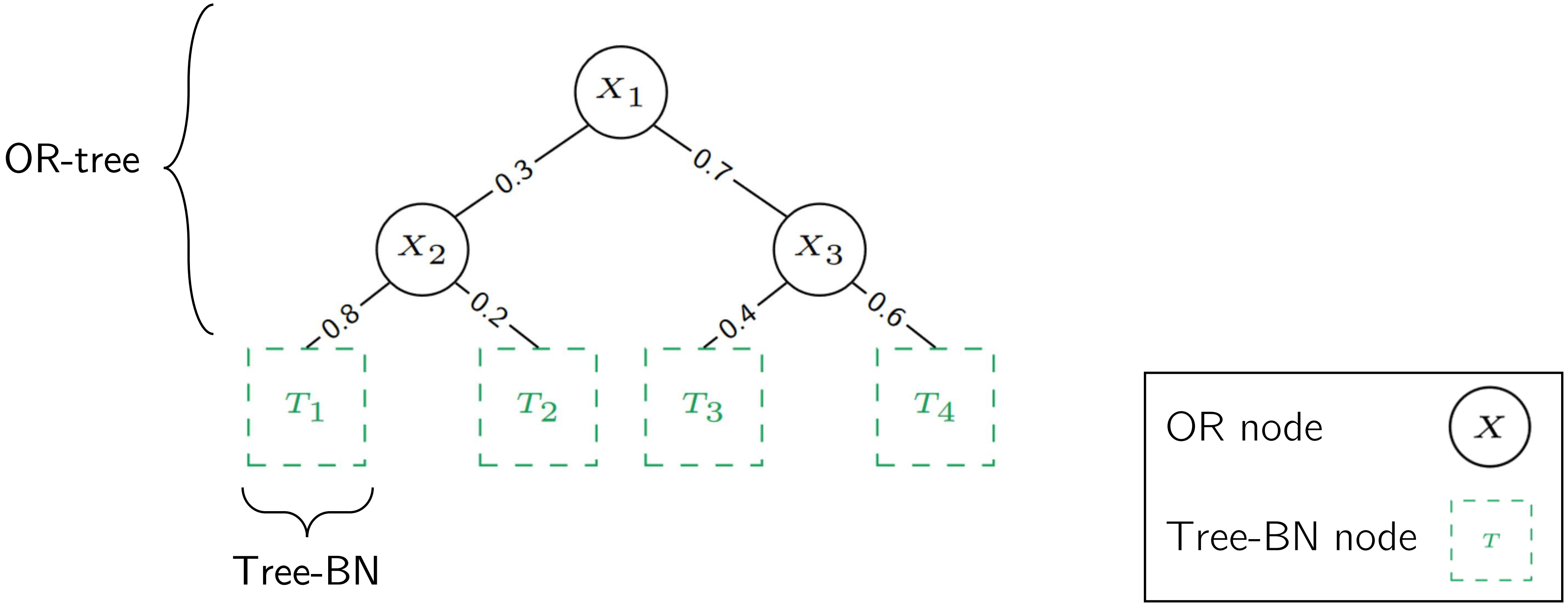
Cutset Networks

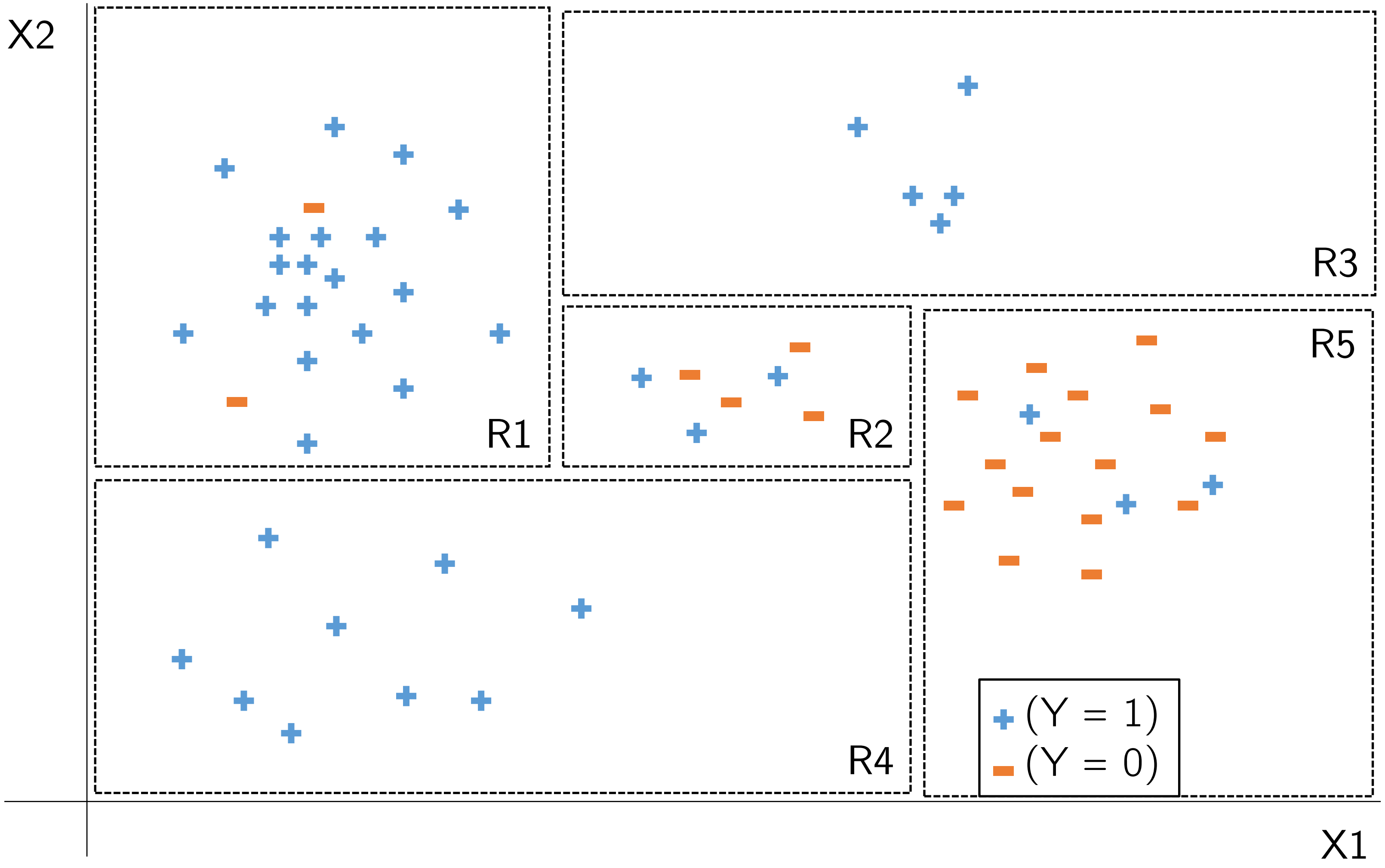
Day 2

Cutset Networks



Rahman, Tahrima, Prasanna Kothalkar, and Vibhav Gogate. "Cutset networks: A simple, tractable, and scalable approach for improving the accuracy of chow-liu trees." ECML PKDD 2014,

