Learning Logistic Circuits

Yitao Liang, Guy Van den Broeck

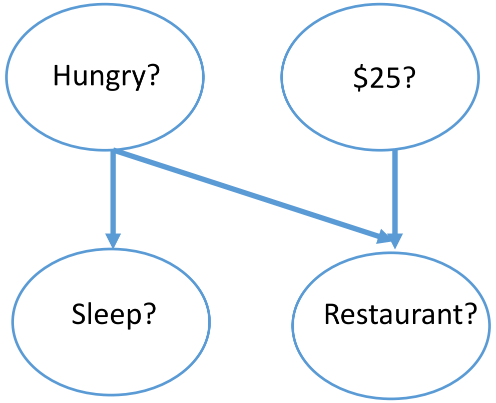
January 31, 2019

STAR AI
RESEARCH LAB
UCLA

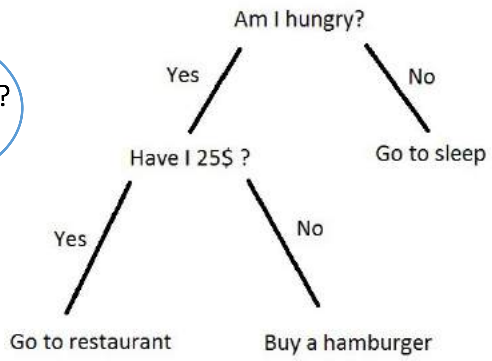
UCLA

Which model to choose

Classical AI Methods:

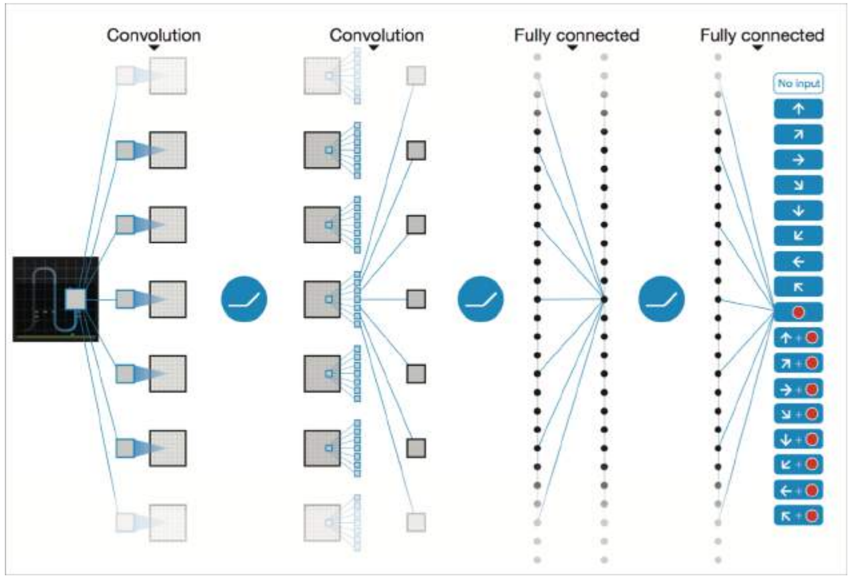


…



Clear Modeling Assumption

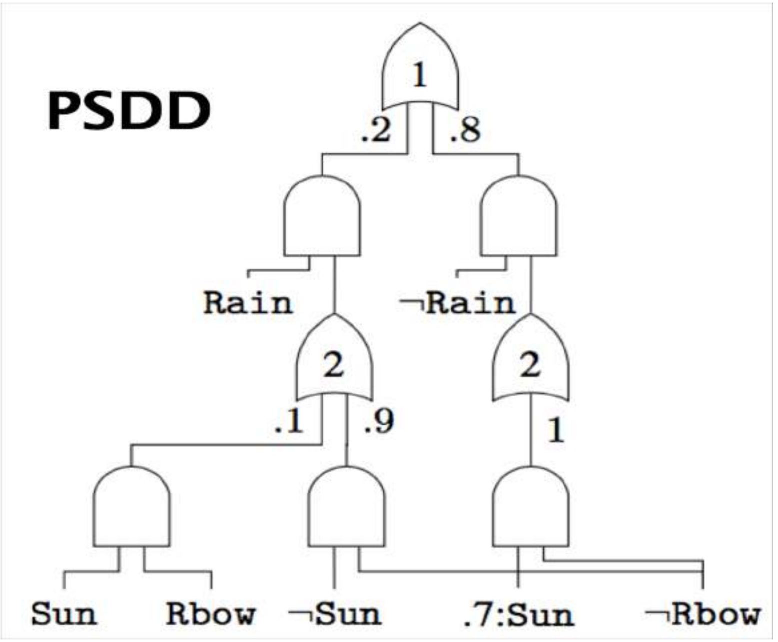
Neural Networks:

