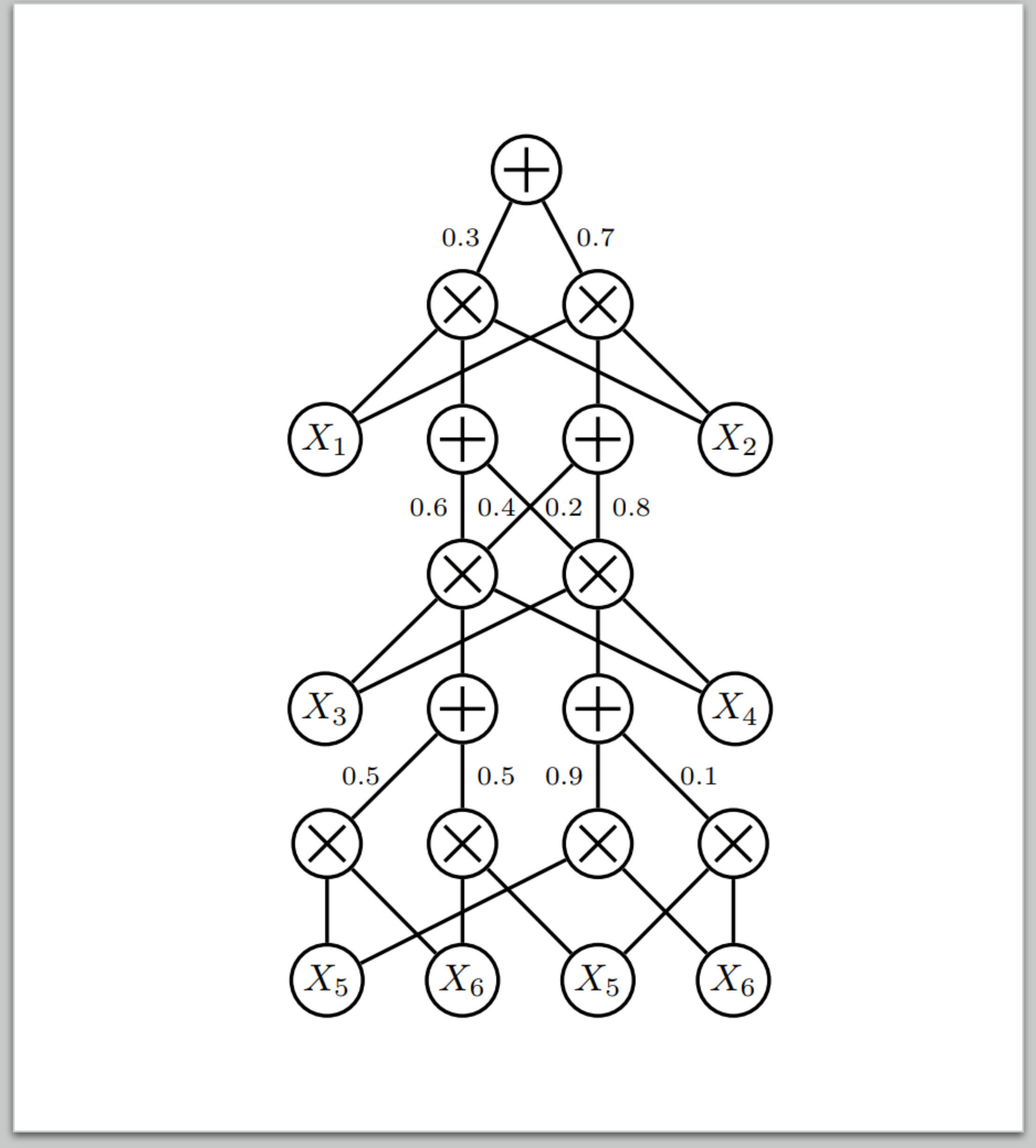
An introduction to Sum-Product Networks (SPNs): A new deep probabilistic architecture

Felix McGregor

Prof Johan du Preez

What are SPNs?

