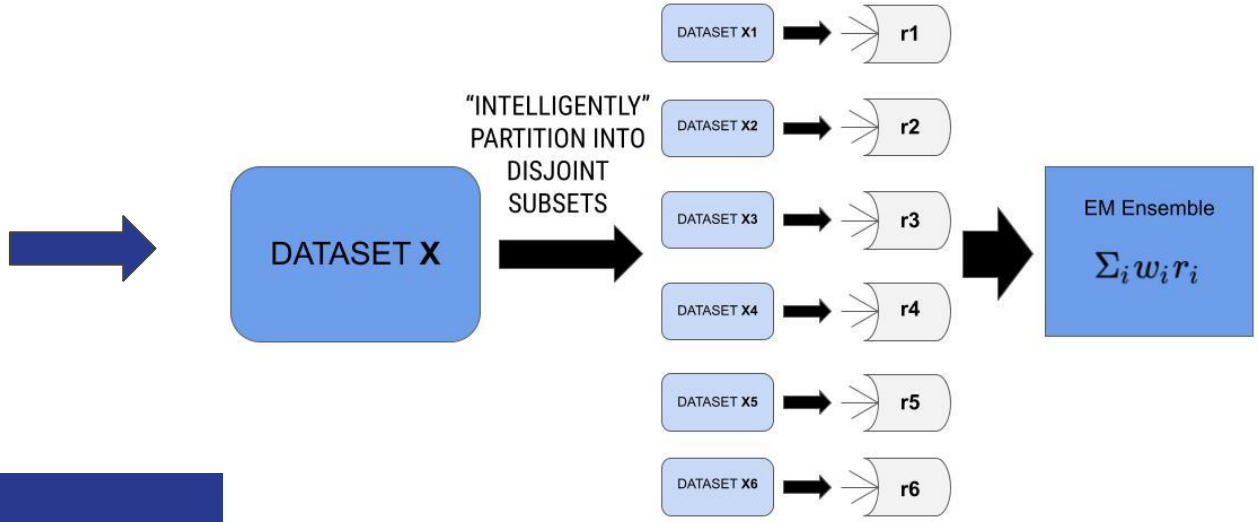
EM Algorithm
Sensitive to Initialization

Ensemble Learning



EM Algorithm
Sensitive to Initialization

What is a good initialization strategy for EM based ensembles?



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

Better on **14/20 datasets**

Significant improvement on the **larger datasets**

Thank You!