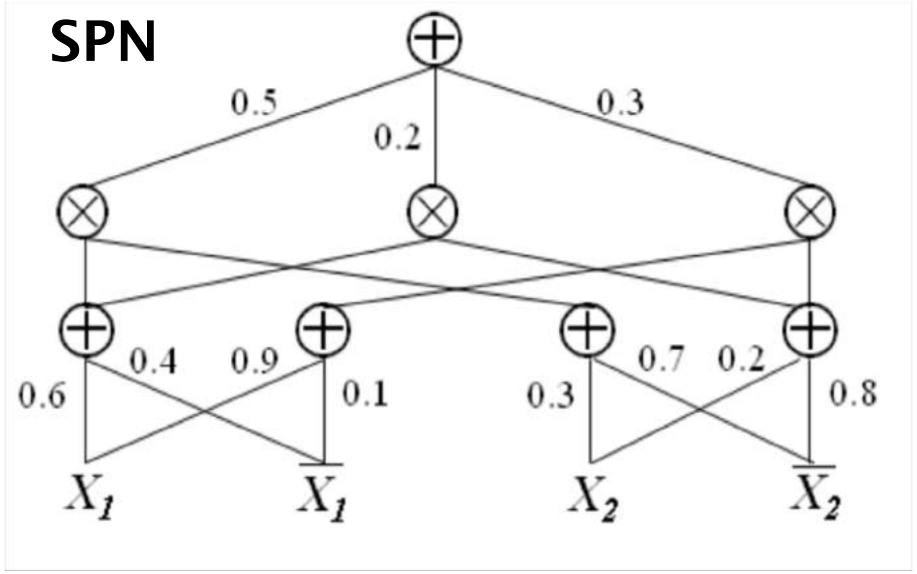


"Black Box"
Good performance on Image Classification

Starting Point: Probabilistic Circuits

A promising synthesis of the two





State-of-the-art on Density Estimation

$$\Pr(X)$$

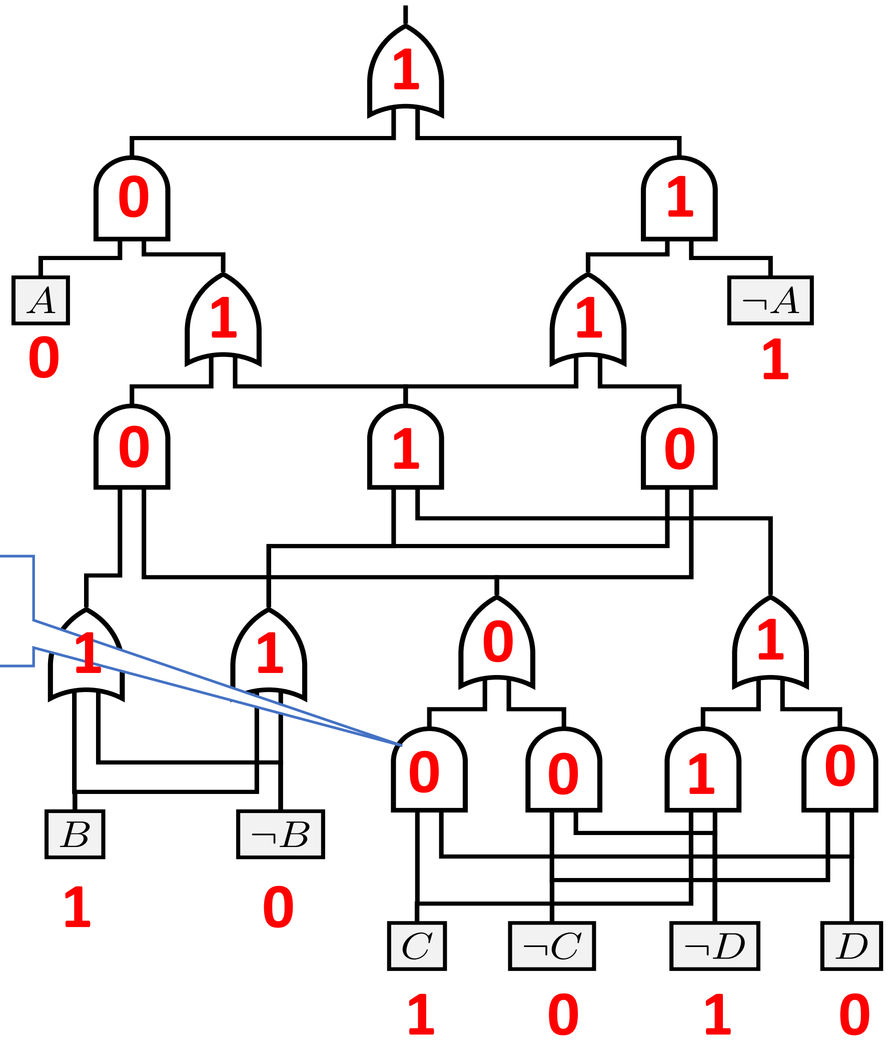
What if we only want to learn a classifier $\Pr(Y|X)$

Logical Circuits

Input:

A	B	C	D
0	1	1	0

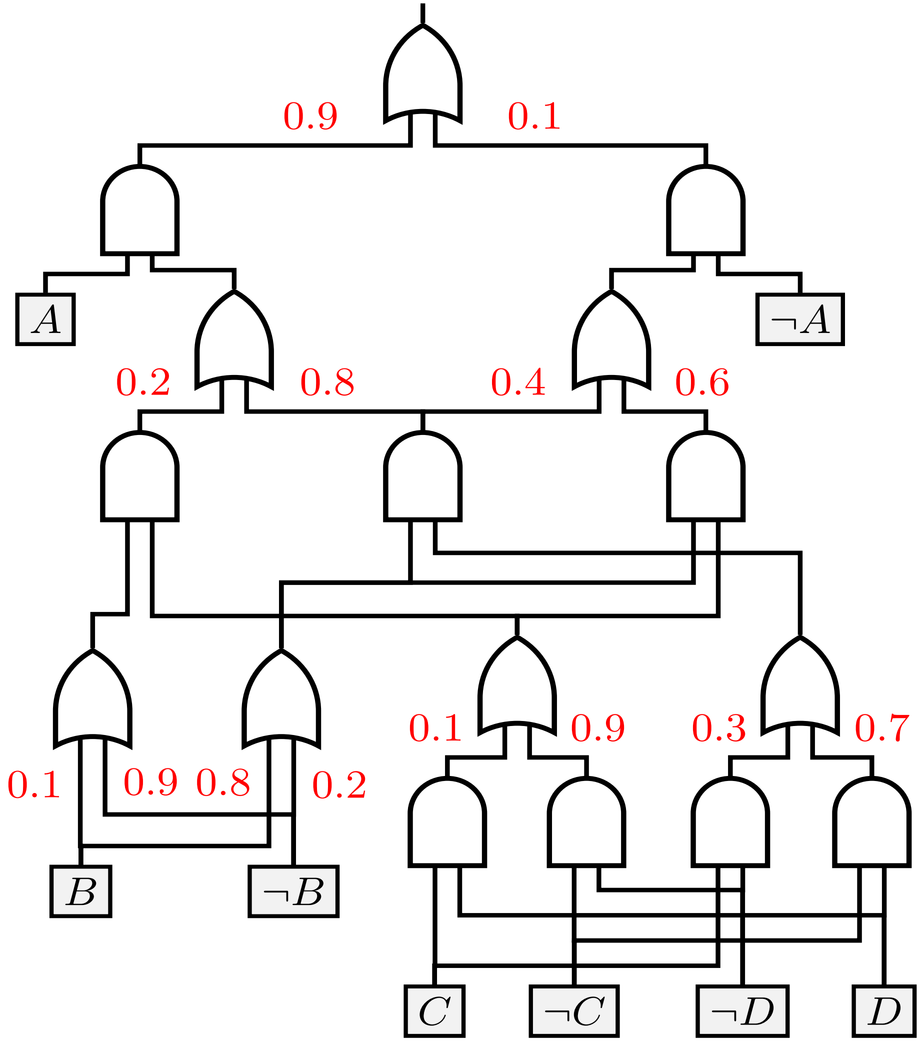
0 = 1 AND 0



Bottom-up Evaluation

Logical -> Probabilistic Circuits

Red Parameters:
Conditional Probabilities

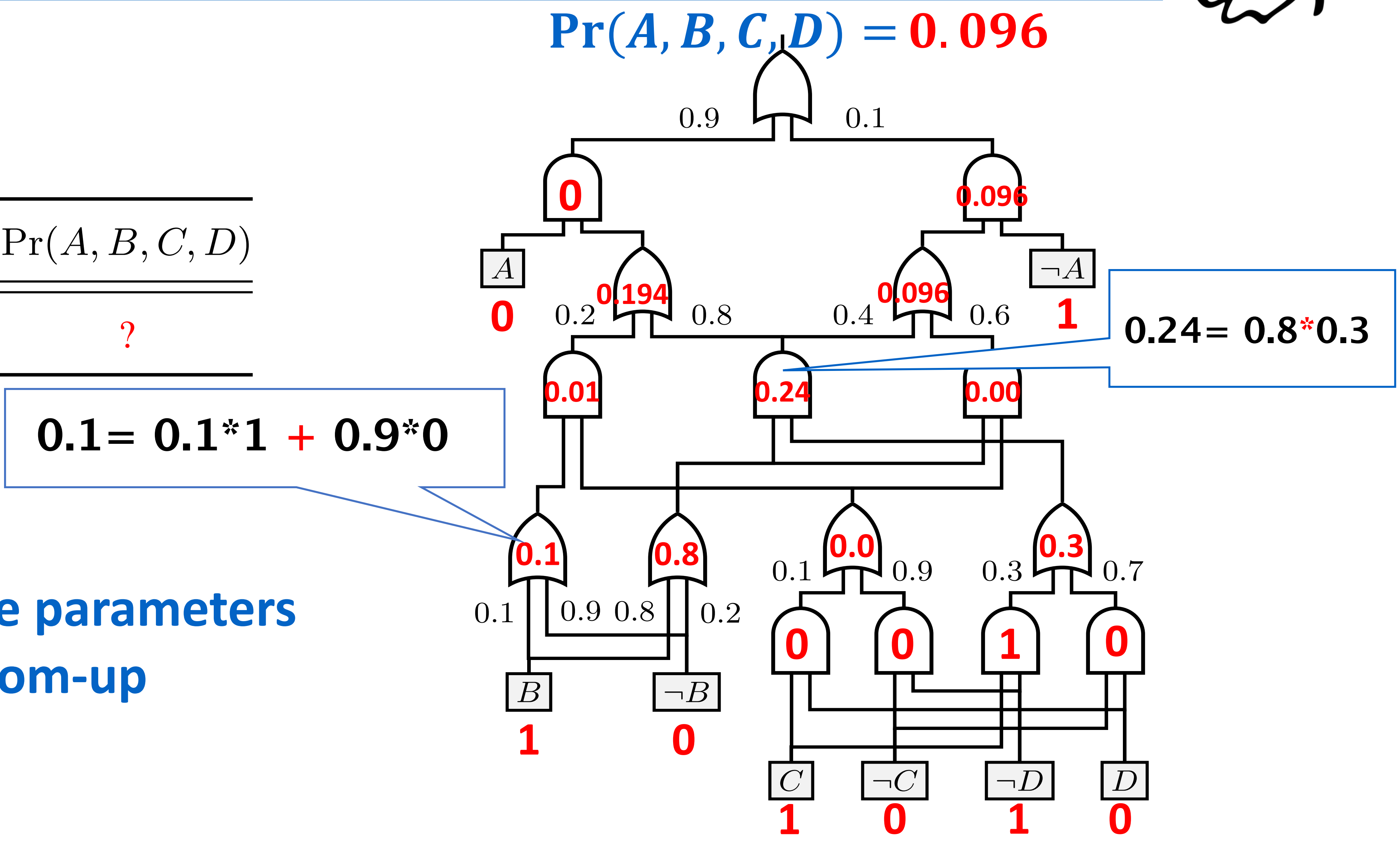


Logical -> Probabilistic Circuits

