Learning The Structure of Probabilistic Sentential Decision Diagrams

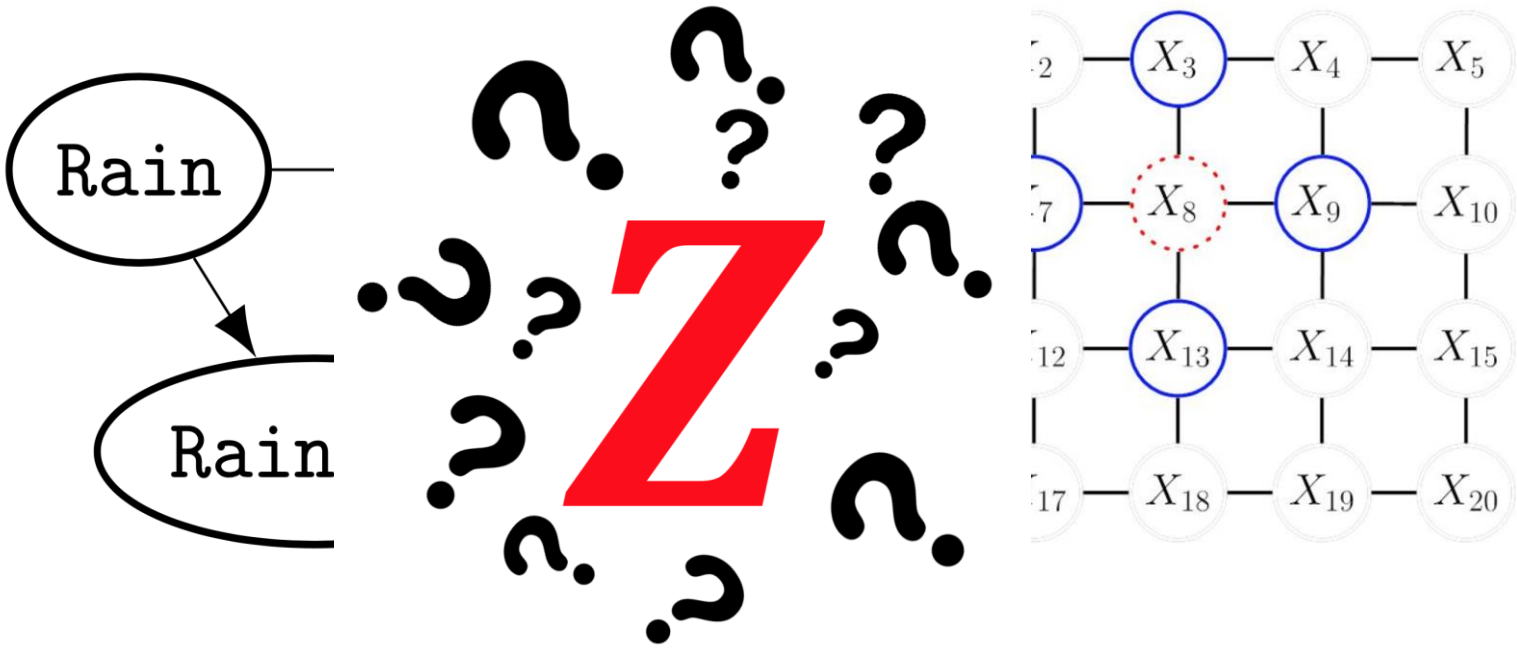
Yitao Liang, Jessa Bekker, Guy Van den Broeck

August 12, 2017

Background: Intractable Representation

Bayesian networks

Markov networks

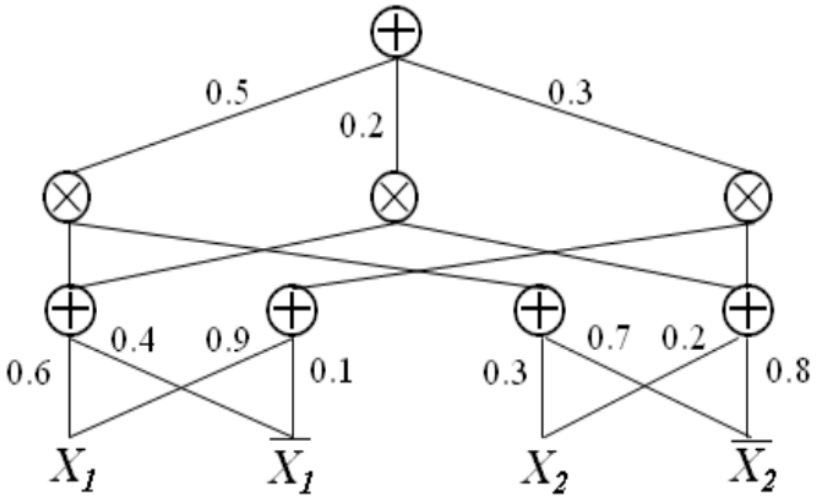


Do not support linear-time exact inference

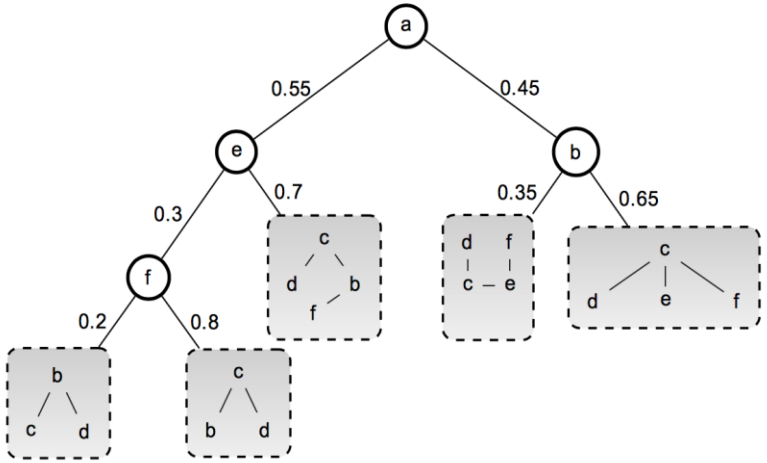
Background: Tractable Representation

Historically: Polytrees, Chow-Liu trees, etc.