Intuition: trees divide space into rectangles



Cutset Networks: Outline

1. Representation

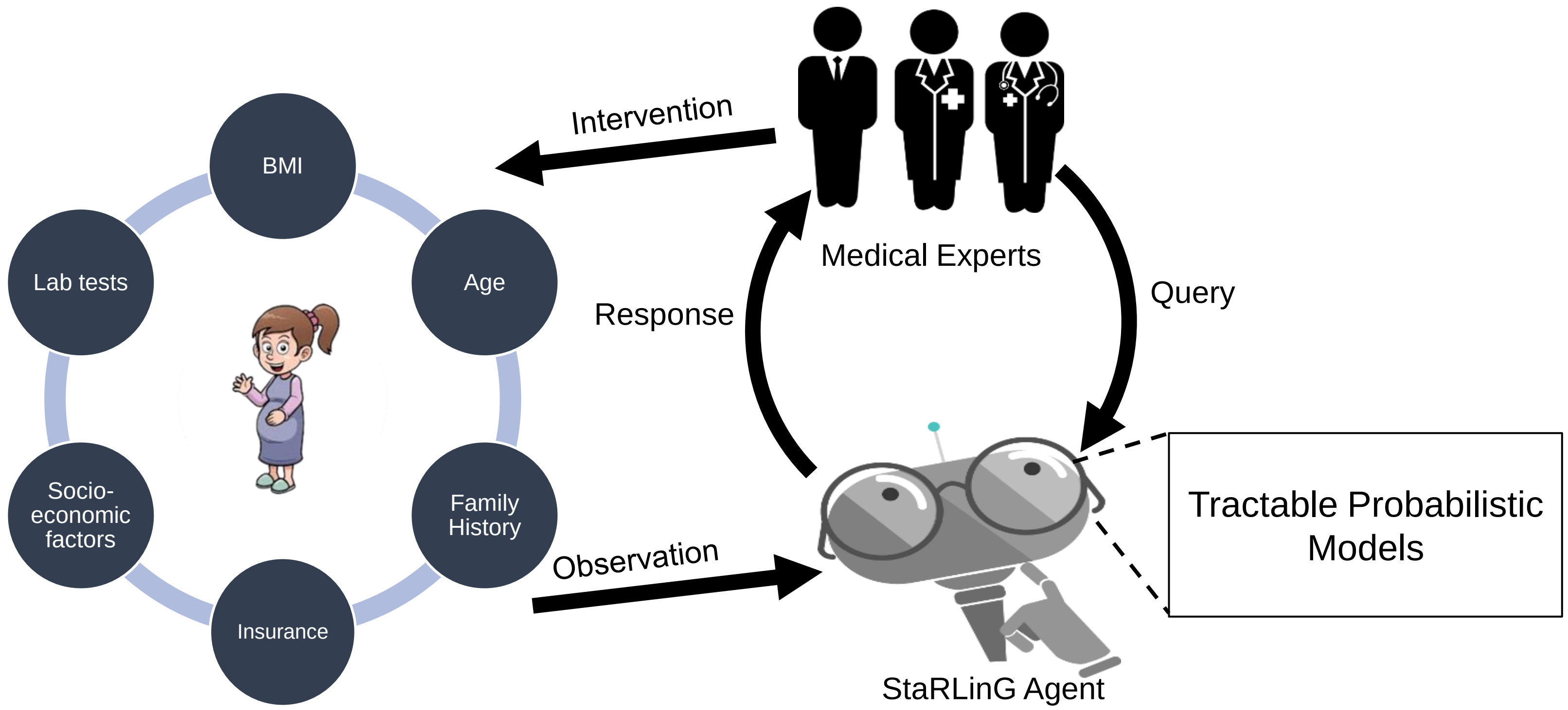
2. Inference

3. Learning

Cutset Networks: Outline

1. **Representation**
2. Inference
3. Learning

Mitigating adverse pregnancy outcomes



Probability table for Gestational diabetes and risk factors

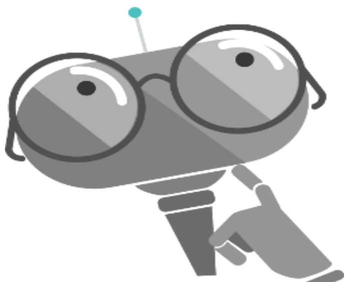
Given,

$$\mathbf{P}(\mathrm{X1} = \mathrm{x1},\ \mathrm{X2} = \mathrm{x2},\ \mathrm{X3} = \mathrm{x3})$$

$$\mathrm{X1},\ \mathrm{X2},\ \mathrm{X3} \in \{0,\ 1\}^3$$

To Do,

Answer queries



StaRLinG Agent

Gestational diabetes — High BP — PCOS

X1	X2	X3	P
0	0	0	p1
0	0	1	p2
0	1	0	p3
0	1	1	p4
1	0	0	p5
1	0	1	p6
1	1	0	p7
1	1	1	p8

Probability table for Gestational diabetes and risk factors

"What is the probability of all risk factors and gestational diabetes?"

$$\boxed{\mathbf{P}(\mathrm{X1} = 1, \mathrm{X2} = 1, \mathrm{X3} = 1)}$$

$$= \mathrm{p8}$$

X1	X2	X3	P
0	0	0	p1
0	0	1	p2
0	1	0	p3
0	1	1	p4
1	0	0	p5
1	0	1	p6
1	1	0	p7
1	1	1	p8

Problem: Lookup time for one entry is $\mathrm{O}(2^n)$

Gestational diabetes — High BP — PCOS

X1	X2	X3	P
0	0	0	p1
0	0	1	p2
0	1	0	p3
0	1	1	p4
1	0	0	p5
1	0	1	p6
1	1	0	p7
1	1	1	p8

n, number of variables

Solution: Tree. Lookup time is $\mathrm{O}(\log 2^n) = \mathrm{O}(n)$

Gestational diabetes   High BP   PCOS

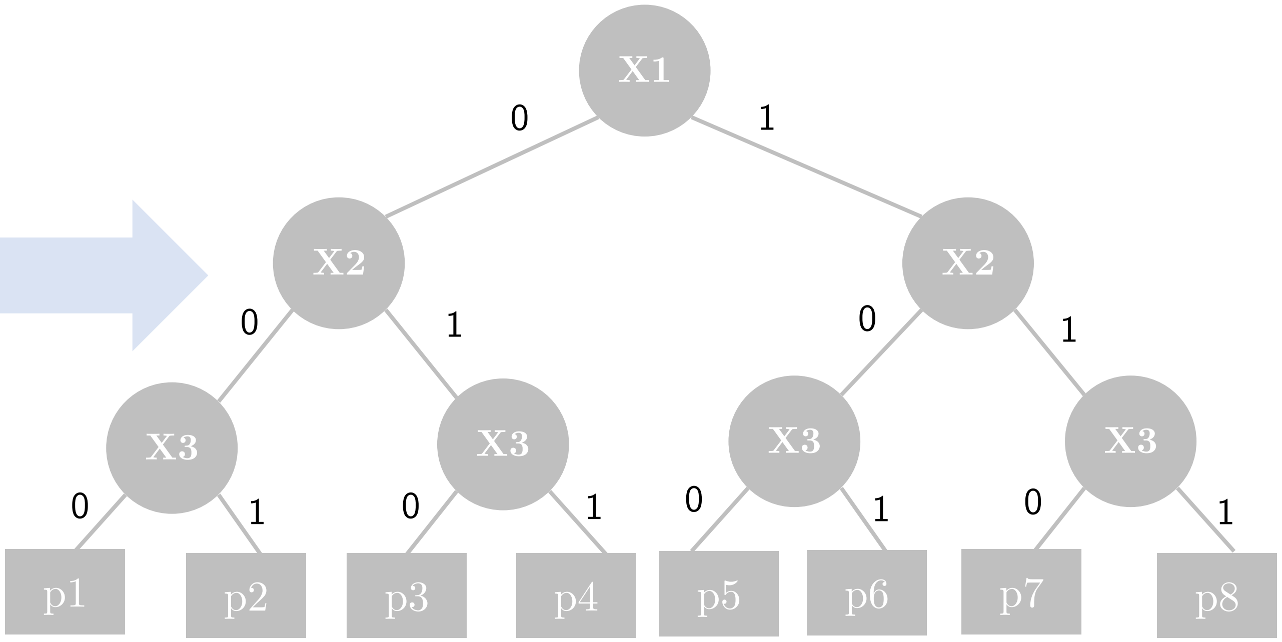
X1	X2	X3	P
0	0	0	p1
0	0	1	p2
0	1	0	p3
0	1	1	p4
1	0	0	p5
1	0	1	p6
1	1	0	p7
1	1	1	p8