Input:

A	B	C	D	$\Pr(A,B,C,D)$
0	1	1	0	?

Multiply the parameters bottom-up

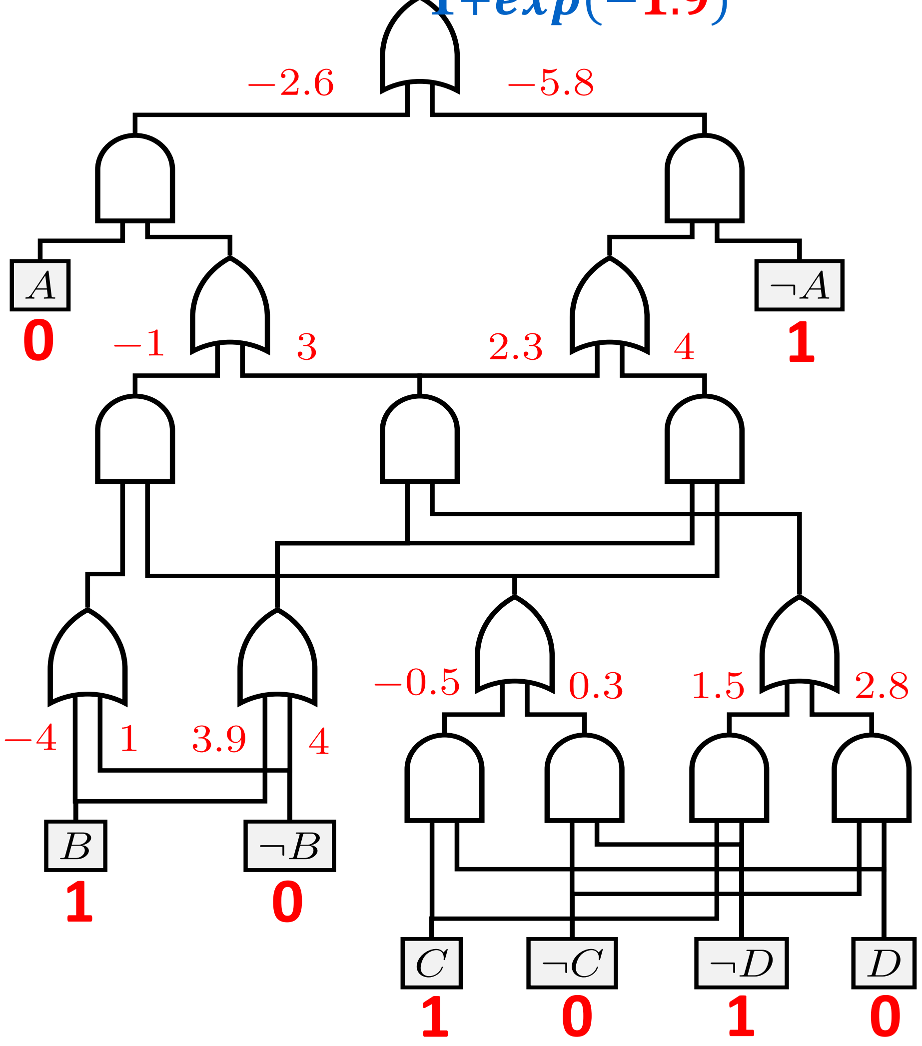


Evaluate Logistic Circuits

$$\Pr(Y=1 \mid A,B,C,D) = \frac{1}{1+exp(-1.9)} = 0.869$$

Input:

A	B	C	D	$\Pr(Y \mid A,B,C,D)$
0	1	1	0	?