SPNs

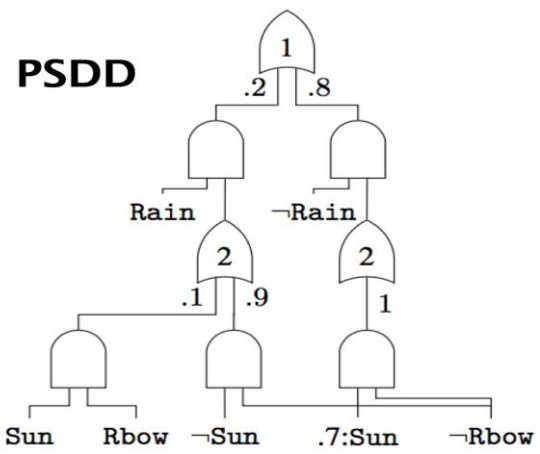


Cutset Networks

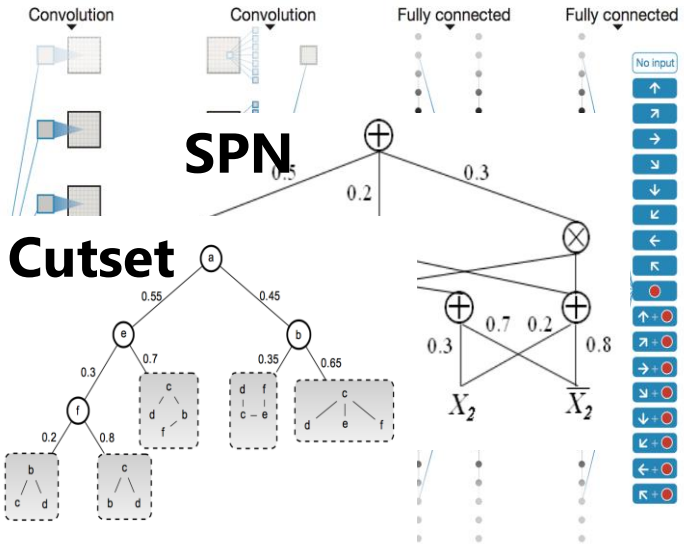


Both are Arithmetic Circuits (ACs)

Probabilistic Sentential Decision Diagrams



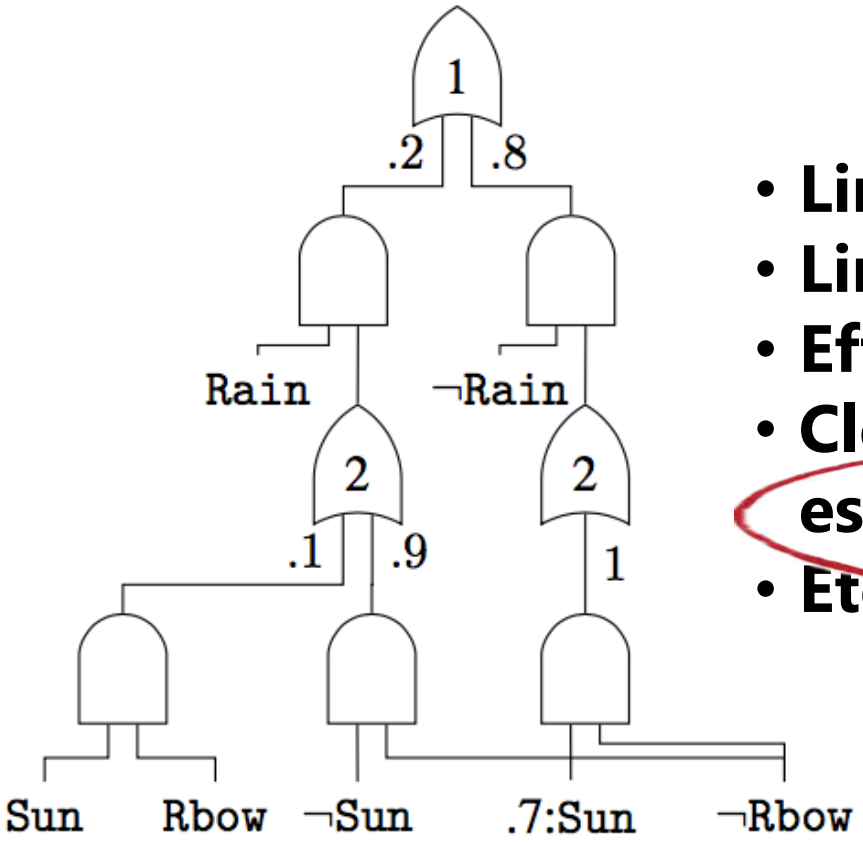
DNN



Strong Properties ⟷ **Representational Freedom**

Perhaps the most powerful circuit proposed to date

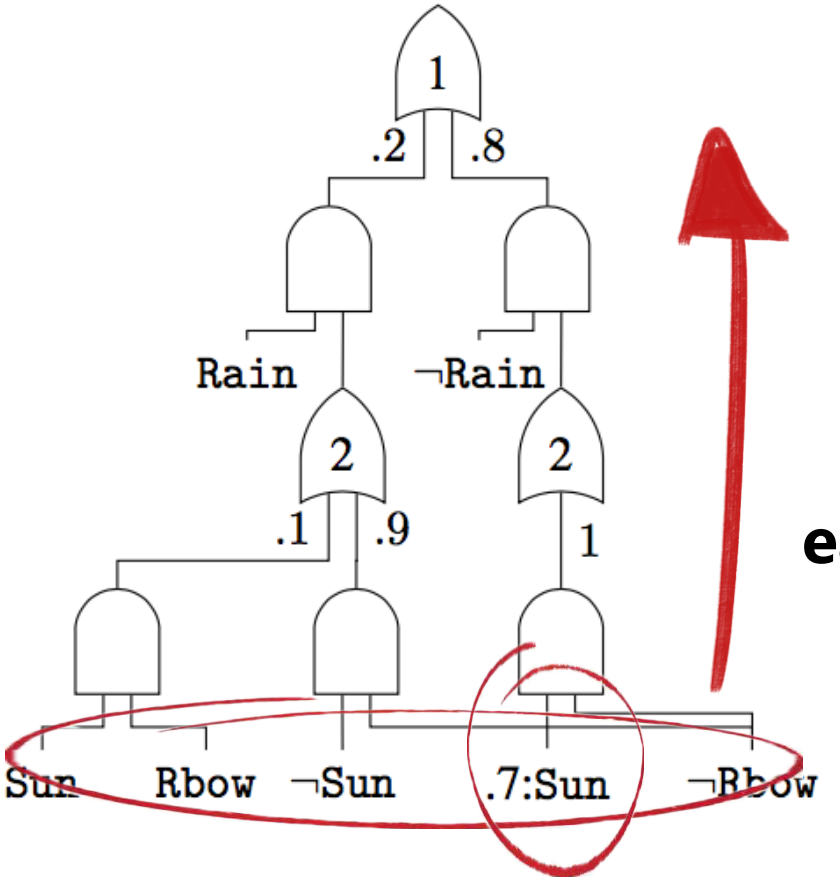
Probabilistic Sentential Decision Diagrams



- **Linear MPE inference**
- **Linear conditional marginals**
- **Efficient multiplication**
- **Closed-form parameter estimation**
- **Etc.**