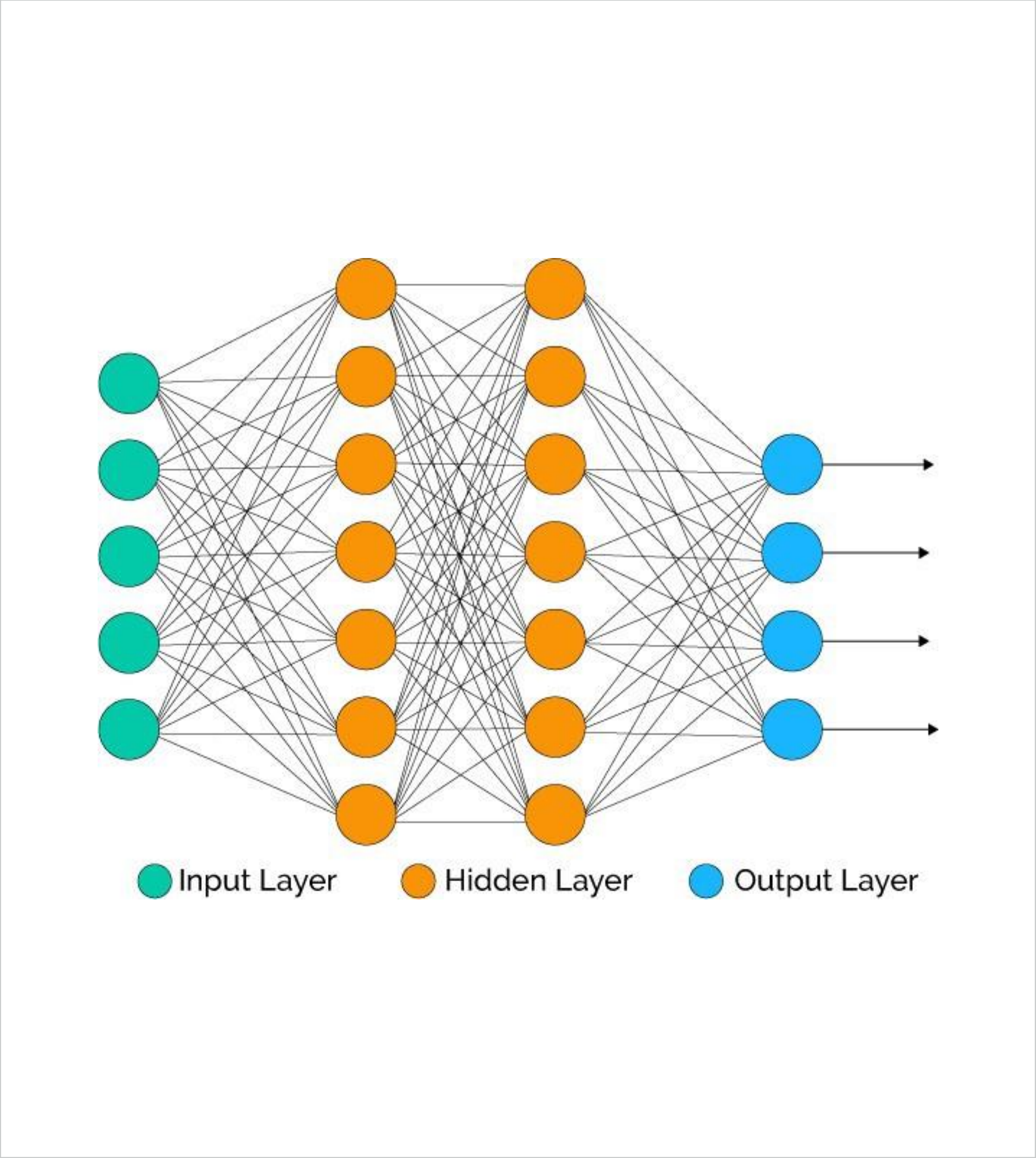
- Poon and Domingos, (UAI 2011 Best Paper)
- Acyclic directed graphs os sums and products



Two views of SPNs

1. Deep architecture

- Product node as activation function
- Clear semantics
- Reason meaningfully about relationships between variables as we are calculating probabilities with respect to some features

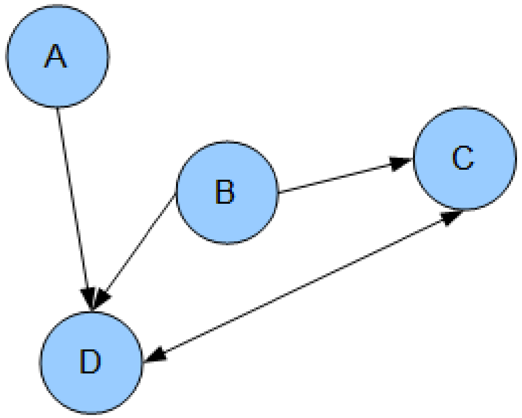


Two views of SPNs

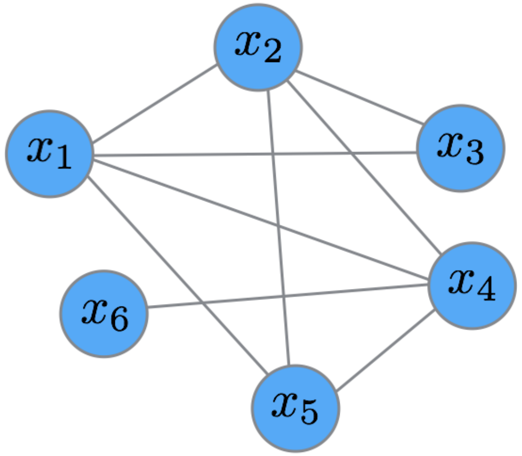
2. Probabilistic graphical models

- Tractable inference can calculate partition
- Inference in linear time to the size of the network