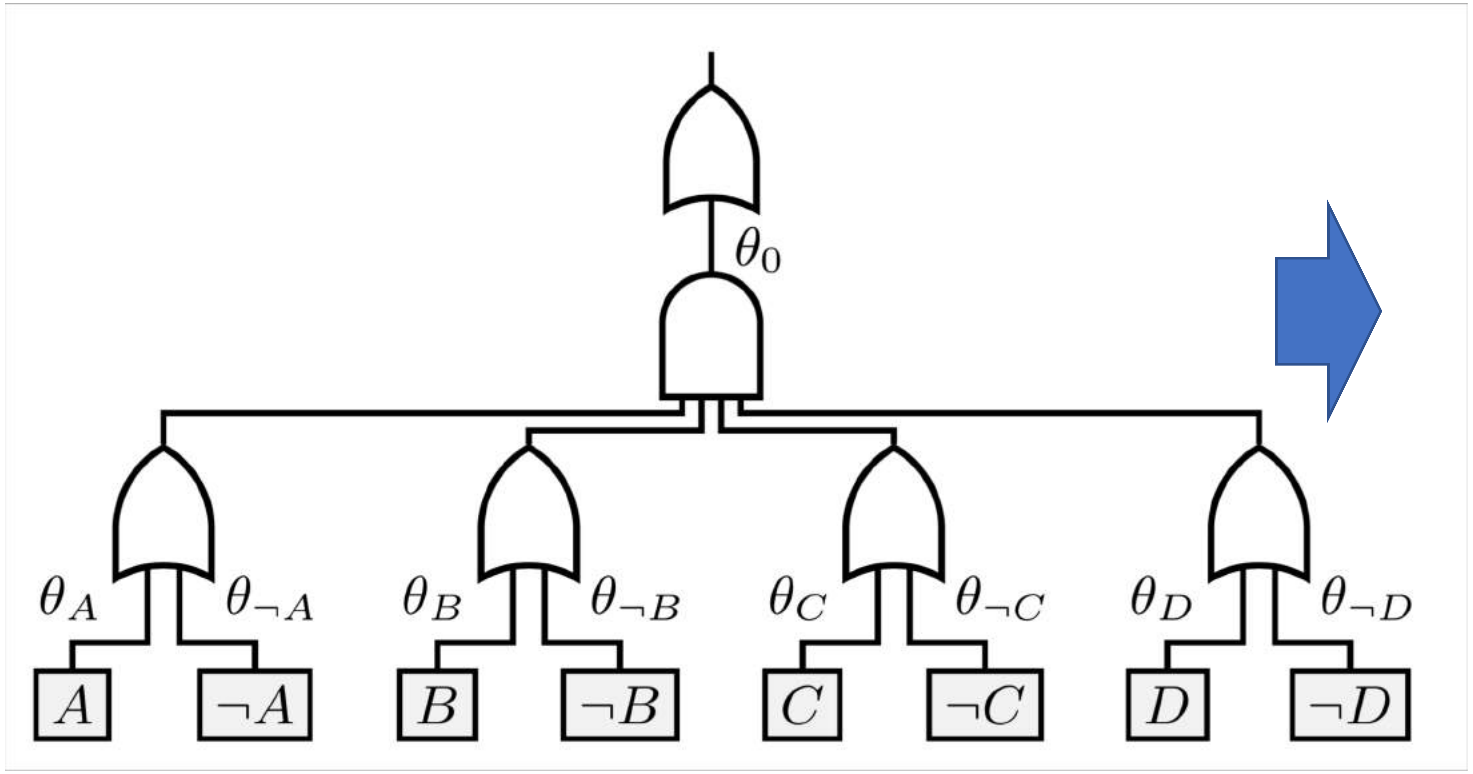


Multiply the parameters bottom-up

Logistic function on final output

Are logistic circuits amenable to tractable learning

Special Case: Logistic Regression

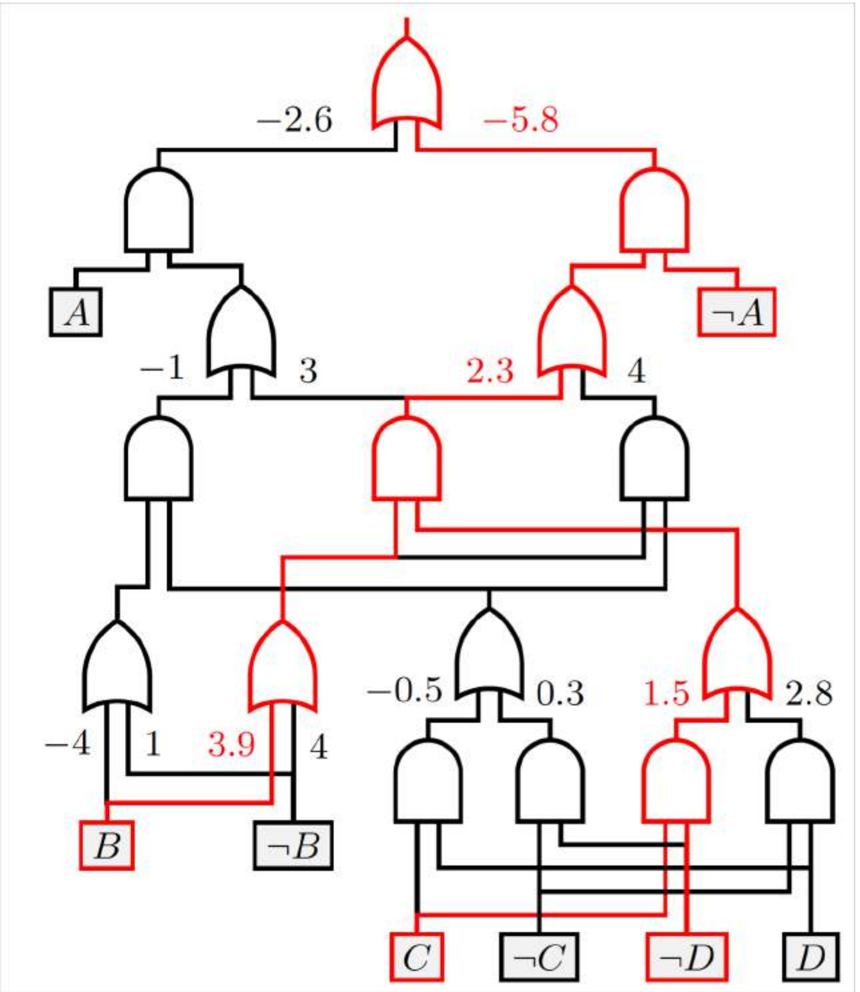


Logistic Regression

$$\Pr(Y = 1 | A, B, C, D) = \frac{1}{1 + \exp(-A * \theta_A - \neg A * \theta_{\neg A} - B * \theta_B - \cdots)}$$

What about other logistic circuits in more general forms?

