Sum-Product Networks

CS886 Topics in Natural Language Processing

Guest Lecture by Pascal Poupart

University of Waterloo

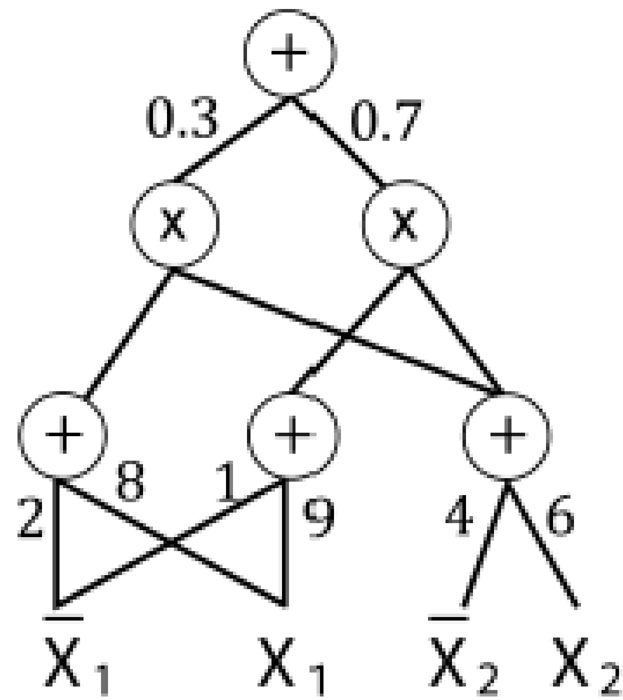
July 22, 2015

Outline

- What is a Sum-Product Network?
- Inference
- Language modeling

What is a Sum-Product Network?

- Poon and Domingos, UAI 2011
- Acyclic directed graph of sums and products
- Leaves can be indicator variables or univariate distributions



Two Views

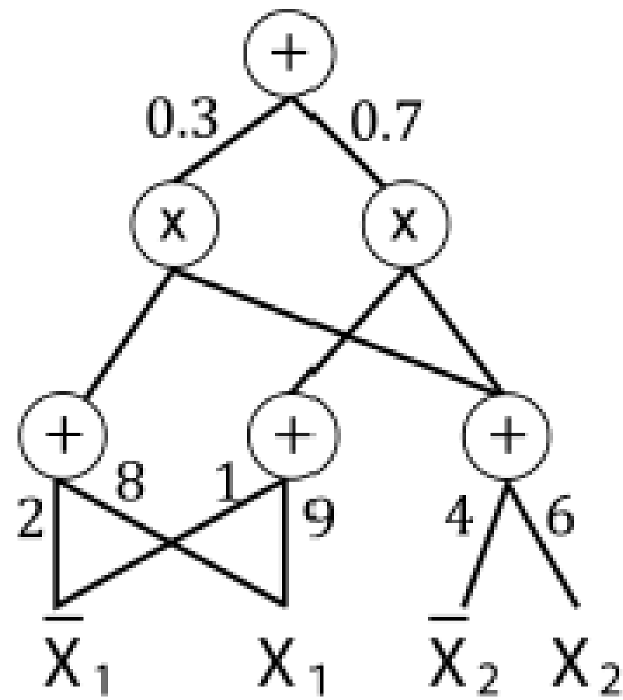
Deep architecture with clear semantics

Tractable probabilistic graphical model

Deep Architecture

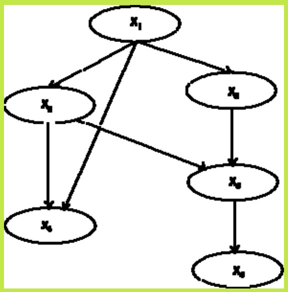
- Specific type of deep neural network
  - Activation function: product