Bayesian Networks

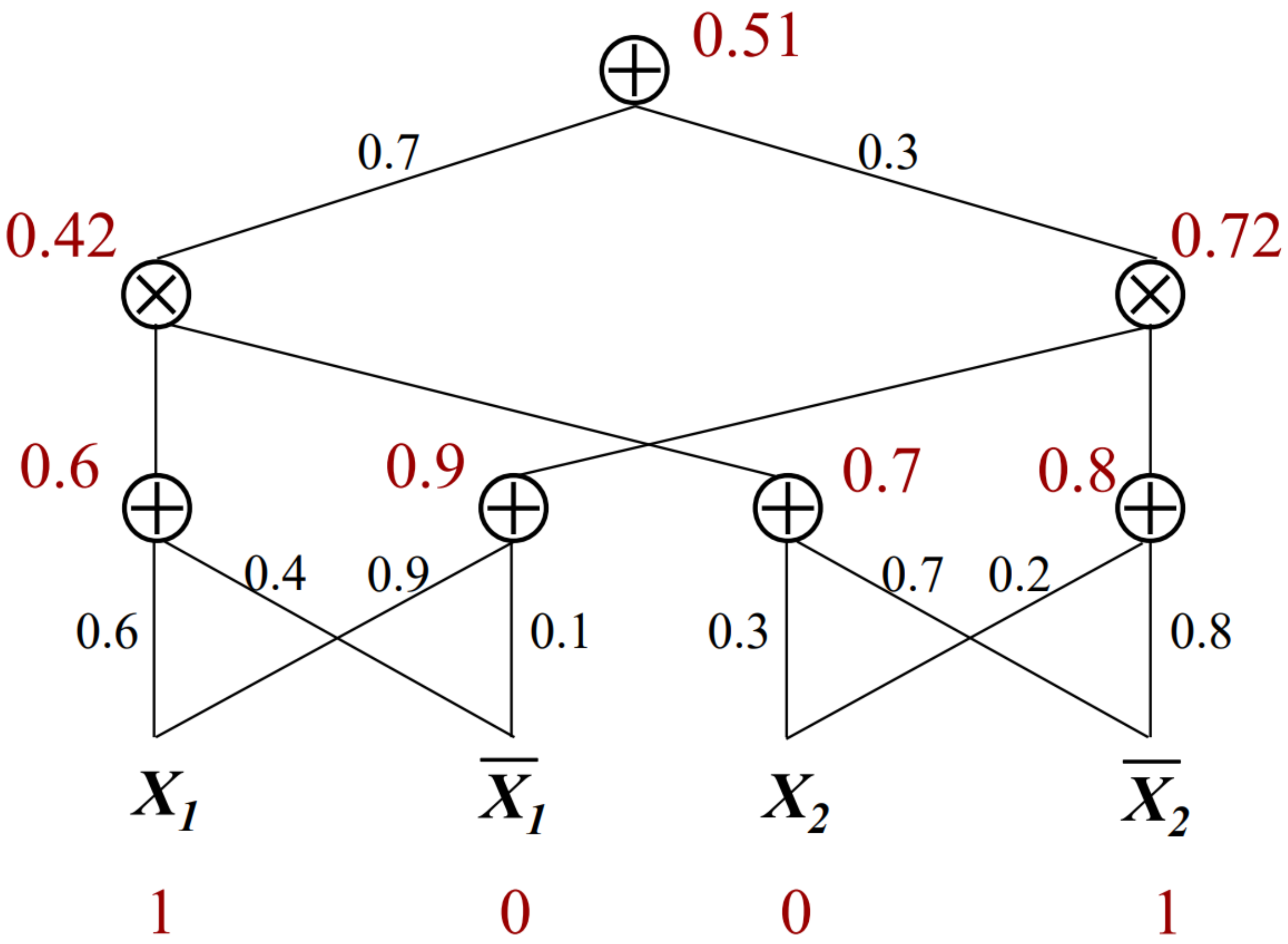


Markov Random Fields



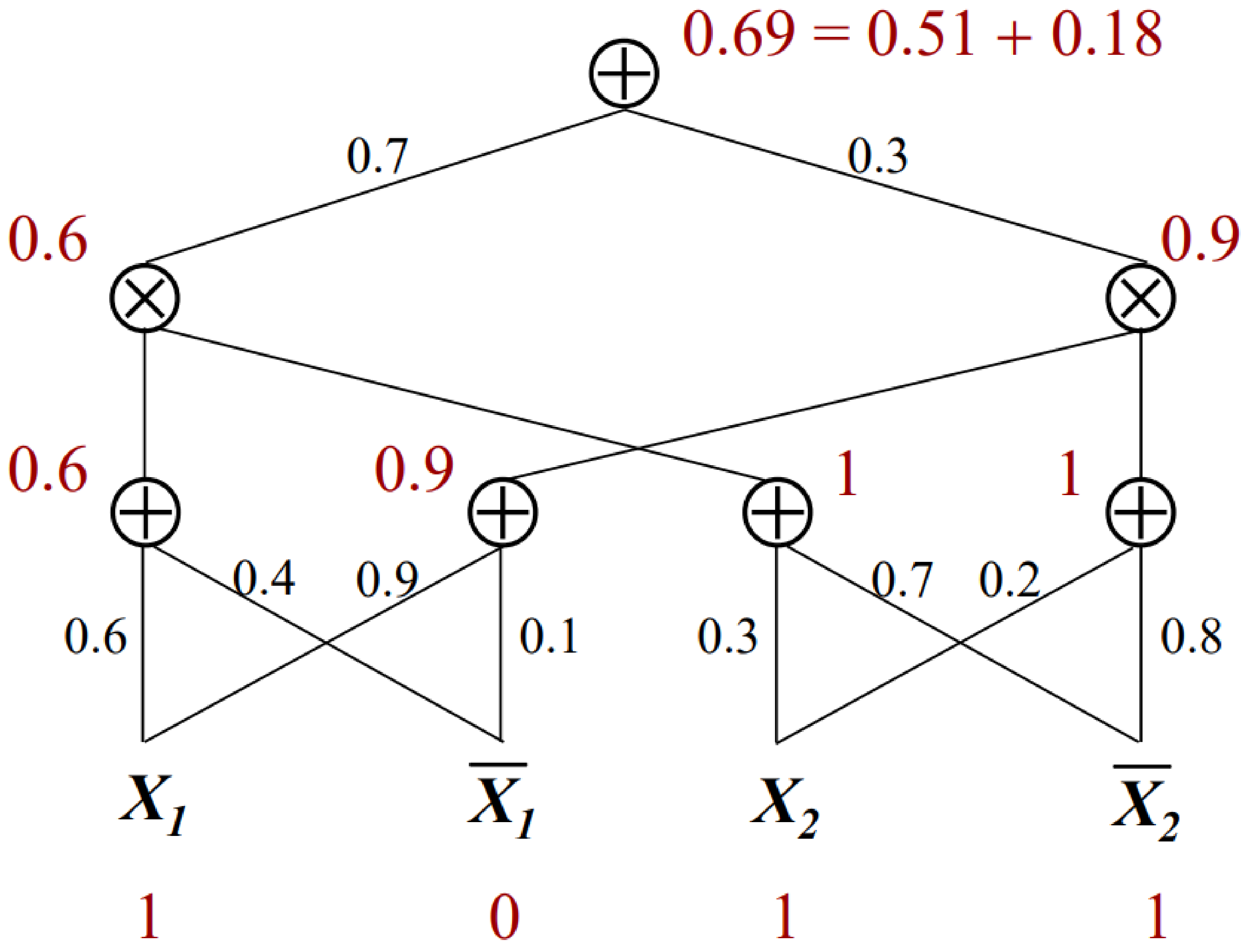
Probabilistic inference

- An SPN represents a joint distribution over a set of variables