Parameter Learning



$$\Pr(Y = 1 \mid A = 0, B = 1, C = 1, D = 0)$$

"Hot" wires are active features

Parameter Learning

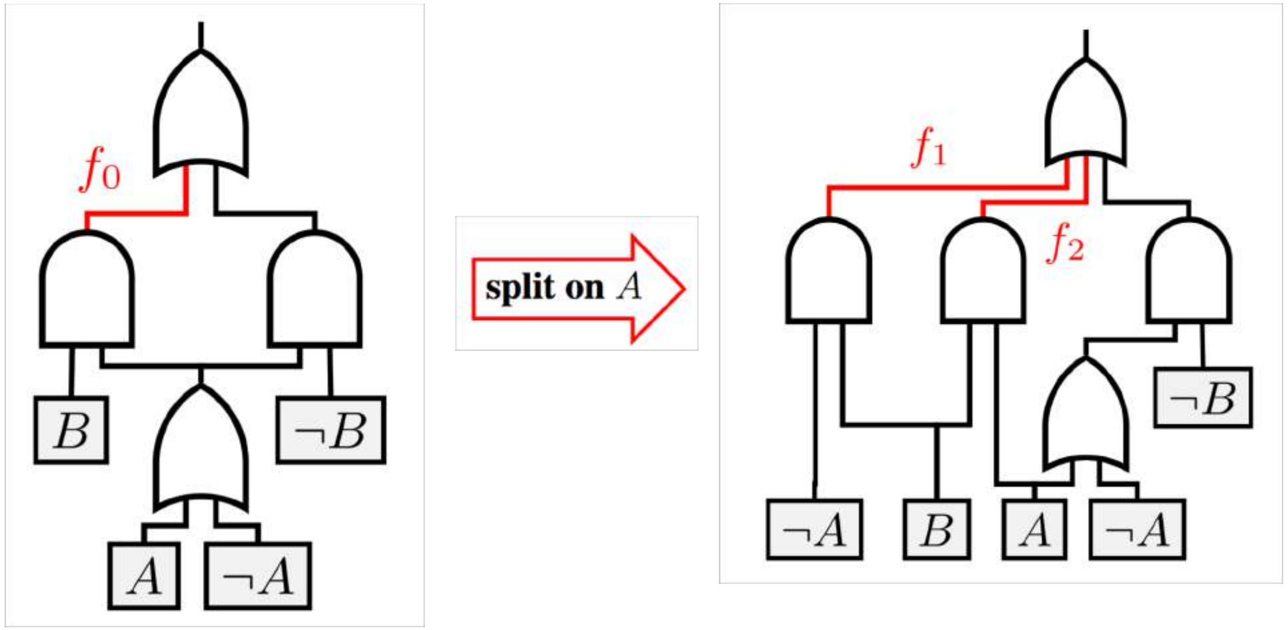
Due to decomposability and determinism, reduce to logistic regression

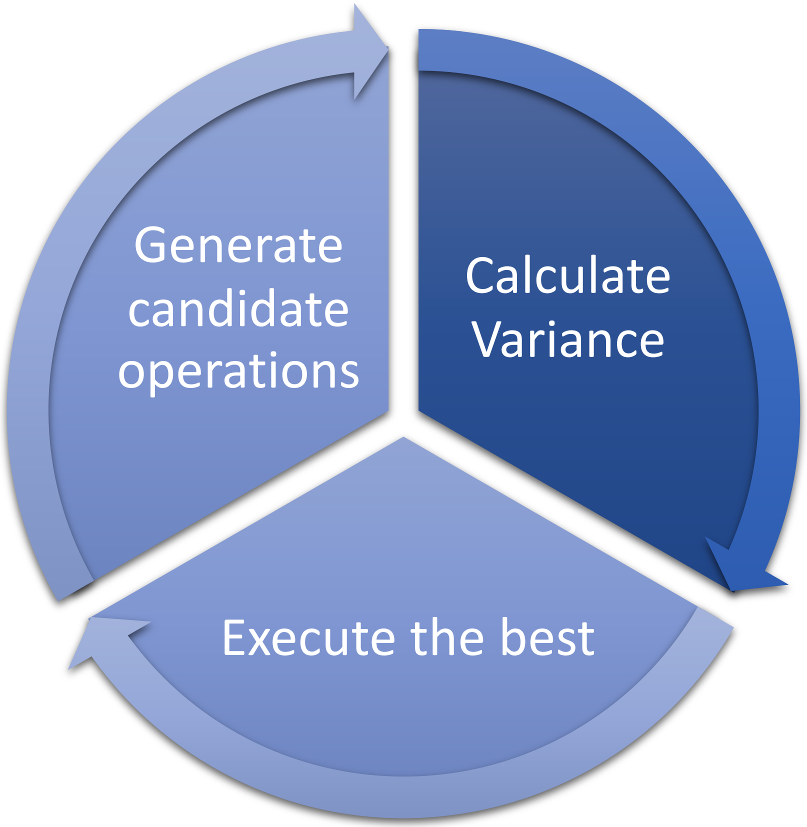
$$\Pr(Y = 1 \mid \mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{x} \cdot \boldsymbol{\theta})}$$