Structure learning

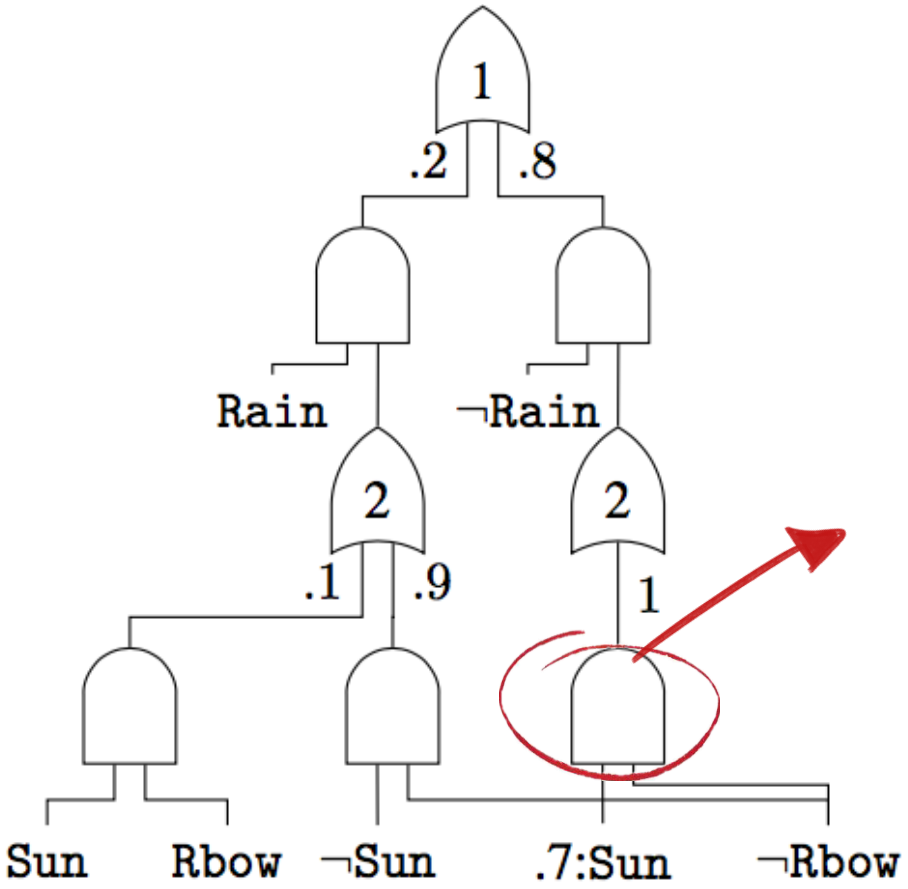
What is a PSDD



Bottom-up

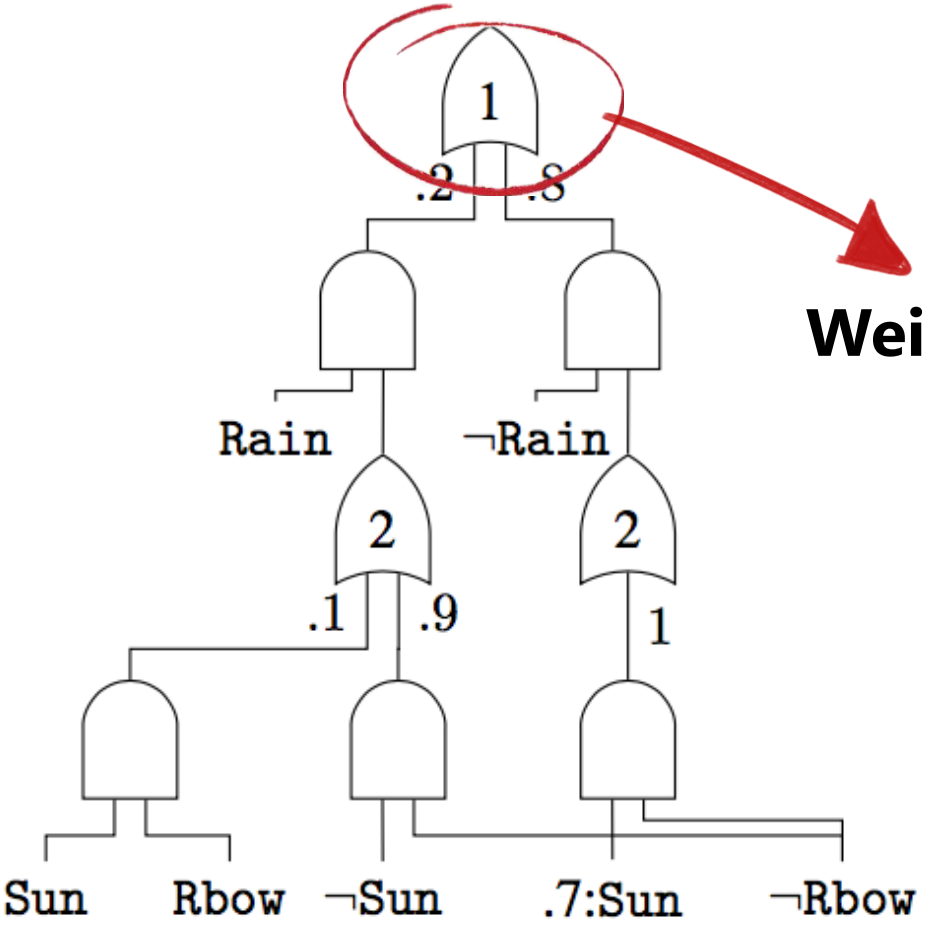
each node is a distribution

What is a PSDD



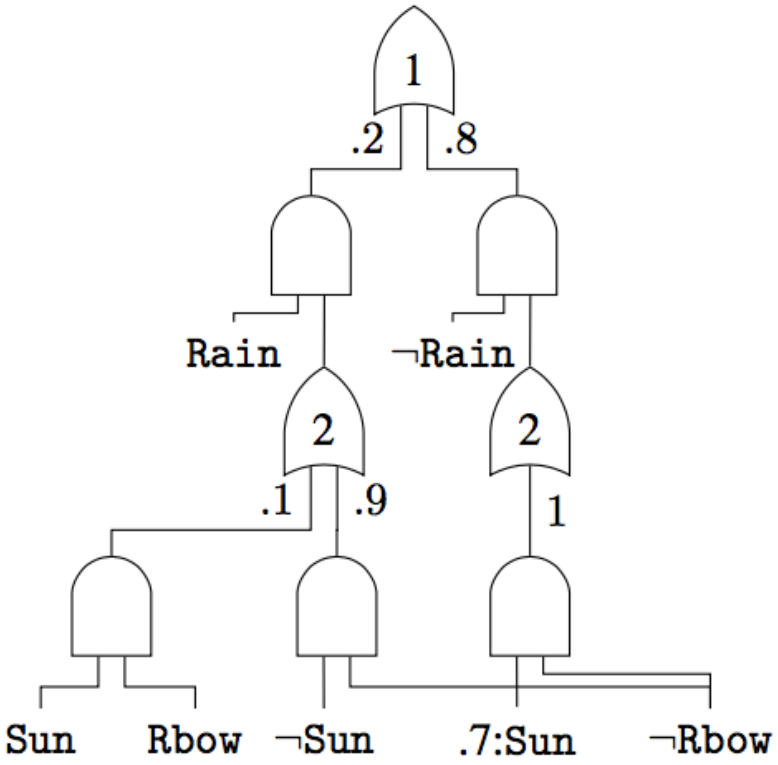
Multiply independent distributions

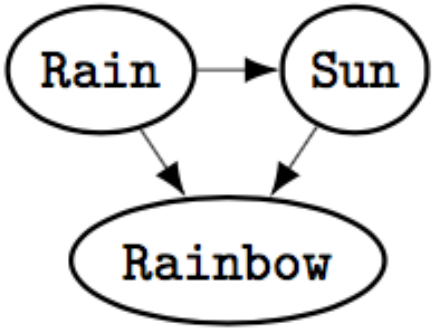
What Is a PSDD



Weighted mixture of lower level distributions