- P(X_1= 1, X_2= 0) ?



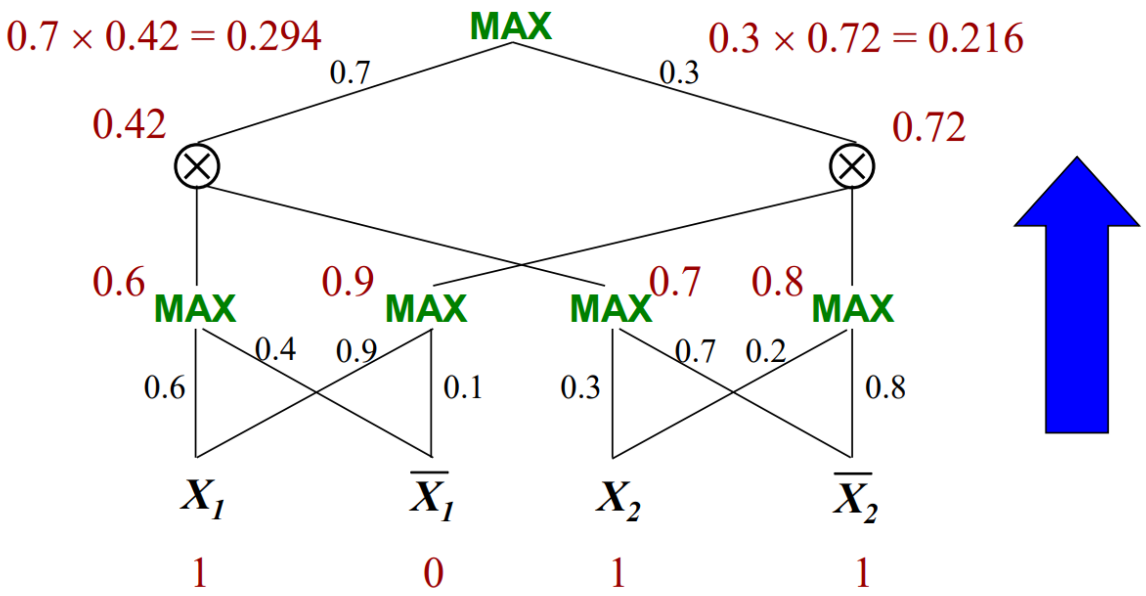
Marginal Inference

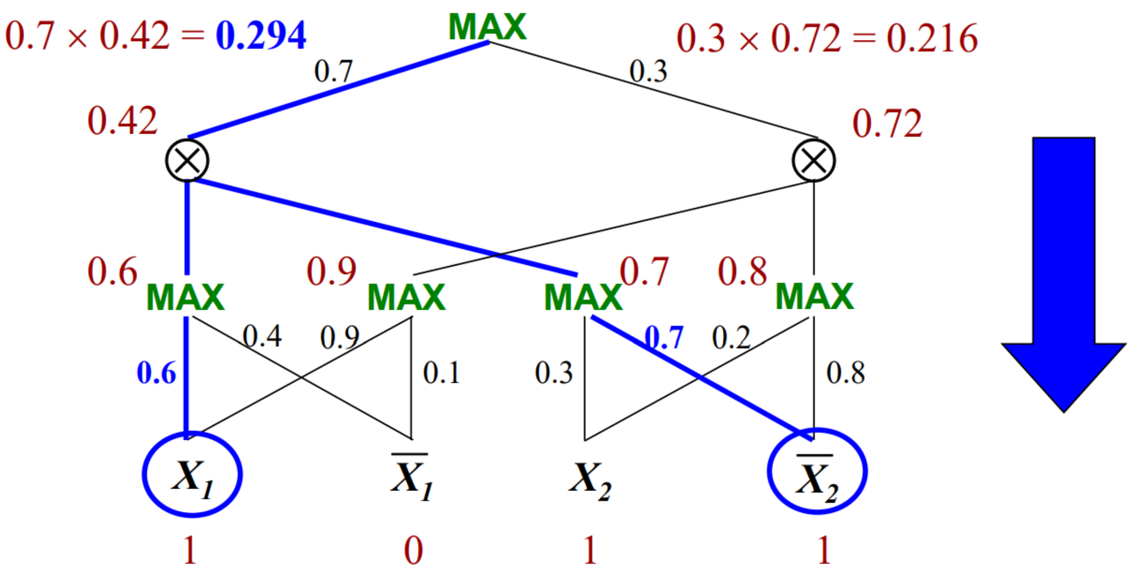
- P(X_1= 1) ?



MPE Inference

- What is the most likely state?