- Advantage:
  - Clear semantics and well understood theory



Probabilistic Graphical Models

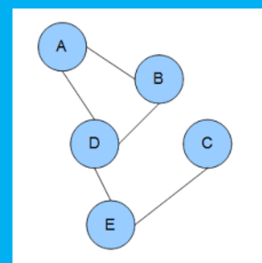
Bayesian Network



Graphical view of direct dependencies

Inference
#P: intractable

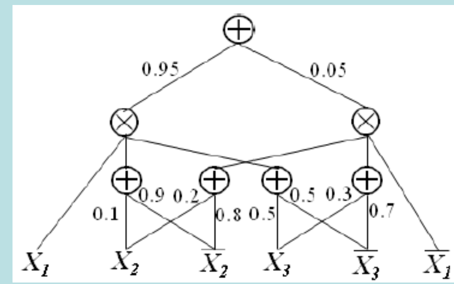
Markov Network



Graphical view of correlations

Inference
#P: intractable

Sum-Product Network

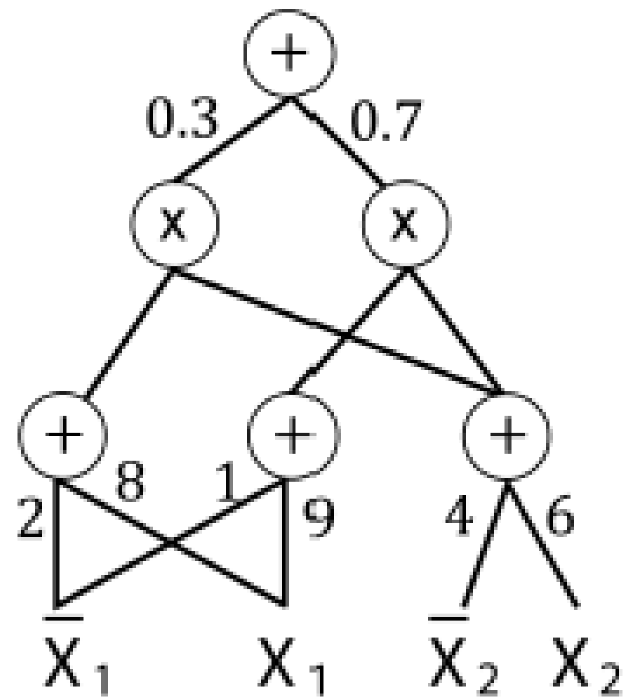


Graphical view of computation

Inference
P: tractable

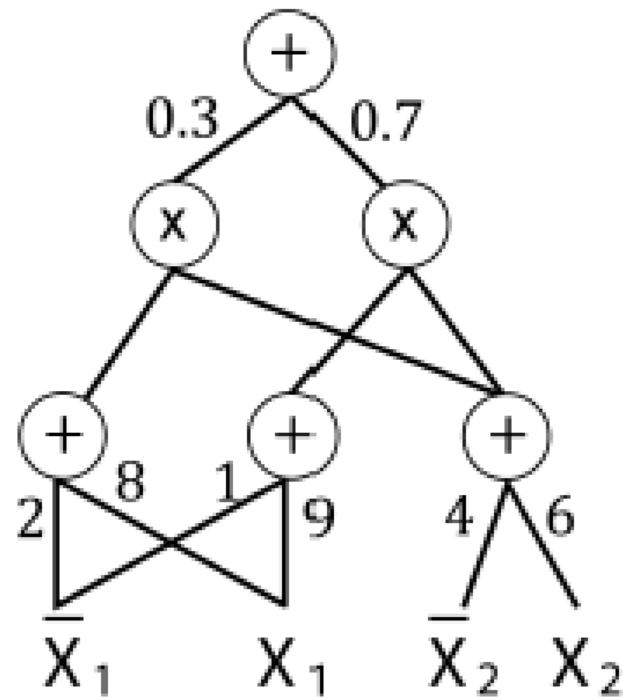
Probabilistic Inference

- SPN represents a joint distribution over a set of random variables
- Example:
  $\Pr(X_1 = true, X_2 = false)$



Marginal Inference

- Example:
  $\Pr(X_2 = false)$



Conditional Inference

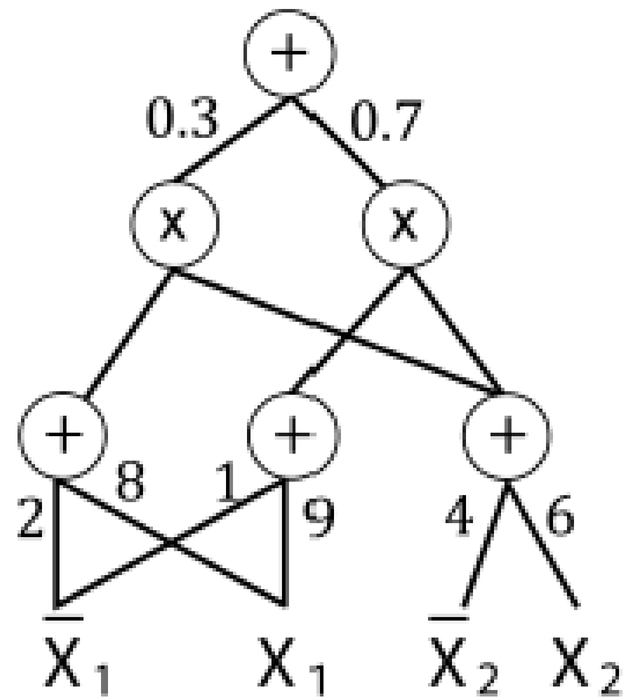
- Example:

$$\Pr(X_1 = true | X_2 = false)$$
$$= \frac{\Pr(X_1 = true, X_2 = false)}{\Pr(X_2 = false)}$$
$$=$$