Features associated with each wire
"Global Circuit Flow"

Convex Parameter learning

Structure Learning





Similar to LearnPsdd

Split nodes to reduce variance of gradients

Comparable Accuracy with Neural Nets

Accuracy % on Dataset	Mnist	Fashion
Baseline: Logistic Regression	85.3	79.3
Baseline: Kernel Logistic Regression	97.7	88.3
Random Forest	97.3	81.6
3-layer MLP	97.5	84.8
RAT-SPN (Peharz et al. 2018)	98.1	89.5
SVM with RBF Kernel	98.5	87.8
5-Layer MLP	99.3	89.8
Logistic Circuit (binary)	97.4	87.6
Logistic Circuit (real-valued)	99.4	91.3
CNN with 3 conv layers	99.1	90.7
Resnet (He et al. 2016)	99.5	93.6

Significantly Smaller in Size

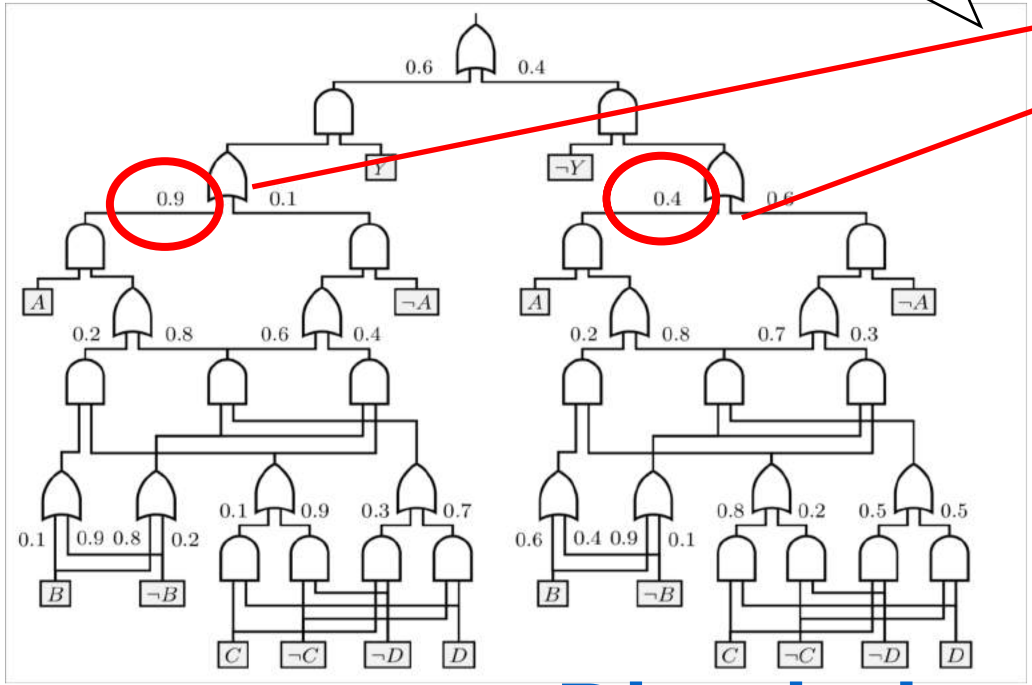
Number of Parameters	MNIST	Fashion
Baseline: Logistic Regression	<1K	<1K
Baseline: Kernel Logistic Regression	1,521 K	3,930K
Logistic Circuit (real-valued)	182K	467K
Logistic Circuit (binary)	268K	614K
3-layer MLP	1,411K	1,411K
RAT-SPN (Peharz et al. 2018)	8,500K	650K
CNN with 3 conv layers	2,196K	2,196K
5-Layer MLP	2,411K	2,411K
ResNet (He et al. 2016)	4,838K	4,838K