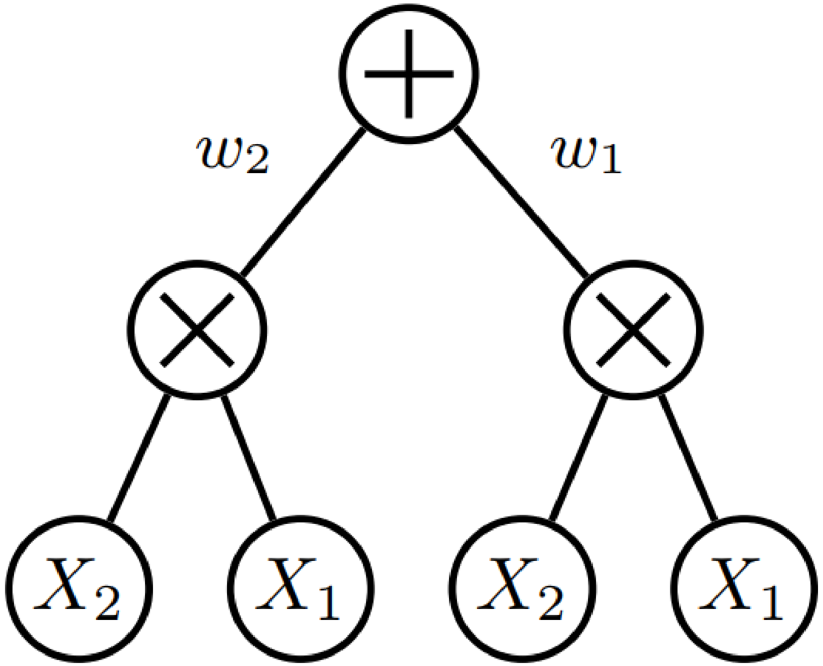
Conditions for a valid SPN

- Complete: children of sum are of the same scope

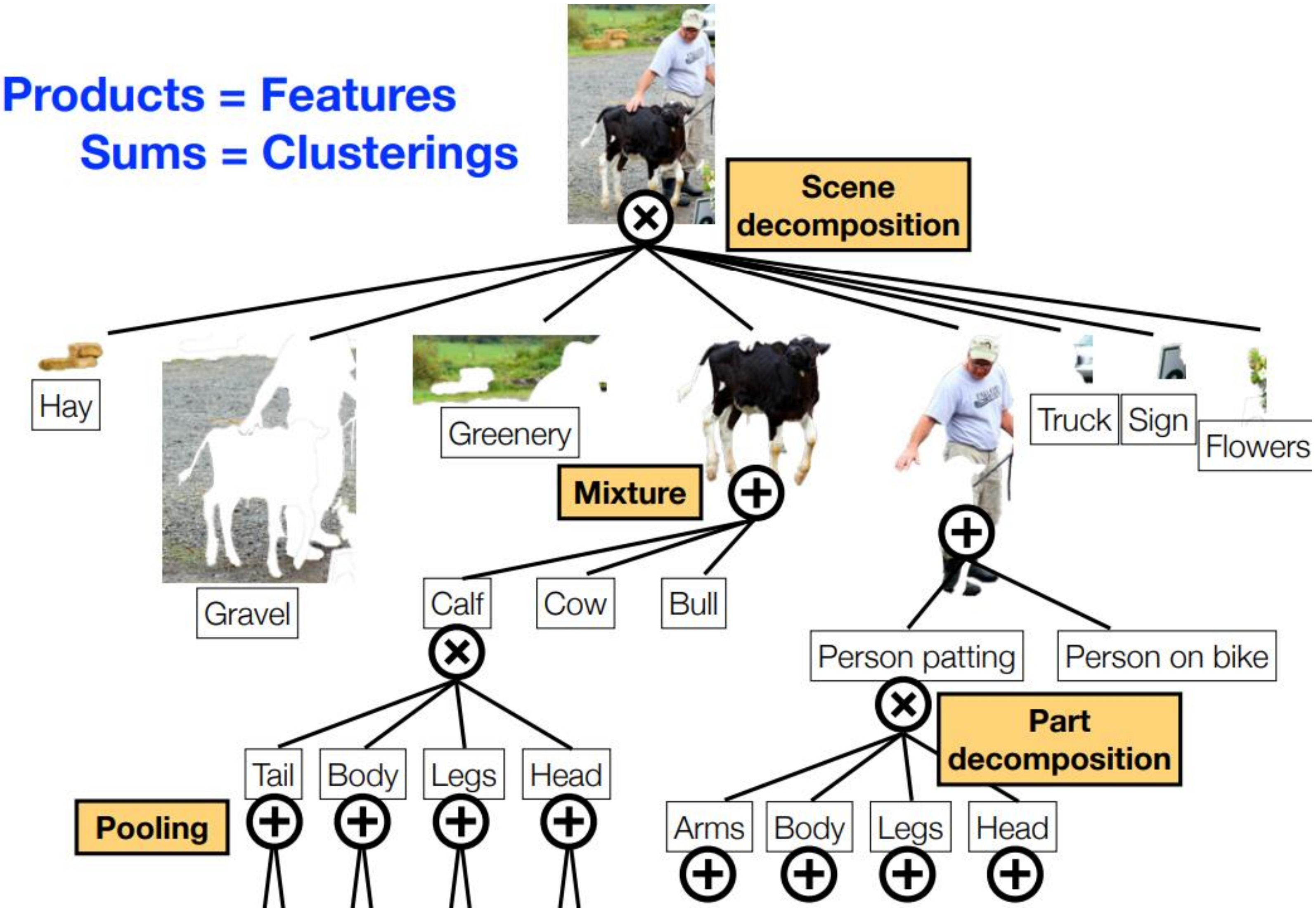
  (Mixtures of distributions)