- Hence any inference query can be answered in two bottom passes of the network
  - **Linear complexity!**

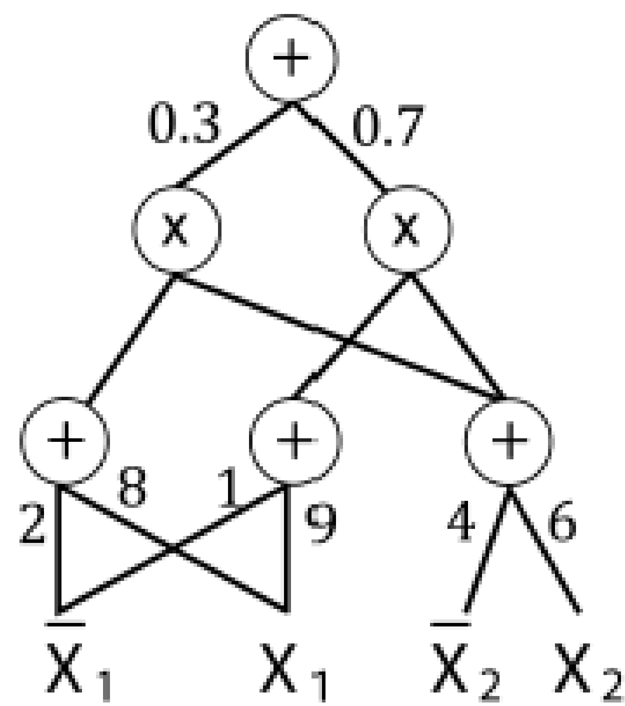
Semantics

- A **valid** SPN encodes a hierarchical mixture distribution
  - Sum nodes: hidden variables (mixture)
  - Product nodes: factorization (independence)



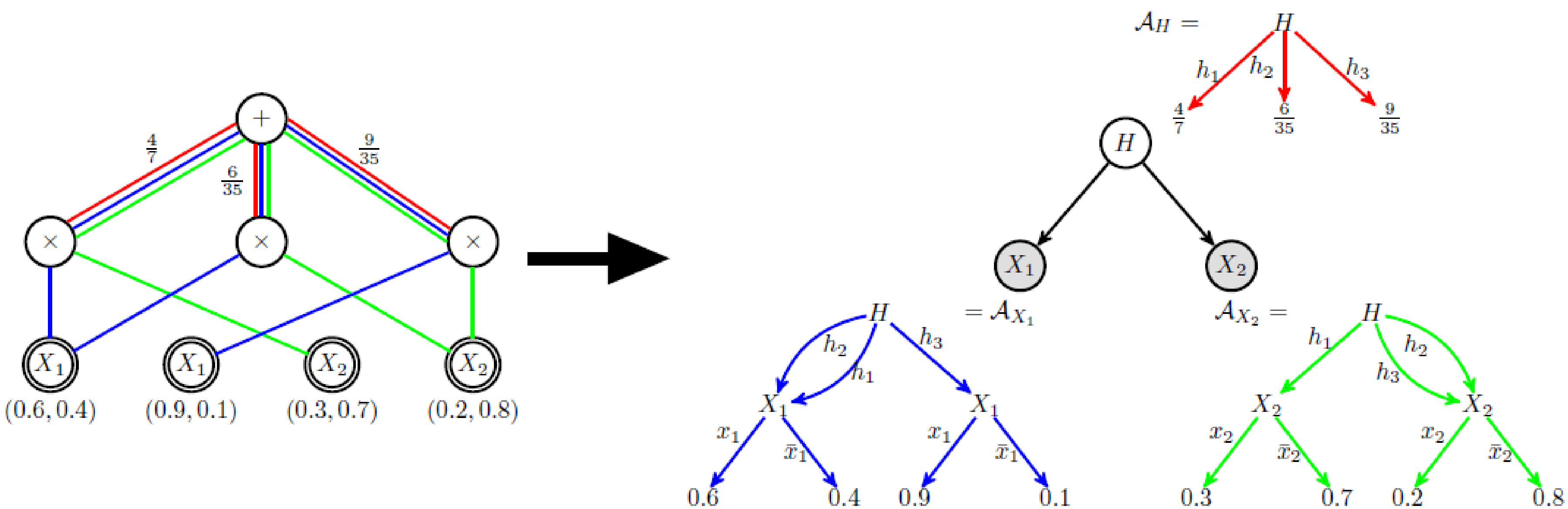
Definitions

- The scope of a node is the set of variables that appear in the sub-SPN rooted at the node
- An SPN is decomposable when each product node has children with disjoint scopes
- An SPN is complete when each sum node has children with identical scopes
- A decomposable and complete SPN is a valid SPN