What Does a PSDD Represent





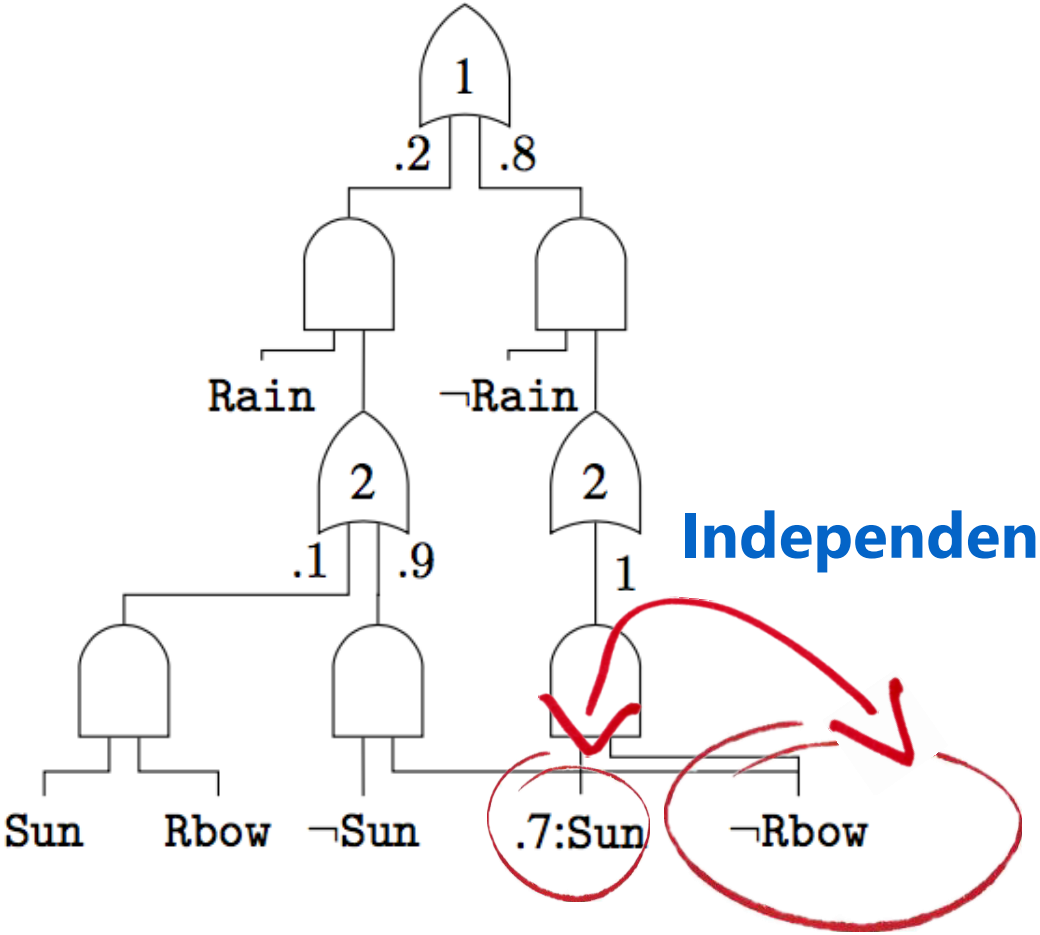
$$\Pr(\texttt{Rain}) = 0.2,$$

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

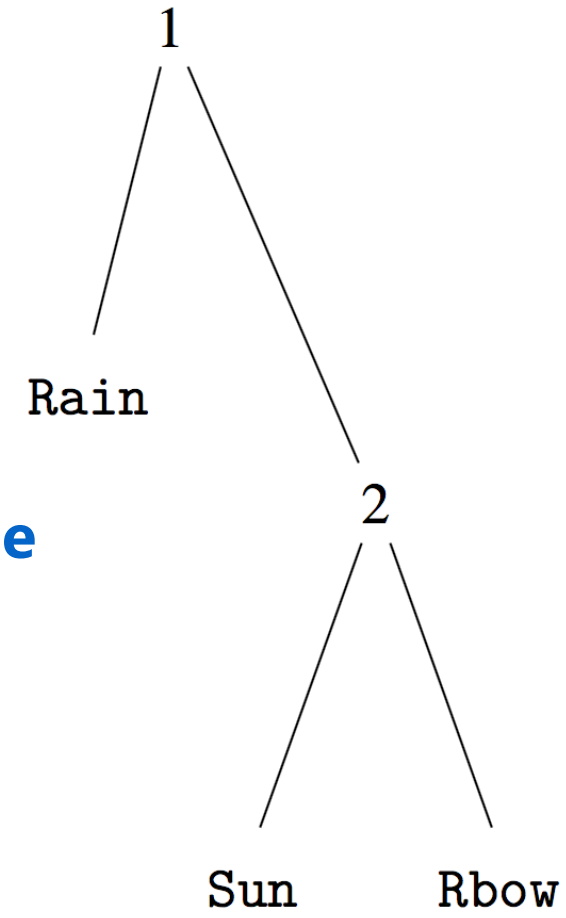
$$\Pr(\texttt{Rbow} \mid \texttt{R}, \texttt{S}) = \begin{cases} 1 \text{ if } \texttt{Rain} \wedge \texttt{Sun} \\ 0 \text{ otherwise} \end{cases}$$