n, number of variables

Solution: Tree. Lookup time is $\mathrm{O}(\log\, 2^n) = \mathrm{O}(n)$

Gestational diabetes High BP PCOS

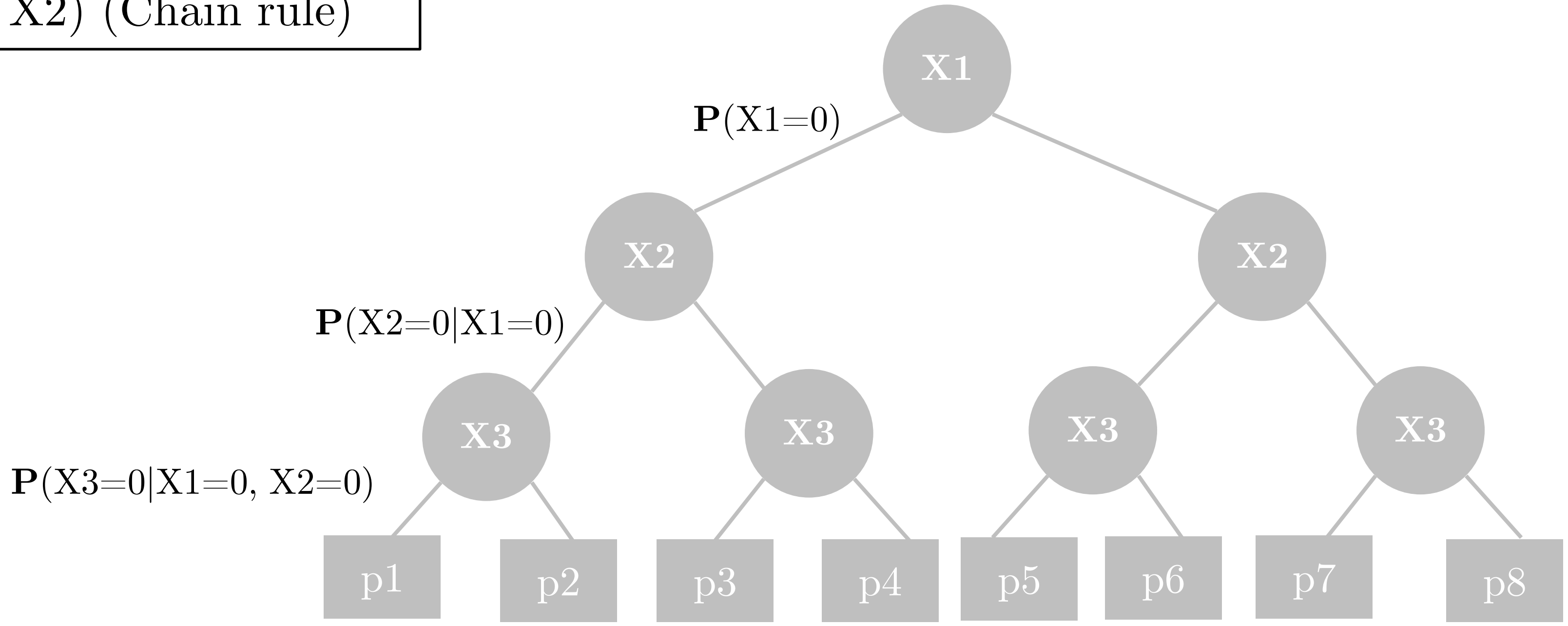
X1	X2	X3	P
0	0	0	p1
0	0	1	p2
0	1	0	p3
0	1	1	p4
1	0	0	p5
1	0	1	p6
1	1	0	p7
1	1	1	p8



n, number of variables

OR-tree: Joint Probability Tree with edge labels

$\mathbf{P}(\text{X1, X2, X3})$
$= \mathbf{P}(\text{X1}) \cdot \mathbf{P}(\text{X1}|\text{X2}) \cdot \mathbf{P}(\text{X3}|\text{X1, X2})$ (Chain rule)



n, number of variables

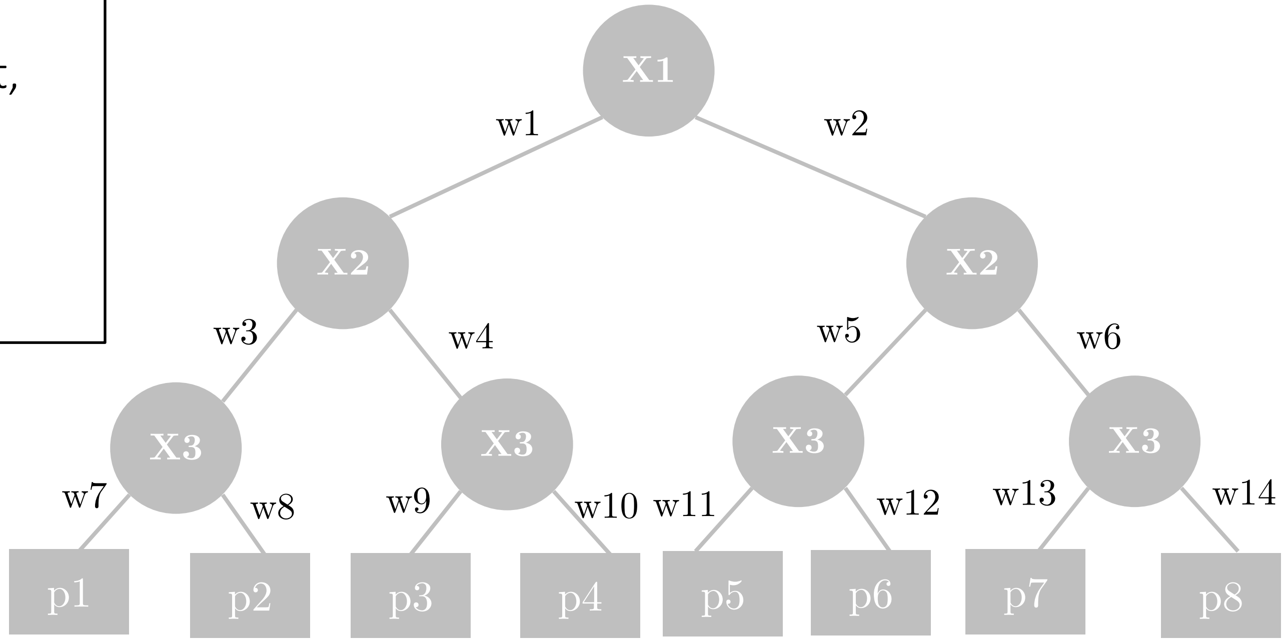
OR-tree: Joint Probability Tree with edge labels