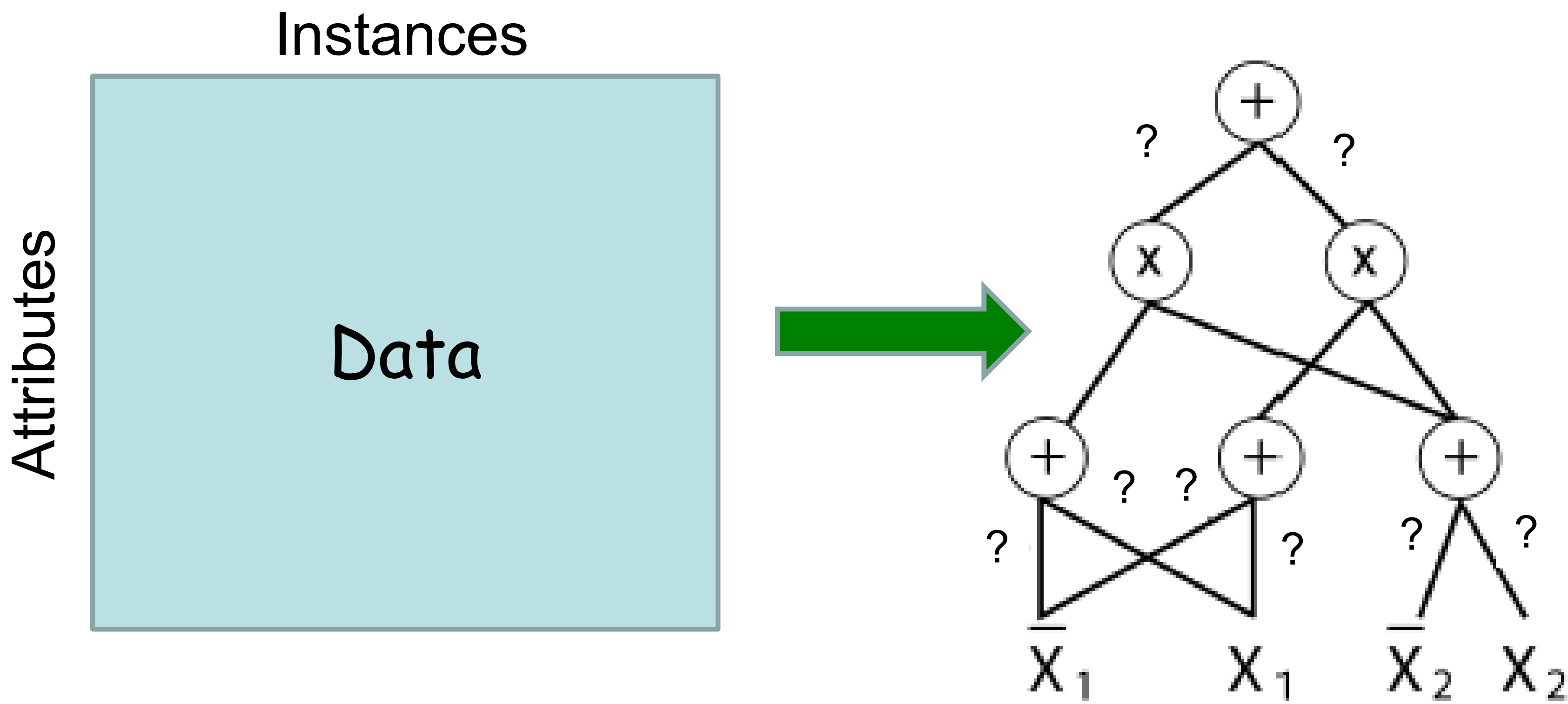


Relationship with Bayes Nets

- Any SPN can be converted into a bipartite Bayesian network (Zhao, Melibari, Poupart, ICML 2015)



Parameter Learning



- Parameter Learning: estimate the weights
  – Expectation-Maximization, Gradient descent