Are PSDDs amenable to tractable structure learning

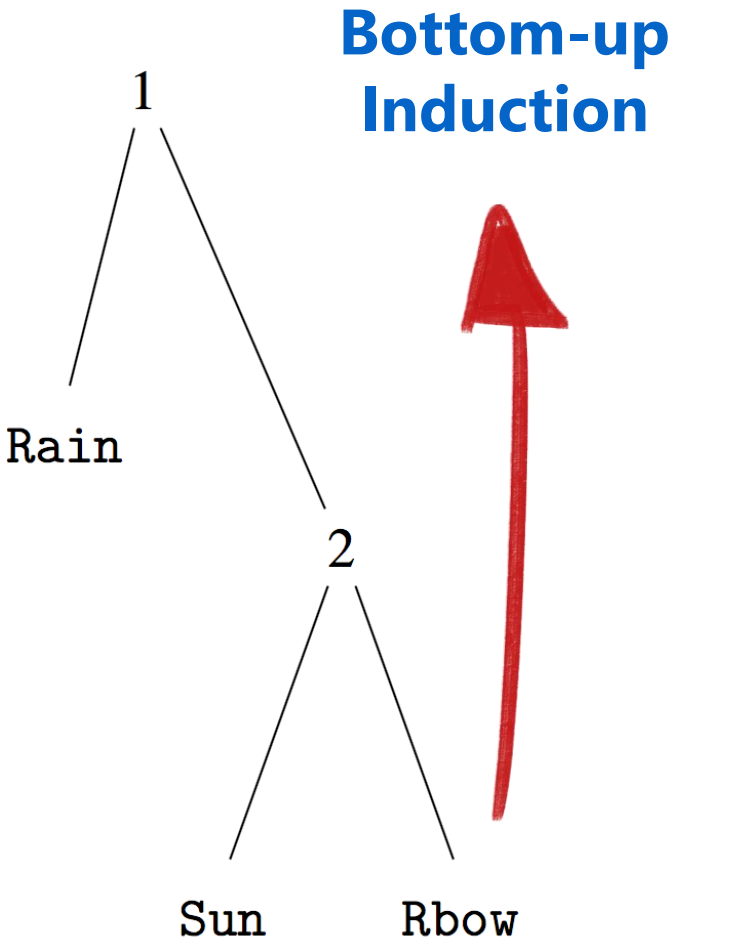
Independence & Variable Tree

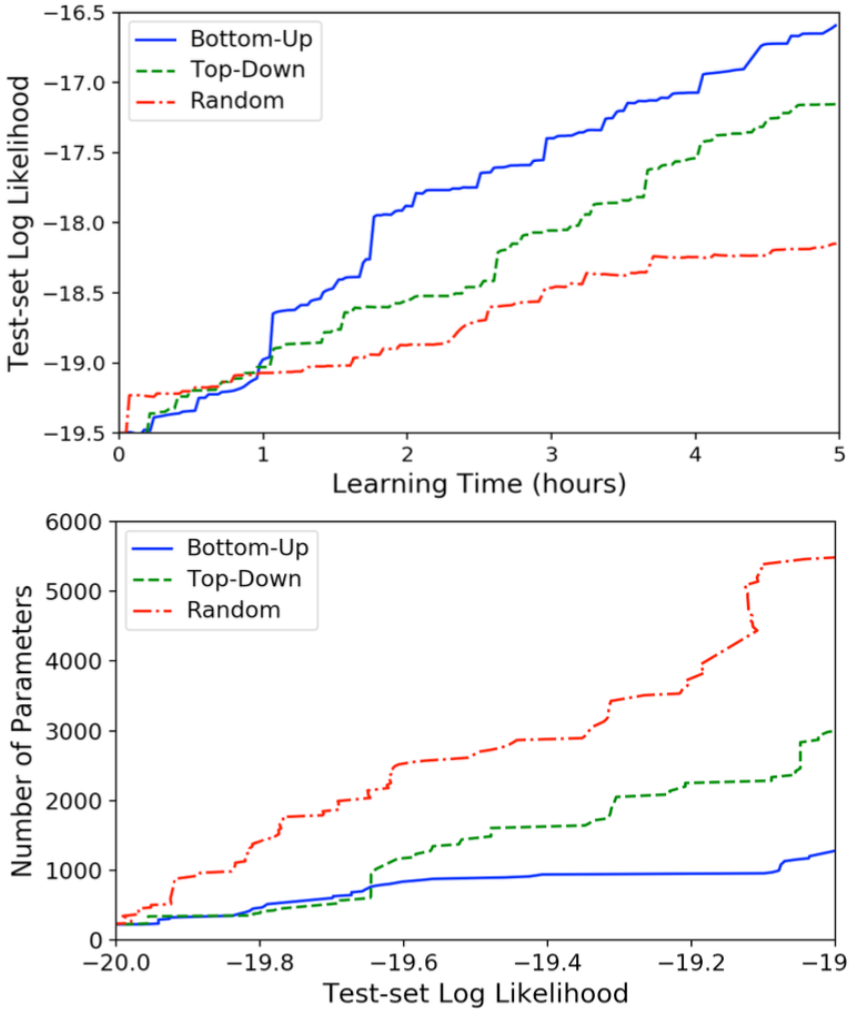


Independence

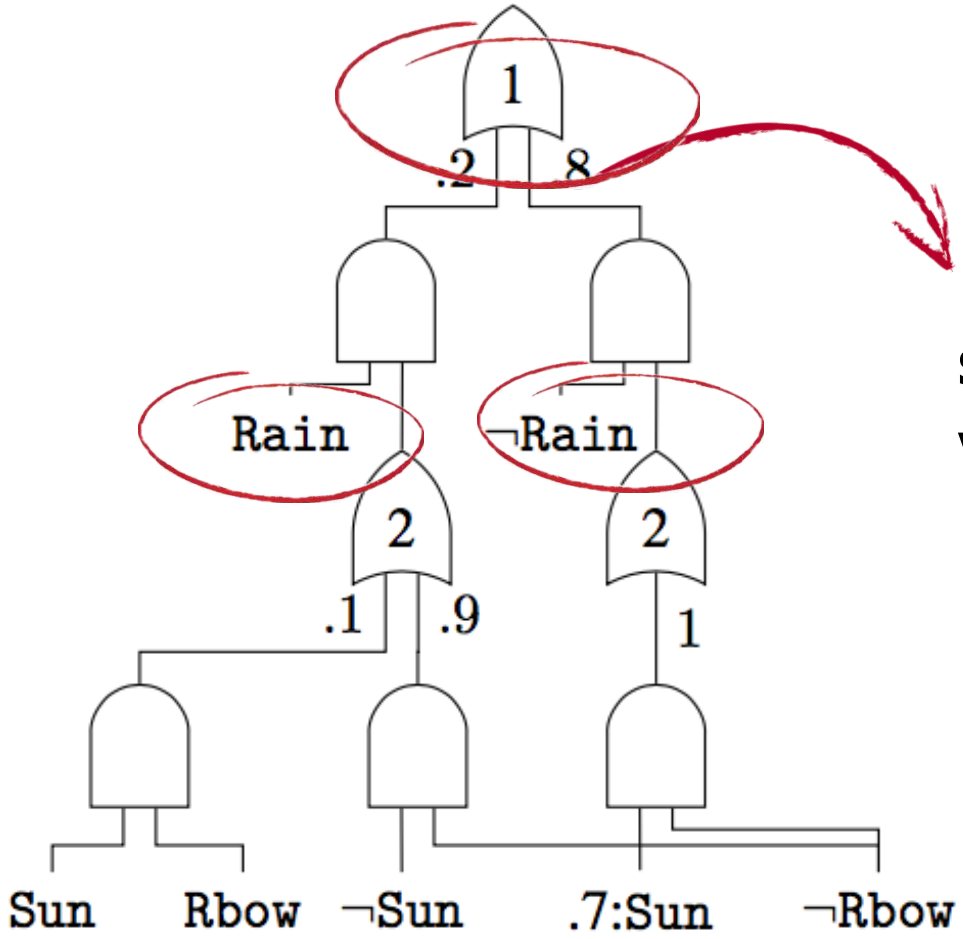