OR-tree $\mathrm{O} = (\mathrm{E},\, \mathrm{w})$ where,

E is a set of edges

$w : \mathrm{E} \mapsto (0,\, 1)$ is the edge weight function such that,

$$\forall\, u, \sum_{(u,v)\in E} w(u,v) = 1$$



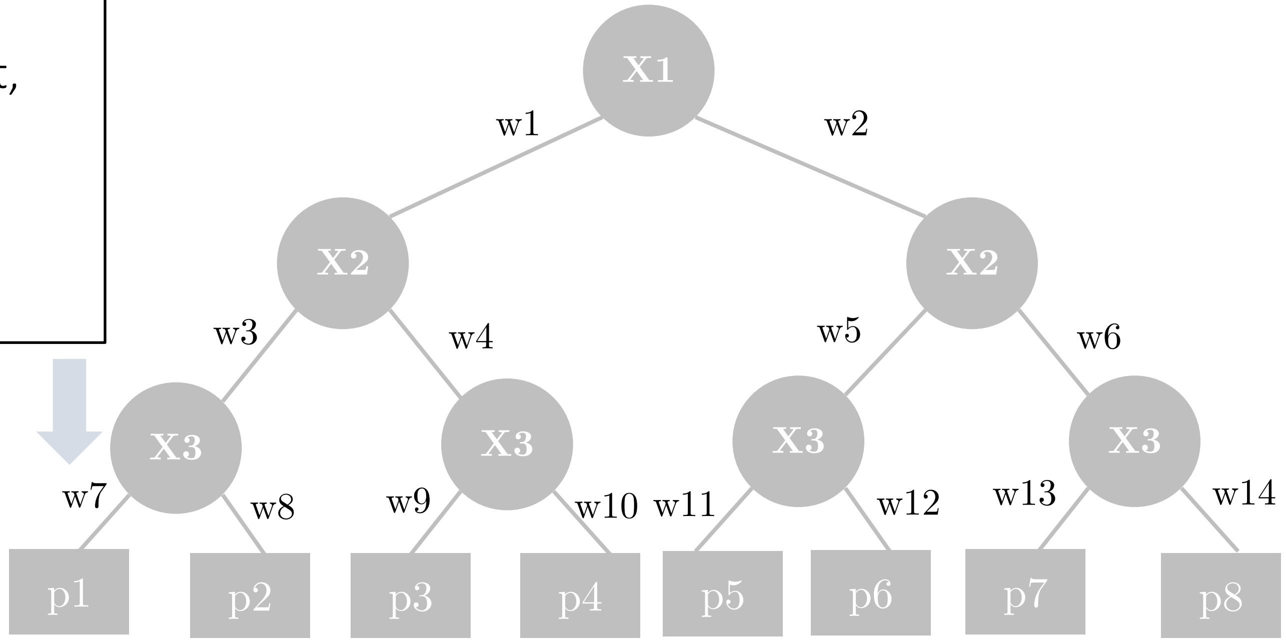
OR-tree: Joint Probability Tree with edge labels

OR-tree $\mathrm{O} = (\mathrm{E},\, \mathrm{w})$ where,

E is a set of edges

$w : \mathrm{E} \mapsto (0,\, 1)$ is the edge weight function such that,

$$\forall\, u,\ \sum_{(u,v)\in E} w(u,v) = 1$$

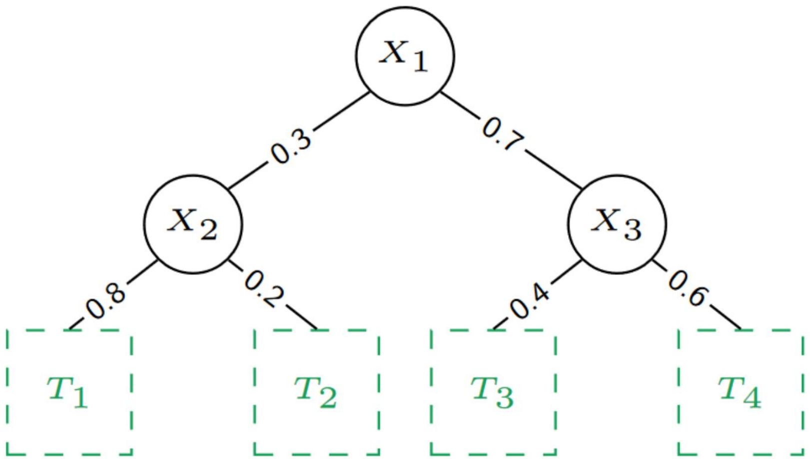


$$\mathrm{w7} = \mathrm{P(X3{=}0 \mid X1{=}0,\, X2{=}0)} = \frac{Count(X1{=}0, X2{=}0, X3{=}0)}{Count(X1{=}0, X3{=}0)}$$

$$(\mathrm{X1},\ \mathrm{X2},\ \mathrm{X3},\ \underbrace{\mathrm{X4},\ \mathrm{X5}}_{\text{Additional risk factors}})$$