- Decomposable: children of a product node are of different scopes

  (Distributions that factorise)



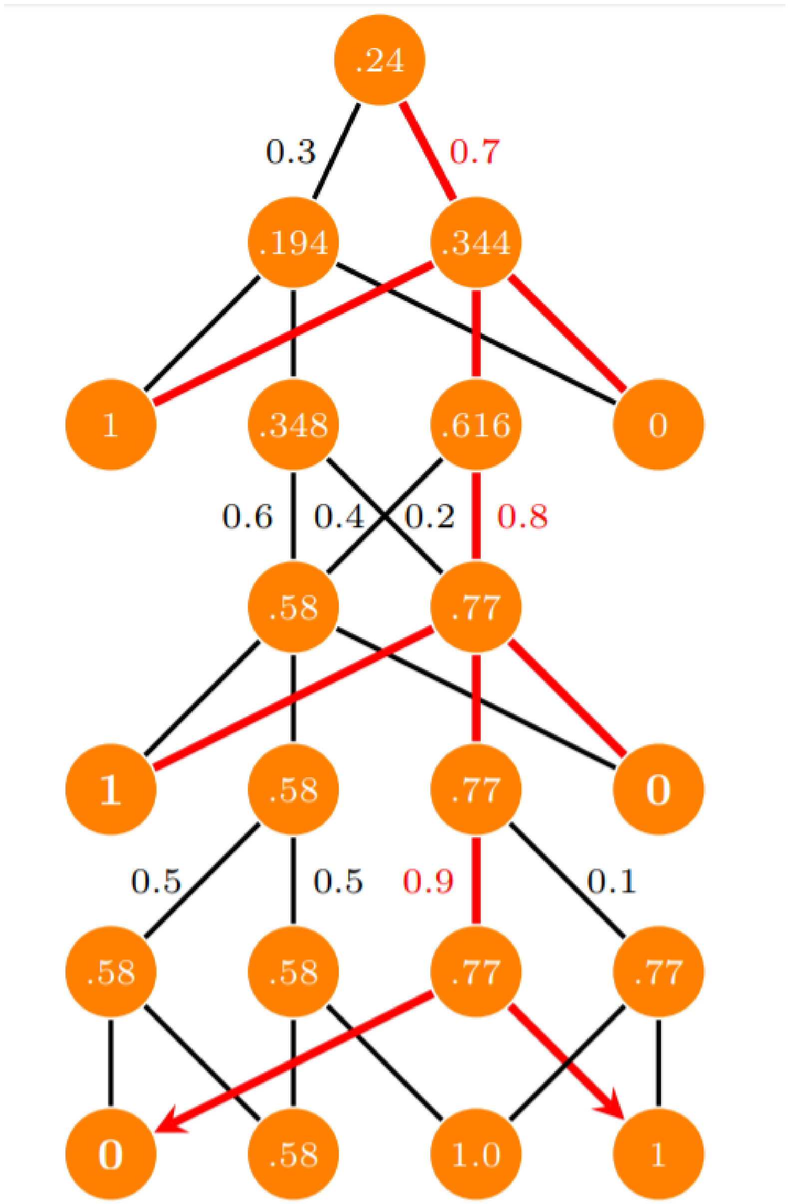
What does this mean?



Gens ICML 2013

Parameter estimation

- Lends itself naturally to backpropagation
- Vanishing gradient / gradient diffusion
- “Hard” gradient