Induce a Vtree from Data



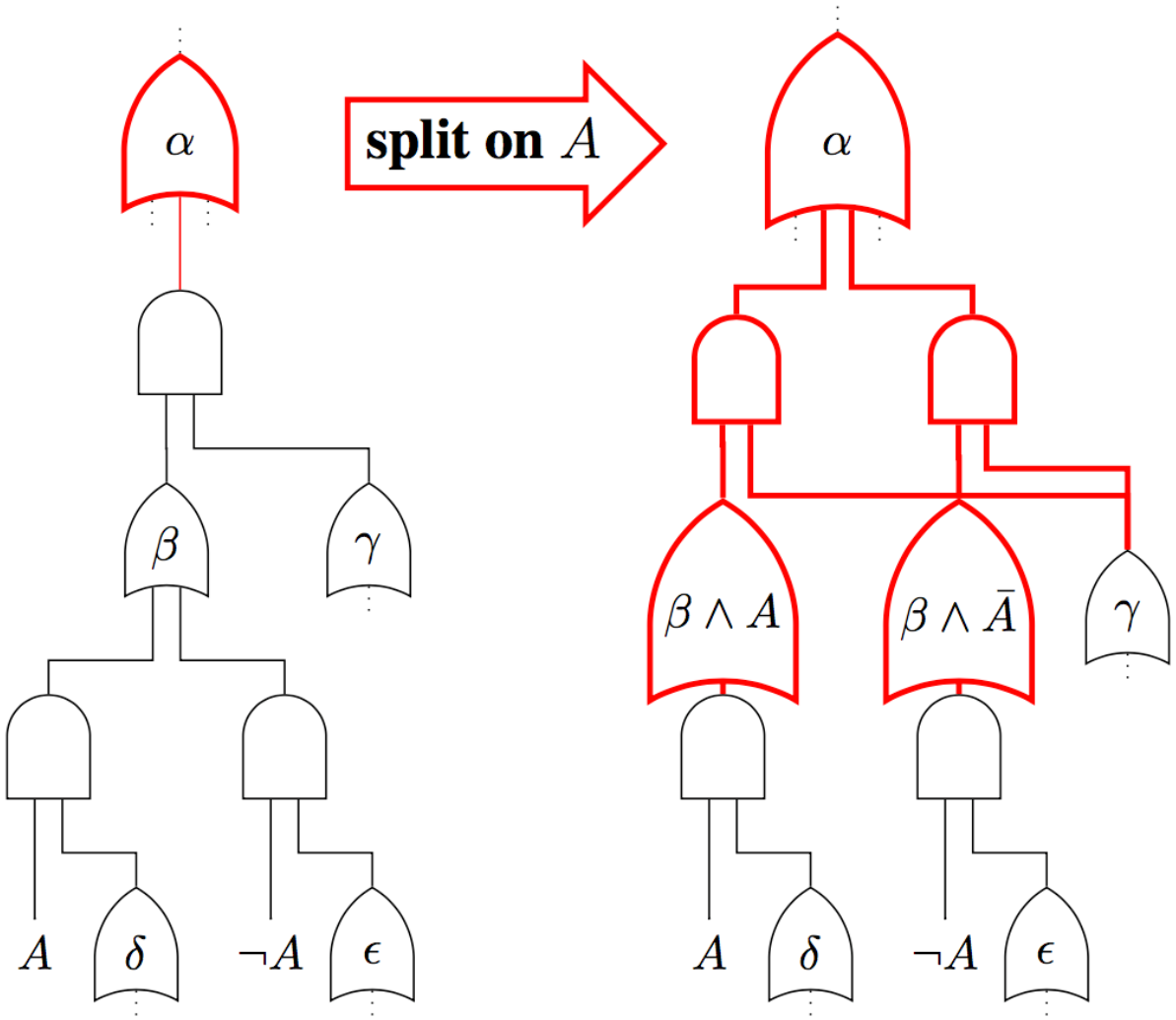


PSDD: Determinism

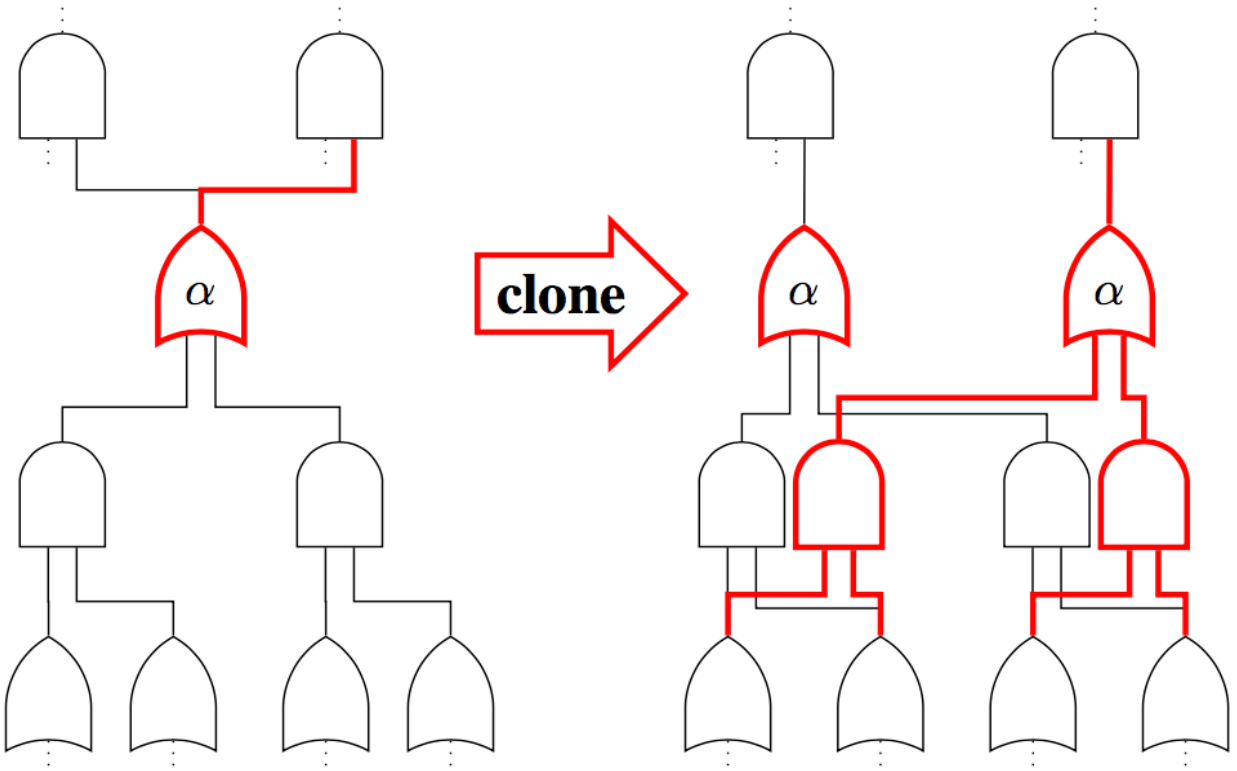


Branch over sentences on left variables

Search for Structure: LearnPSDD Operations



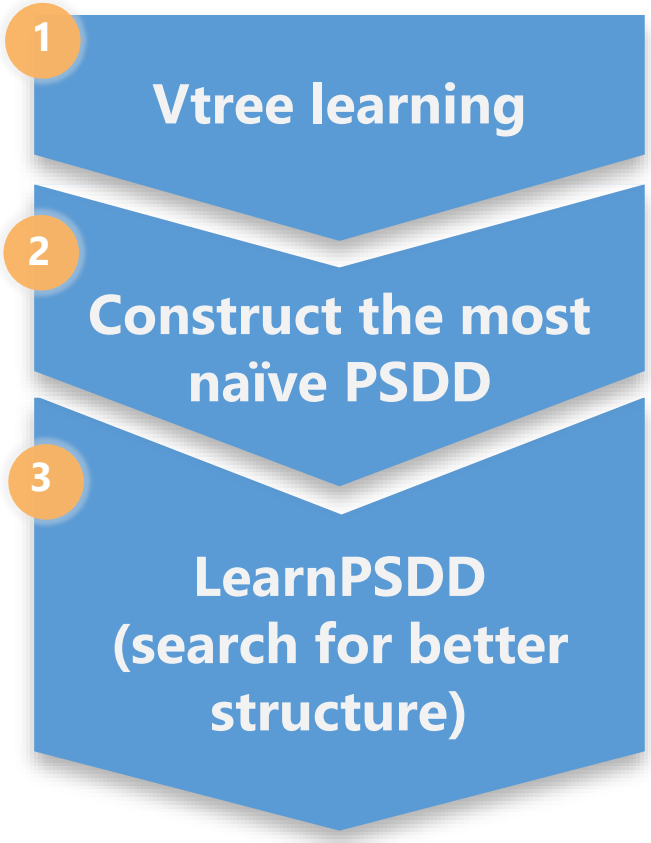