Cutset network $\mathcal{M}$ of depth 2 over $(\mathrm{X1, X2, X3, X4, X5})$

OR-tree

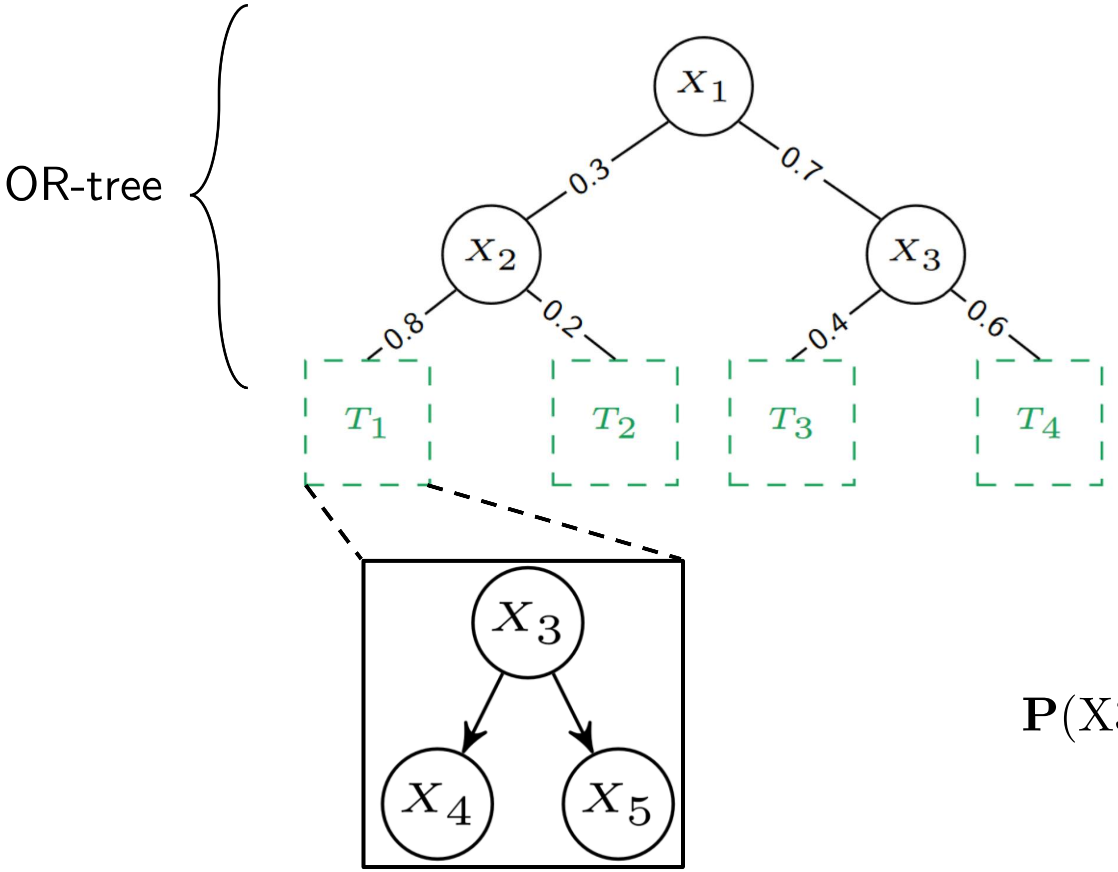


$\mathcal{M} = (\mathrm{O, T})$ where,

O is a rooted OR-Tree

T is a set of Tree-BNs

Cutset network $\mathcal{M}$ of depth 2 over $(\mathrm{X1}, \mathrm{X2}, \mathrm{X3}, \mathrm{X4}, \mathrm{X5})$



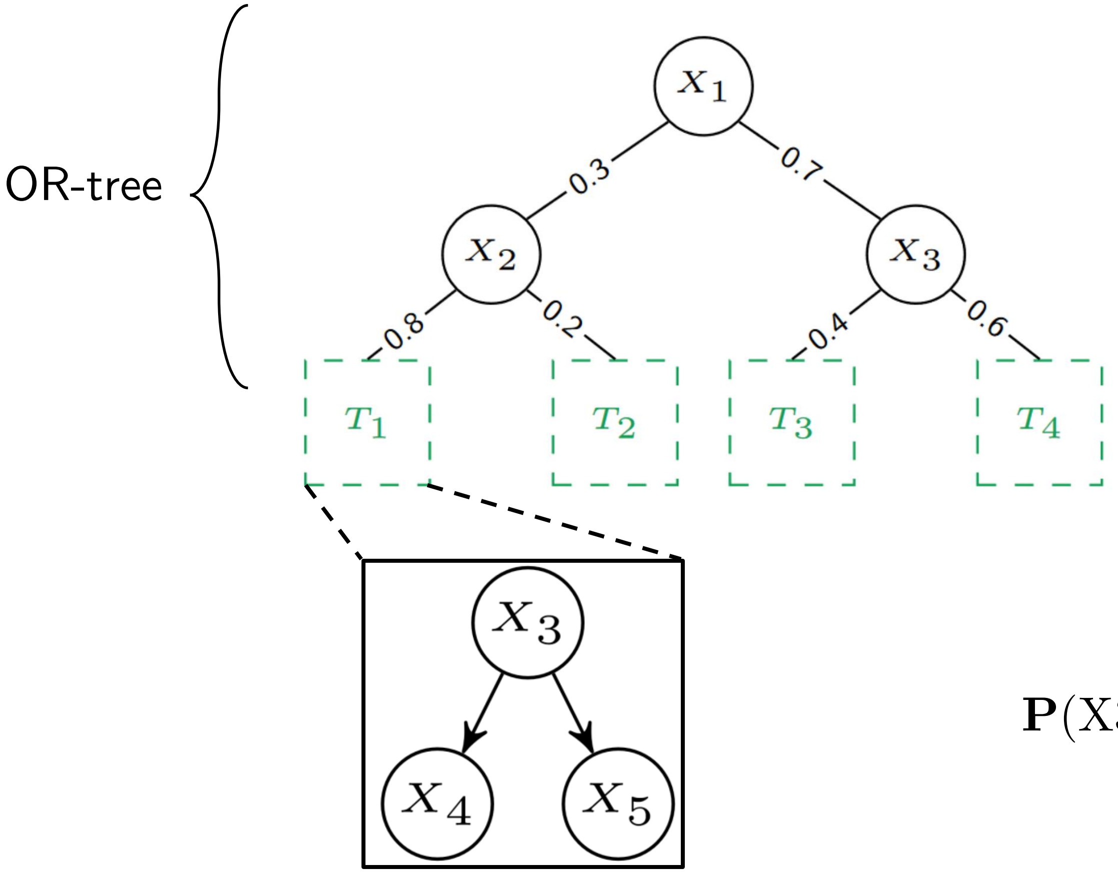
$\mathcal{M} = (\mathrm{O}, \mathrm{T})$ where,

O is a rooted OR-Tree

T is a set of Tree-BNs

$$\mathbf{P}(\mathrm{X3,X4,X5} \mid \mathrm{X1{=}0,X2{=}0}) = \mathbf{P}(\mathrm{X3} \mid \mathrm{X1{=}0,X2{=}0}) \cdot \mathbf{P}(\mathrm{X4} \mid \mathrm{X3,X1{=}0,X2{=}0}) \cdot \mathbf{P}(\mathrm{X5} \mid \mathrm{X3,X1{=}0,X2{=}0})$$

Cutset network $\mathcal{M}$ of depth 2 over $(\mathrm{X1}, \mathrm{X2}, \mathrm{X3}, \mathrm{X4}, \mathrm{X5})$

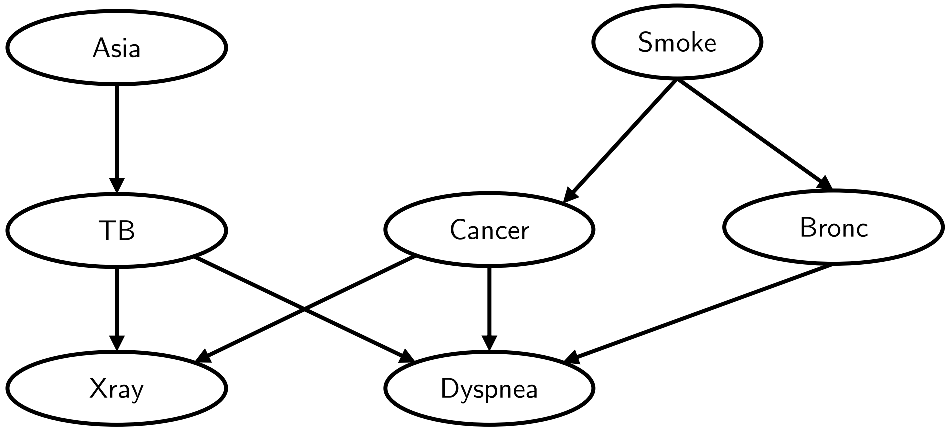


Context-specific independence:

$$\mathrm{X4} \perp\!\!\!\perp \mathrm{X5} \mid \mathrm{X3} \text{ If } (\mathrm{X1} = 0) \wedge (\mathrm{X2} = 0)$$

$$\begin{aligned}\mathbf{P}(\mathrm{X3,X4,X5} \mid \mathrm{X1{=}0,X2{=}0}) = {} & \mathbf{P}(\mathrm{X3} \mid \mathrm{X1{=}0,X2{=}0}) \cdot \\ & \mathbf{P}(\mathrm{X4} \mid \mathrm{X3,X1{=}0,X2{=}0}) \cdot \\ & \mathbf{P}(\mathrm{X5} \mid \mathrm{X3,X1{=}0,X2{=}0})\end{aligned}$$