DiMauro 2016

Parameter estimation

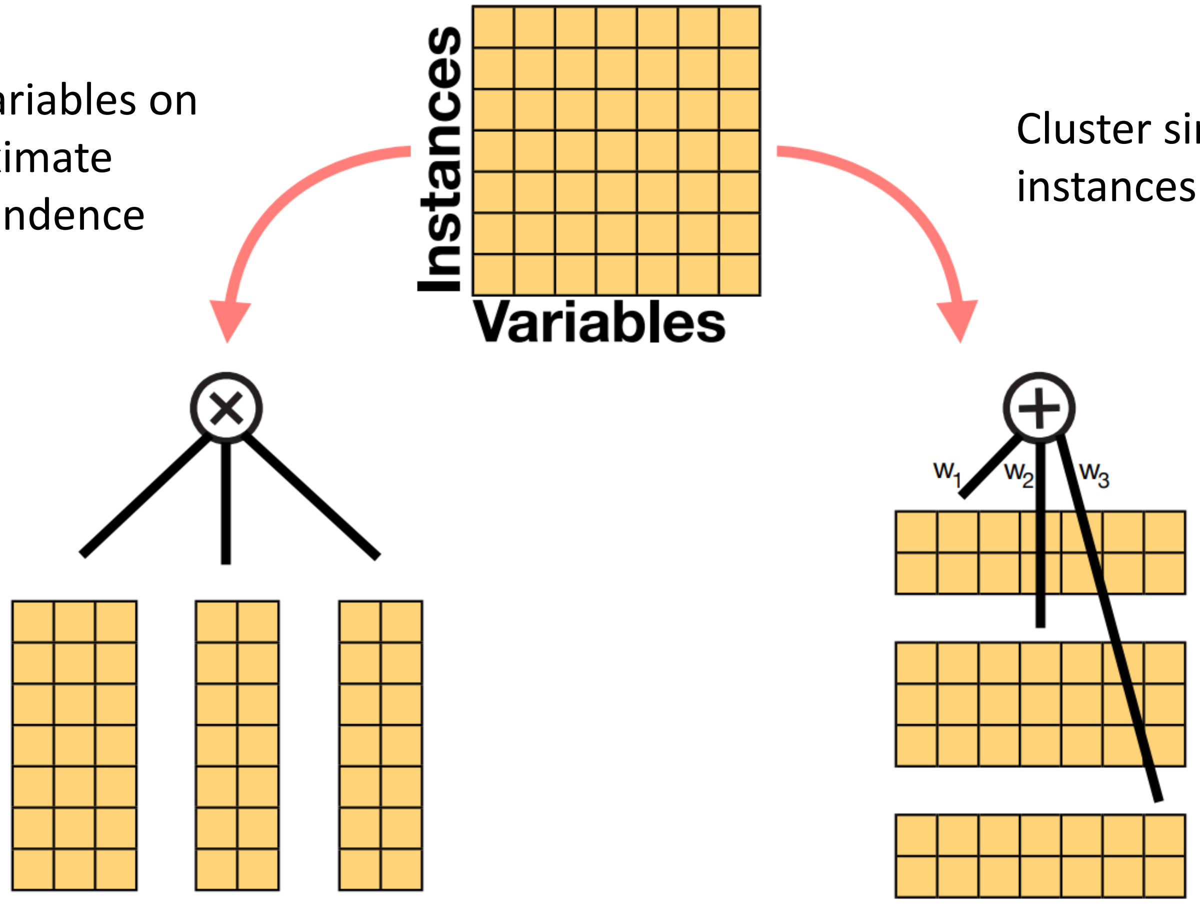
Update	Soft Inference	Hard Inference
Gen. GD	$\Delta w = \eta \frac{\partial S[\boldsymbol{x},\boldsymbol{y}]}{\partial w}$	$\Delta w_i = \eta \frac{c_i}{w_i}$
Gen. EM	$P(\mathrm{H_k} = \mathrm{i} \mid \mathbf{x}, \mathbf{y}) \propto w_{ki} \frac{\partial S[\boldsymbol{x},\boldsymbol{y}]}{\partial S_k}$	$P(\mathrm{H_k} = \mathrm{i} \mid \mathbf{x}, \mathbf{y}) = \begin{cases} 1 & : w_{ki} \in W \\ 0 & : \text{otherwise} \end{cases}$
Disc. GD	$\Delta w = \eta \left(\frac{1}{S[\boldsymbol{y},\mathbf{1}\mid\boldsymbol{x}]} \frac{\partial S[\boldsymbol{y},\mathbf{1}\mid\boldsymbol{x}]}{\partial w} - \frac{1}{S[\mathbf{1},\mathbf{1}\mid\boldsymbol{x}]} \frac{\partial S[\mathbf{1},\mathbf{1}\mid\boldsymbol{x}]}{\partial w} \right)$	$\Delta w_i = \eta \frac{\Delta c_i}{w_i}$