Search for Structure: LearnPSDD Operations



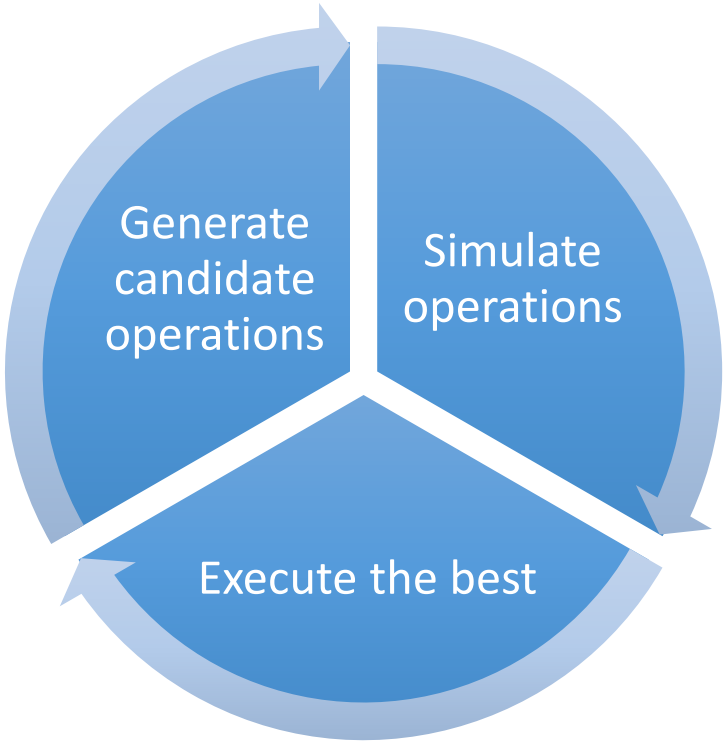
$$\text{score} = \frac{\ln \mathcal{L}(r' \mid \mathcal{D}) - \ln \mathcal{L}(r \mid \mathcal{D})}{\text{size}(r') - \text{size}(r)}$$