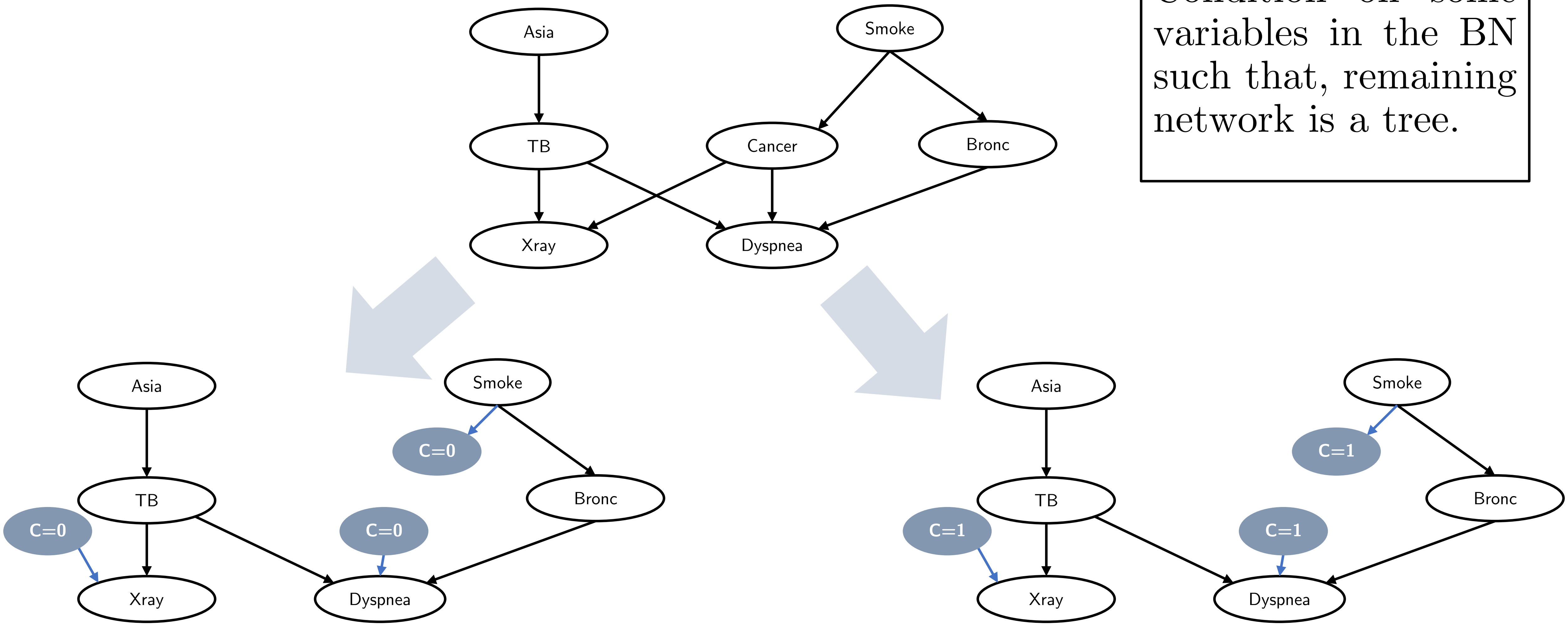
Connection to cutset conditioning



Condition on some variables in the BN such that, remaining network is a tree.

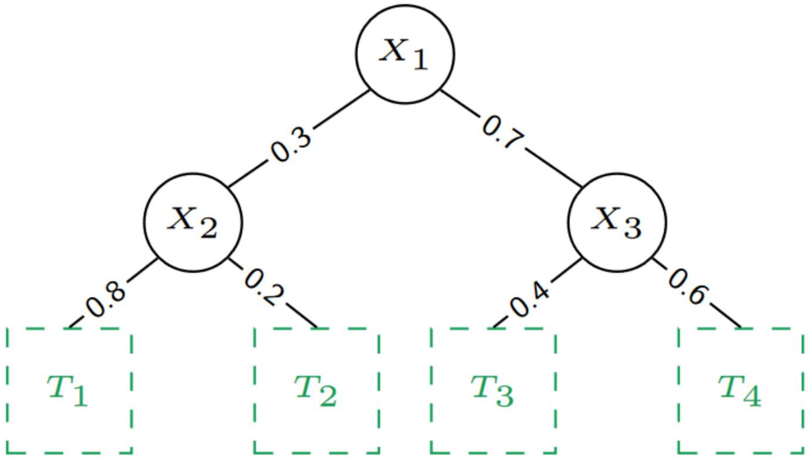
Robert Mateescu and Rina Dechter. 2005. AND/OR cutset conditioning. IJCAI (2005)

Condition on $Cancer \in \{0, 1\}$



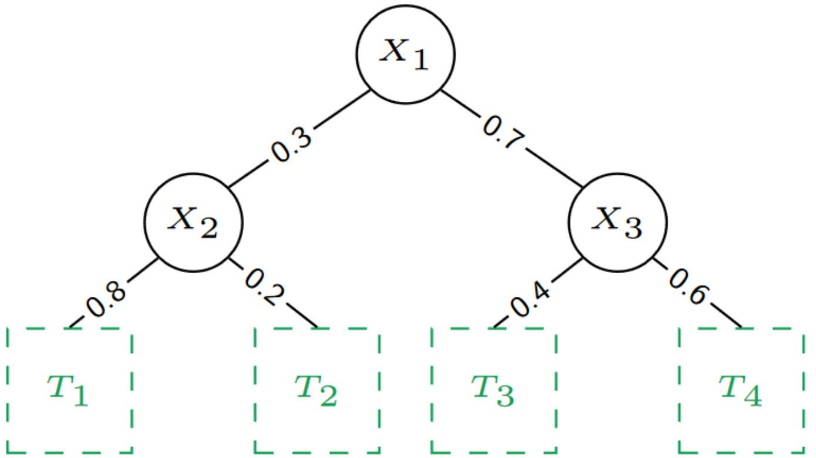
Condition on some variables in the BN such that, remaining network is a tree.

Cutset network $\mathcal{M}$ of depth 2 over $(\mathrm{X1},\ \mathrm{X2},\ \mathrm{X3},\ \mathrm{X4},\ \mathrm{X5})$



$$= \begin{cases} 0.3 \cdot 0.8 \cdot \mathbf{P}(\mathrm{X3},\ \mathrm{X4},\ \mathrm{X5} \mid \mathrm{X1} = 0,\ \mathrm{X2} = 0) & \text{If X1} = 0,\ \mathrm{X2} = 0 \\ 0.3 \cdot 0.2 \cdot \mathbf{P}(\mathrm{X3},\ \mathrm{X4},\ \mathrm{X5} \mid \mathrm{X1} = 0,\ \mathrm{X2} = 1) & \text{If X1} = 0,\ \mathrm{X2} = 1 \\ 0.7 \cdot 0.4 \cdot \mathbf{P}(\mathrm{X2},\ \mathrm{X4},\ \mathrm{X5} \mid \mathrm{X1} = 1,\ \mathrm{X3} = 0) & \text{If X1} = 1,\ \mathrm{X3} = 0 \\ 0.7 \cdot 0.6 \cdot \mathbf{P}(\mathrm{X3},\ \mathrm{X4},\ \mathrm{X5} \mid \mathrm{X1} = 1,\ \mathrm{X3} = 1) & \text{If X1} = 1,\ \mathrm{X3} = 1 \end{cases}$$