Structure Learning

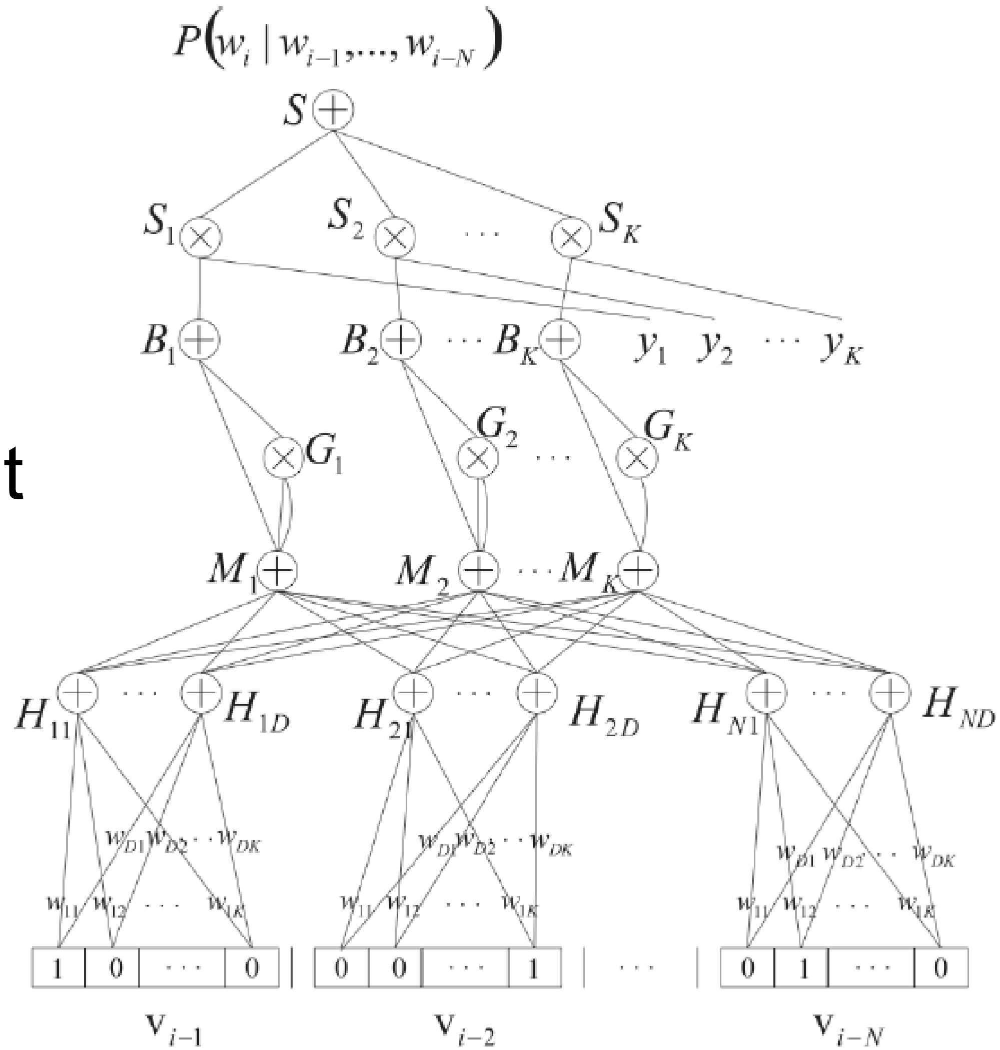
- Alternate between
  - Data Clustering: sum nodes
  - Variable partitioning: product nodes

Applications

- Image completion (Poon, Domingos; 2011)
- Activity recognition (Amer, Todorovic; 2012)
- **Language modeling (Cheng et al.; 2014)**
- Speech modeling (Perhaz et al.; 2014)

Language Model

- An SPN-based n-gram model
- Fixed structure
- Discriminative weight learning by gradient descent



Results

- From Cheng et al. 2014

Table 1: Perplexity scores (PPL) of different language models.

Model	Individual PPL	+KN5
TrainingSetFrequency	528.4	
KN5 [3]	141.2	
Log-bilinear model [4]	144.5	115.2
Feedforward neural network [5]	140.2	116.7
Syntactical neural network [8]	131.3	110.0
RNN [6]	124.7	105.7
LDA-augmented RNN [9]	113.7	98.3
SPN-3	**104.2**	**82.0**
SPN-4	**107.6**	**82.4**
SPN-4'	**100.0**	**80.6**

Conclusion

- Sum-Product Networks
  - Deep architecture with clear semantics
  - Tractable probabilistic graphical model
- Work in progress at Waterloo
  - Improved structure learning: H. Zhao
  - Online parameter learning: H. Zhao, A. Rashwan
  - SPNs for sequence data: M. Melibari
  - Decision SPNs: M. Melibari
- Open problem:
  - Thorough comparison of SPNs to other deep networks