Learn a PSDD from Data

Roadmap



LearnPSDD

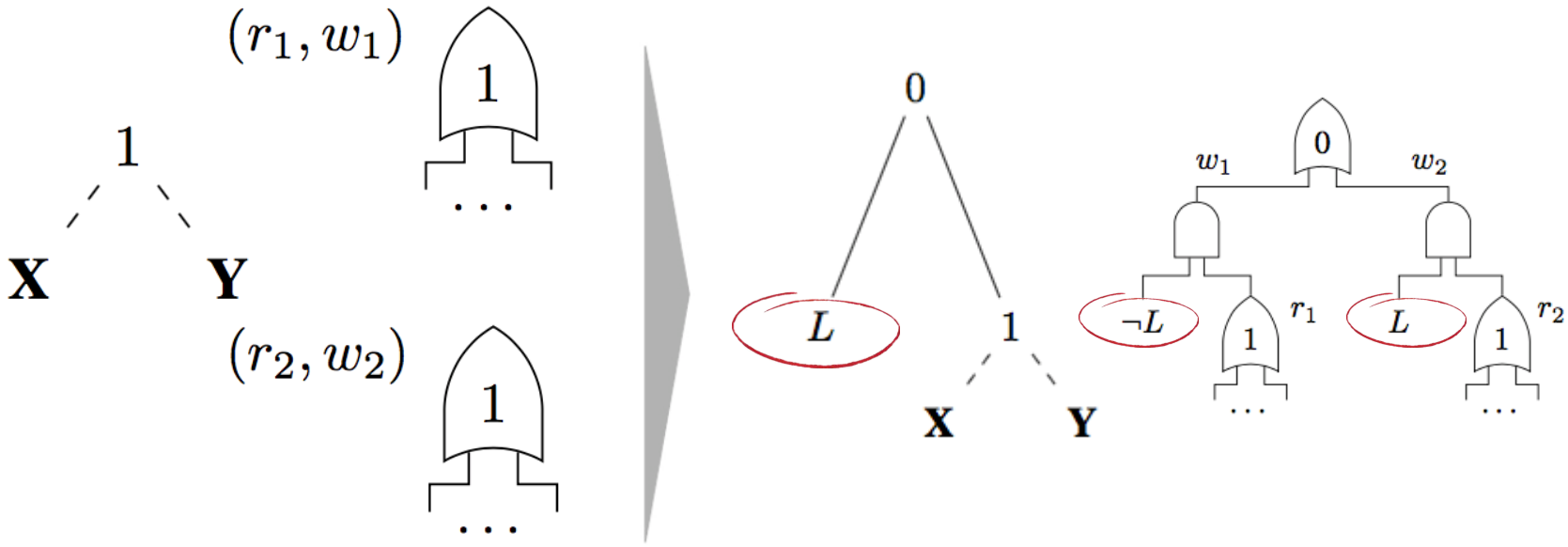


Experiments

Datasets	|Var|	|Train|	|Valid|	|Test|	LearnPSDD		EM-LearnPSDD		SearchSPN	Merged L-SPN		Merged O-SPN	
					LL	Size	LL	Size	LL	LL	Size	LL	Size
NLTCS	16	16181	2157	3236	$-6.03^{\dagger *}$	3170	$\mathbf{-6.03}^{*}$	2147	−6.07	−6.04	3988	−6.05	1152
MSNBC	17	291326	38843	58265	$-6.05^{\dagger}$	8977	$\mathbf{-6.04}^{*}$	3891	−6.06	−6.46	2440	−6.08	9478
KDD	64	1800992	19907	34955	$-2.16^{\dagger}$	14974	$\mathbf{-2.12}^{*}$	9182	**−2.16**	−2.14	6670	−2.19	16608
Plants	69	17412	2321	3482	−14.93	13129	-13.79^{*}	13951	$-13.12^{\dagger}$	**−12.69**	47802	−13.49	36960
Audio	100	15000	2000	3000	−42.53	13765	-41.98^{*}	9721	$-40.13^{\dagger}$	**−40.02**	10804	−42.06	6142
Jester	100	9000	1000	4116	−57.67	11322	-53.47^{*}	7014	$-53.08^{\dagger}$	**−52.97**	10002	−55.36	4996
Netflix	100	15000	2000	3000	−58.92	10997	-58.41^{*}	6250	$-56.91^{\dagger}$	**−56.64**	11604	−58.64	6142
Accidents	111	12758	1700	2551	−34.13	10489	-33.64^{*}	6752	$-30.02^{\dagger}$	**−30.01**	13322	−30.83	6846
Retail	135	22041	2938	4408	−11.13	4091	$\mathbf{-10.81}^{*}$	7251	$-10.97^{\dagger}$	−10.87	2162	−10.95	3158
Pumsb-Star	163	12262	1635	2452	−34.11	10489	-33.67^{*}	7965	$-28.69^{\dagger}$	**−24.11**	17604	−24.34	18338
DNA	180	1600	400	1186	-89.11^{*}	6068	−92.67	14864	$-81.76^{\dagger}$	**−85.51**	4320	−87.49	1430
Kosarek	190	33375	4450	6675	$-10.99^{\dagger}$	11034	-10.81^{*}	10179	−11.00	**−10.62**	5318	−10.98	6712
MSWeb	294	29441	32750	5000	$-10.18^{\dagger}$	11389	-9.97^{*}	14512	−10.25	**−9.90**	16484	−10.06	12770
Book	500	8700	1159	1739	−35.90	15197	-34.97^{*}	11292	$-34.91^{\dagger}$	**−34.76**	11998	−37.44	11916
EachMovie	500	4524	1002							**−57.**	15998	−58.05	19846
WebKB	839	2803	558							**−155.**	20134	−161.17	10046
Reuters-52	889	6532	1028							**−80.**	46232	−87.49	28334
20NewsGrp.	910	11293	3764	3764	−161.41	12222	-161.09^{*}	18431	$-153.63^{\dagger}$	**−154.67**	43684	−161.46	29016
BBC	1058	1670	225	330	−260.83	10185	$\mathbf{-253.19}^{*}$	28927	$-253.13^{\dagger}$	−253.45	21160	−260.59	8454
AD	1556	2461	327	491	−30.49	9666	−31.78	9521	$-16.97^{\dagger}$	−16.77	49790	**−15.39**	31070

Comparable in performance & Smaller in size

Ensembles of PSDDs



EM/Bagging

State-of-the-Art Performance

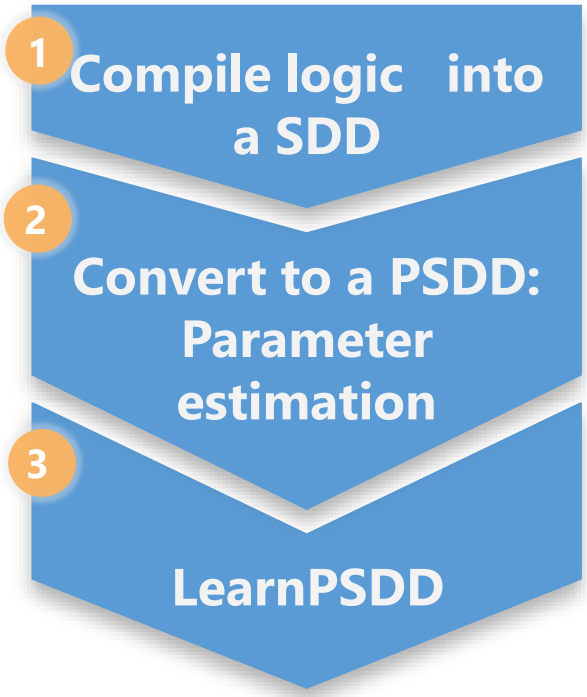
Datasets	Var	LearnPSDD Ensemble	Best-to-Date
NLTCS	16	−5.99†	−6.00
MSNBC	17	−6.04†	−6.04†
KDD	64	−2.11†	−2.12
Plants	69	−13.02	−11.99†
Audio	100	−39.94	−39.49†
Jester	100	−51.29	−41.11†
Netflix	100	−55.71†	−55.84
Accidents	111	−30.16	−24.87†
Retail	135	−10.72†	−10.78
Pumsb-Star	163	−26.12	−22.40†
DNA	180	−88.01	−80.03†
Kosarek	190	−10.52†	−10.54
MSWeb	294	−9.89	−9.22†
Book	500	−34.97	−30.18†
EachMovie	500	−58.01	−51.14†
WebKB	839	−161.09	−150.10†
Reuters-52	889	−89.61	−80.66†
20NewsGrp.	910	−155.97	−150.88†
BBC	1058	−253.19	−233.26†
AD	1556	−31.78	−14.36†

State-of-the-art in 6 datasets

Retain the ability to fit logically constrained distributions

Learning in Logically Constrained Domains

Roadmap



Experiments in Logically Constrained Domains

Discrete multi-valued data

$$A\colon a_1, a_2, a_3$$

$$A_1, A_2, A_3$$

$$(a_1 \vee a_2 \vee a_3) \wedge \begin{cases} a_1 \wedge \neg a_2 \wedge \neg a_3 \\ \vee \\ \neg a_1 \wedge a_2 \wedge \neg a_3 \\ \vee \\ \neg a_1 \wedge \neg a_2 \wedge a_3 \end{cases}$$

Datasets	No Constraint	PSDD	LearnPSDD
Adult	−18.41	−14.14	−12.86
CovType	−14.39	−8.81	−7.32

Never omit domain constraints

Summary

No constraints

1. Determine variable tree
2. Construct the most naïve PSDD
3. LearnPSDD (search for better structure)

With constraints

1. Compile Logics into a SDD
2. Convert to PSDD: Parameter estimation
3. LearnPSDD

State-of-the-art Performance

Thanks

https://github.com/UCLA-StarAI/LearnPSDD