Cutset network $\mathcal{M}$ of depth 2 over $(\mathrm{X1, X2, X3, X4, X5})$

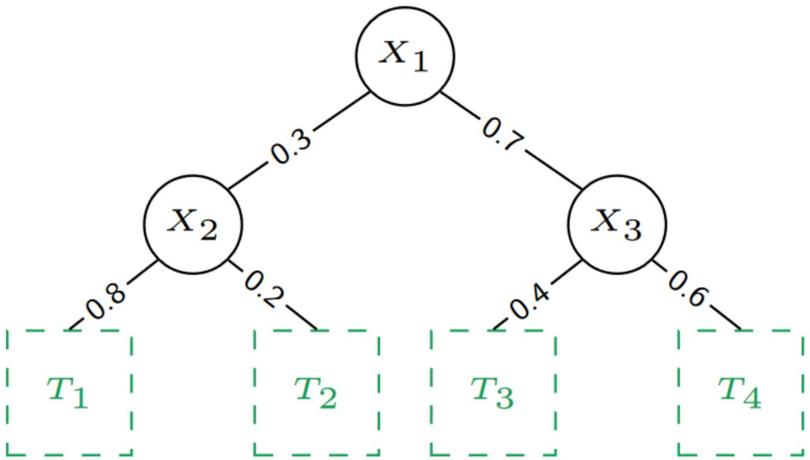


$$= \begin{cases} 0.3 \cdot 0.8 \cdot \mathbf{P}(\mathrm{X3, X4, X5} \mid \mathrm{X1} = 0, \mathrm{X2} = 0) & \text{If } \mathrm{X1} = 0, \mathrm{X2} = 0 \\ 0.3 \cdot 0.2 \cdot \mathbf{P}(\mathrm{X3, X4, X5} \mid \mathrm{X1} = 0, \mathrm{X2} = 1) & \text{If } \mathrm{X1} = 0, \mathrm{X2} = 1 \\ 0.7 \cdot 0.4 \cdot \mathbf{P}(\mathrm{X2, X4, X5} \mid \mathrm{X1} = 1, \mathrm{X3} = 0) & \text{If } \mathrm{X1} = 1, \mathrm{X3} = 0 \\ 0.7 \cdot 0.6 \cdot \mathbf{P}(\mathrm{X3, X4, X5} \mid \mathrm{X1} = 1, \mathrm{X3} = 1) & \text{If } \mathrm{X1} = 1, \mathrm{X3} = 1 \end{cases}$$

Network polynomial:

$$\boxed{P(x) = \underbrace{\left(\prod_{(v_i, v_j) \in path_O(x)} w(v_i, v_j) \right)}_{\text{OR-tree}} \underbrace{\left(T_{l(x)}(x_{V(T_{l(x)})}) \right)}_{\text{Tree-BN}}}$$

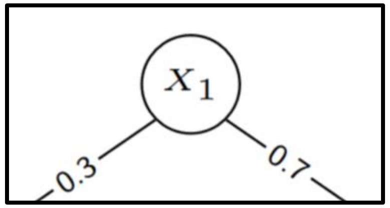
Cutset networks are Deterministic

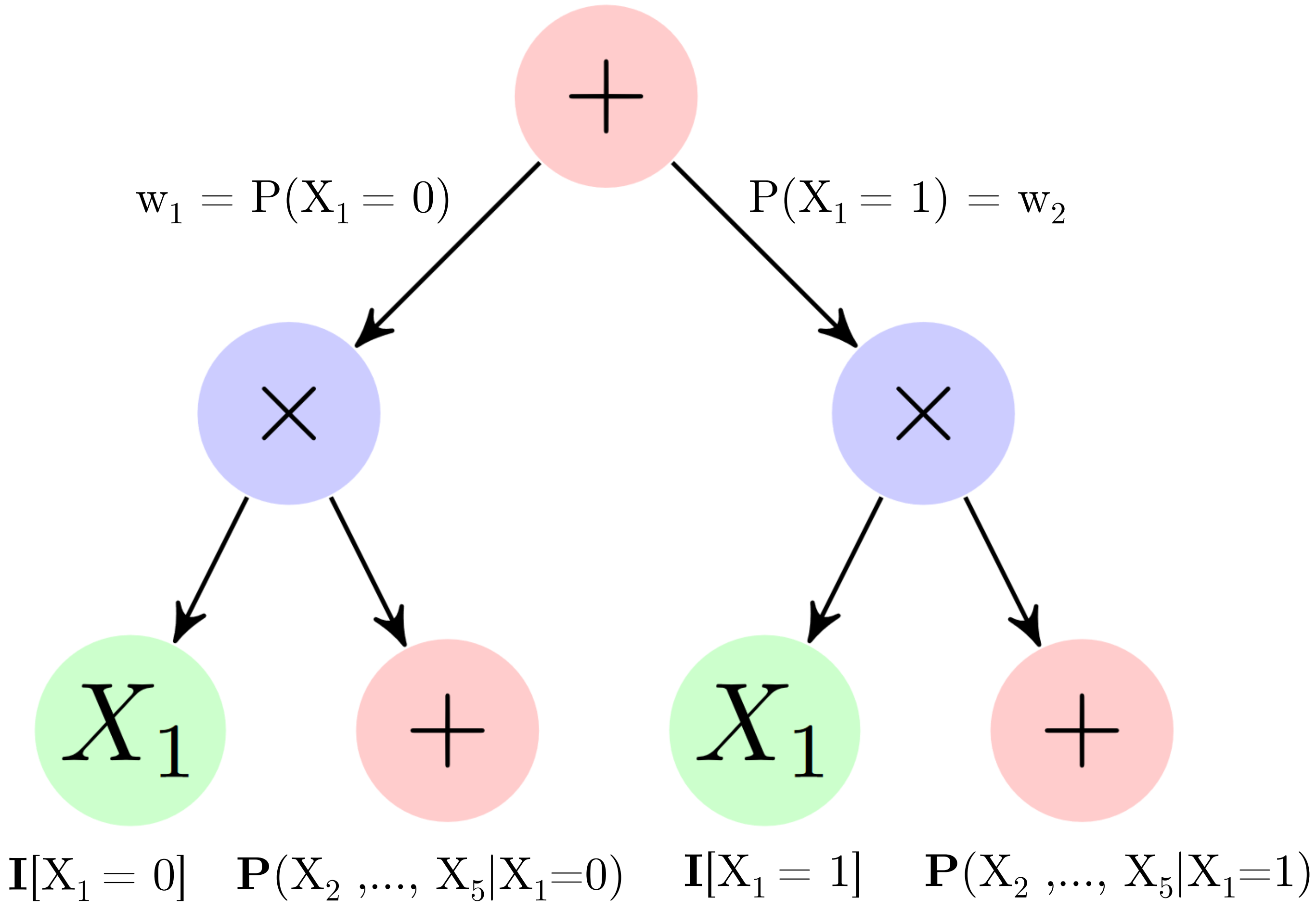


A node n is deterministic if, for any fully-instantiated input, the output of **at most one of its children** is nonzero.

A circuit is deterministic if all its nodes are deterministic.

OR-node as deterministic SPN





Cutset Networks: Outline

1. Representation

2. **Inference**

3. Learning

Answering queries using probability table

"What is the probability of High BP?"