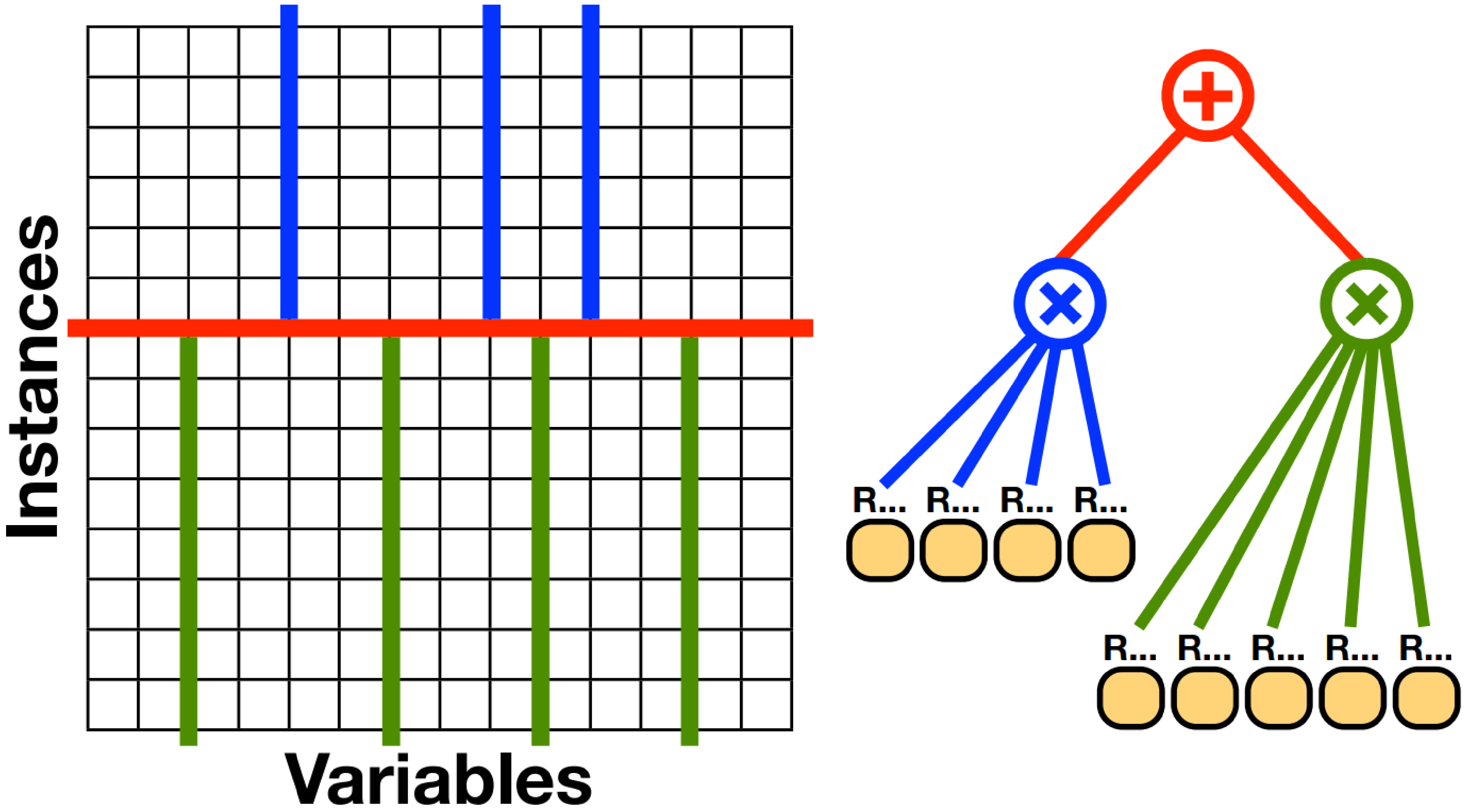
Gens & Domingos 2012

Structure learning: LearnSPN

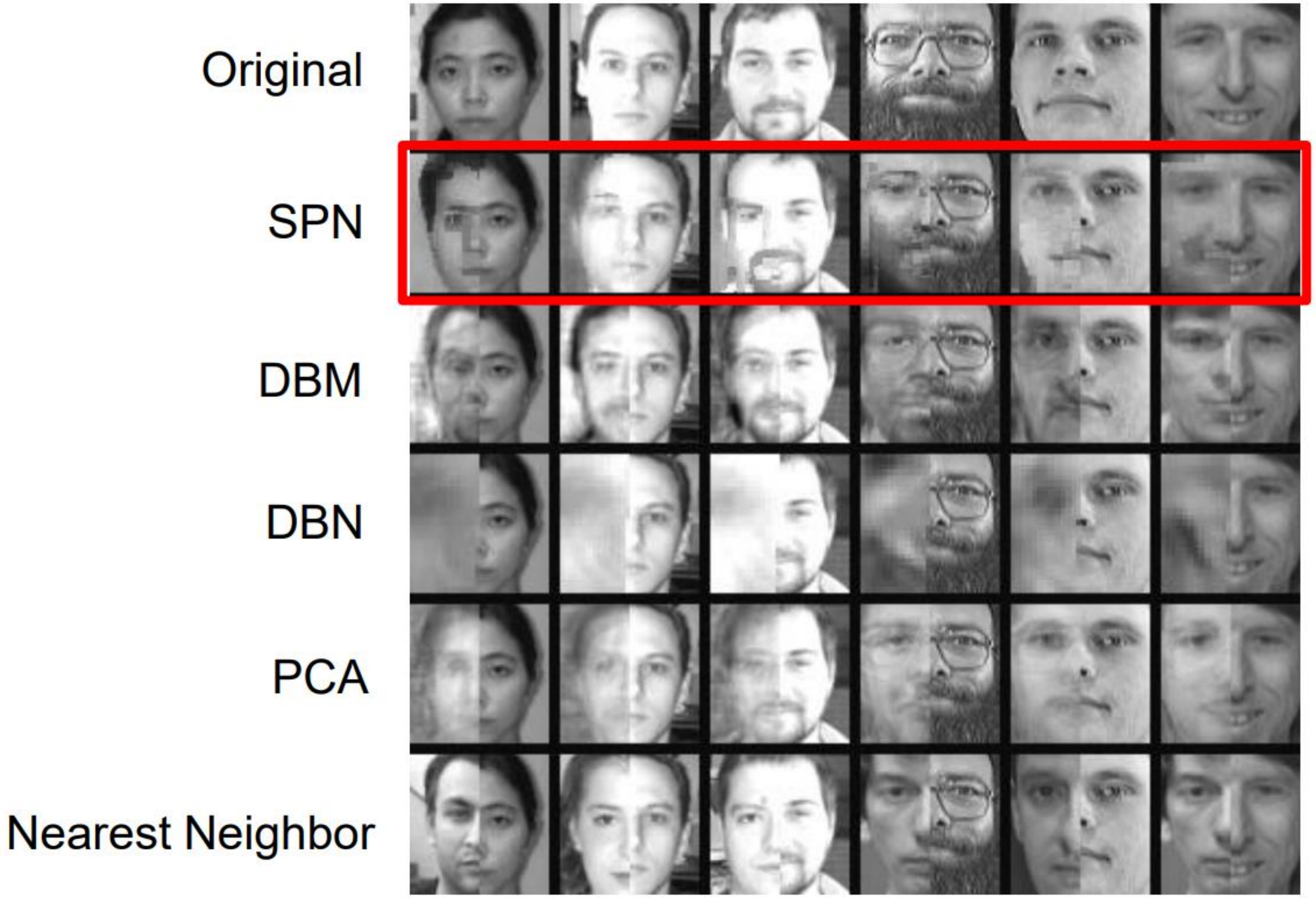
Split variables on approximate independence

Cluster similar instances




Instances
Variables
R...
R...
R...
R...
R...
R...
R...
R...
R...

Cool applications: Face completion



Poon & Domingos 2011

Cool applications

- 83.96% on CIFAR 10 Discriminative Learning of Sum-Product Networks (NIPS 2012)
- Satellite image classification

Resources

- Best place for all things SPN
    - https://github.com/arranger1044/awesome-spn

- Some video lecture sites
    - http://techtalks.tv/
    - http://videolectures.net/

- My Github
    - https://github.com/felixmcgregor/Sum-Product-Networks