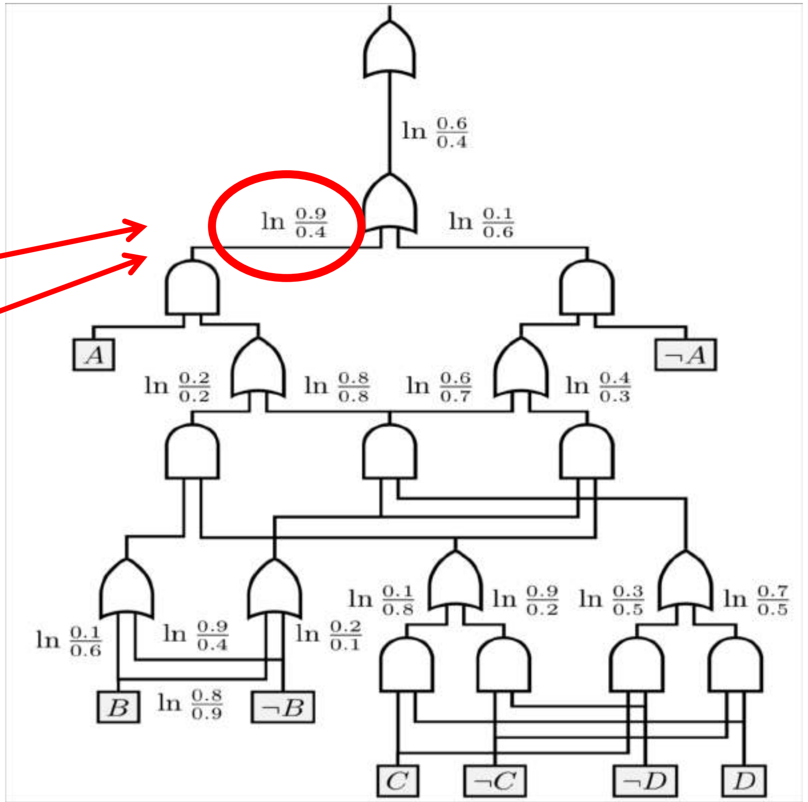
Better Data Efficiency

Accuracy % with % of Training Data	MNIST			Fashion		
	100%	10%	2%	100%	10%	2%
5-layer MLP	99.3	**98.2**	94.3	89.8	86.5	80.9
CNN with 3 Conv Layers	99.1	98.1	95.3	90.7	87.6	83.8
Logistic Circuit (Binary)	97.4	96.9	94.1	87.6	86.7	83.2
Logistic Circuit (Real-Valued)	**99.4**	97.8	**96.1**	**91.3**	**87.8**	**86.0**

Probabilistic -> Logistic Circuits

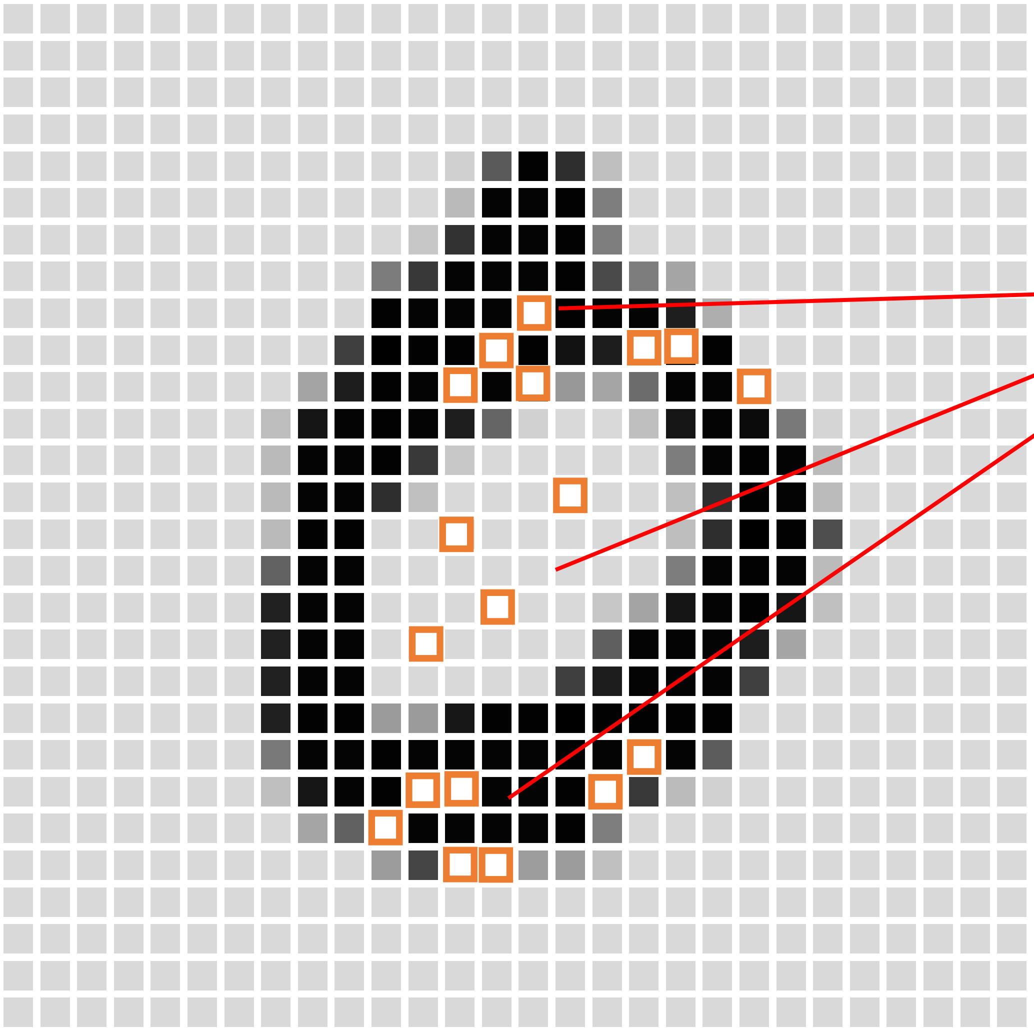
Probabilities become log-odds





Discriminative Counterparts

What do Features Mean



This is the feature that contributes the most to this image's classification probability

feature value : 0.925
feature weight : 3.489

feature interpretation:
curvy lines and hallow center

Conclusion

Logistic circuits:

- Synthesis of symbolic AI and statistical learning
- Discriminative counterparts of probabilistic circuits
- Convex parameter learning
- Simple heuristic for structure learning
- Good performance
- Easy to interpret

Thanks

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