$\mathbf{P}(\mathrm{X2} = 1)$

X1	X2	X3	P
0	0	0	p1
0	0	1	p2
0	1	0	p3
0	1	1	p4
1	0	0	p5
1	0	1	p6
1	1	0	p7
1	1	1	p8

Answering queries using probability table

"What is the probability of High BP?"

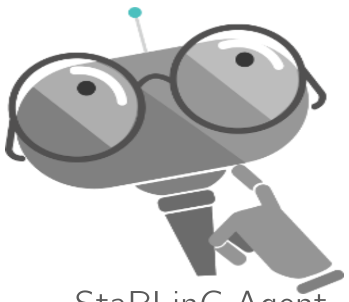
$$\mathbf{P}(X2 = 1)$$

$$\sum_{x1,x3} \mathbf{P}(X1 = x1, X2 = 1, X3 = x3)$$

X1	X2	X3	P
0	0	0	p1
0	0	1	p2
0	1	0	p3
0	1	1	p4
1	0	0	p5
1	0	1	p6
1	1	0	p7
1	1	1	p8

Answering queries using probability table

"What is the probability of High BP?"



$\mathbf{P}(X2 = 1)$

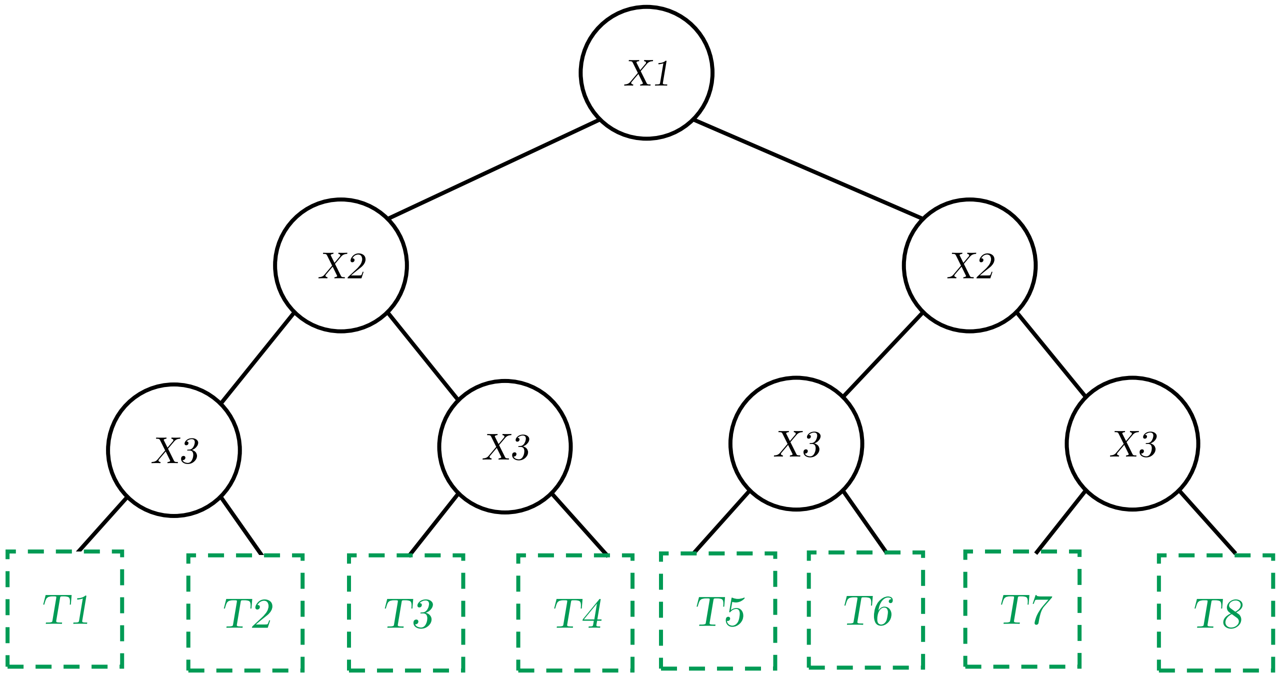
$= p3 + p4 + p7 + p8$

X1	X2	X3	P
0	0	0	p1
0	0	1	p2
0	1	0	p3
0	1	1	p4
1	0	0	p5
1	0	1	p6
1	1	0	p7
1	1	1	p8

Answering queries using Cutset Network (MAR)

"What is the probability of High BP?"

$\mathbf{P}(X2 = 1)$

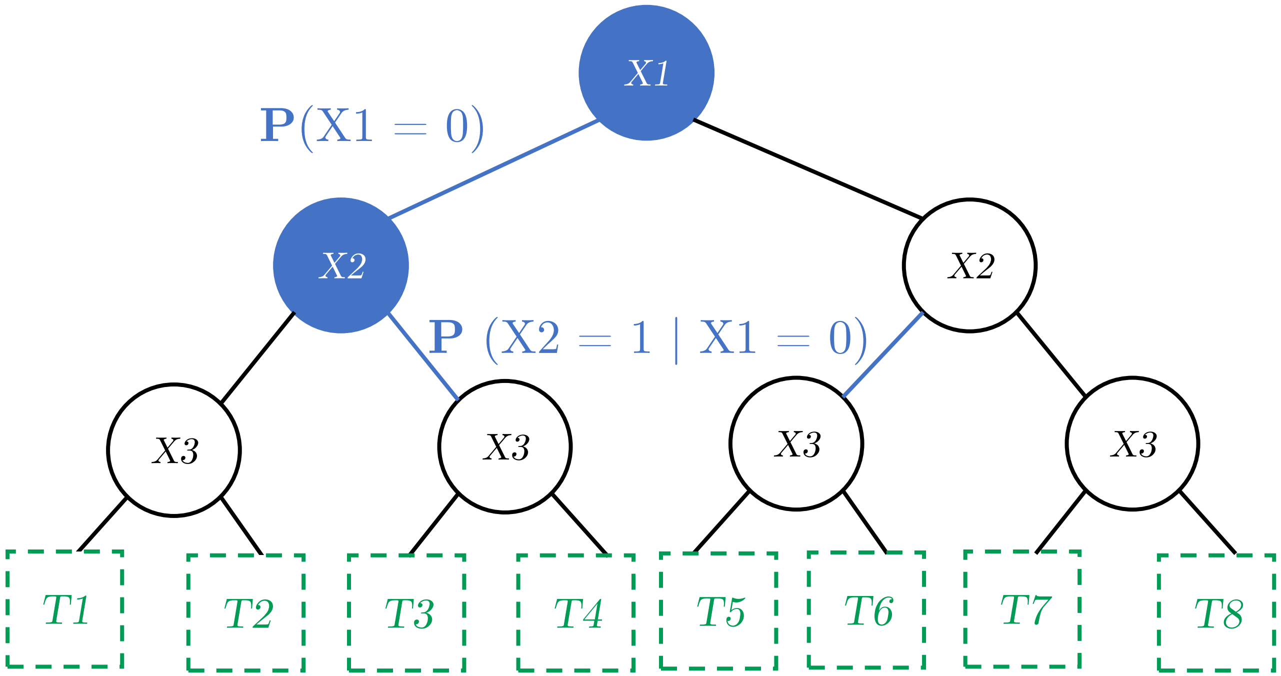


Answering queries using Cutset Network (MAR)

"What is the probability of High BP?"

StaRLinG Agent

$\mathbf{P}(\mathrm{X2} = 1)$



$= \mathbf{P}(\mathrm{X1} = 0) \cdot \mathbf{P}\ (\mathrm{X2} = 1 \mid \mathrm{X1} = 0)$

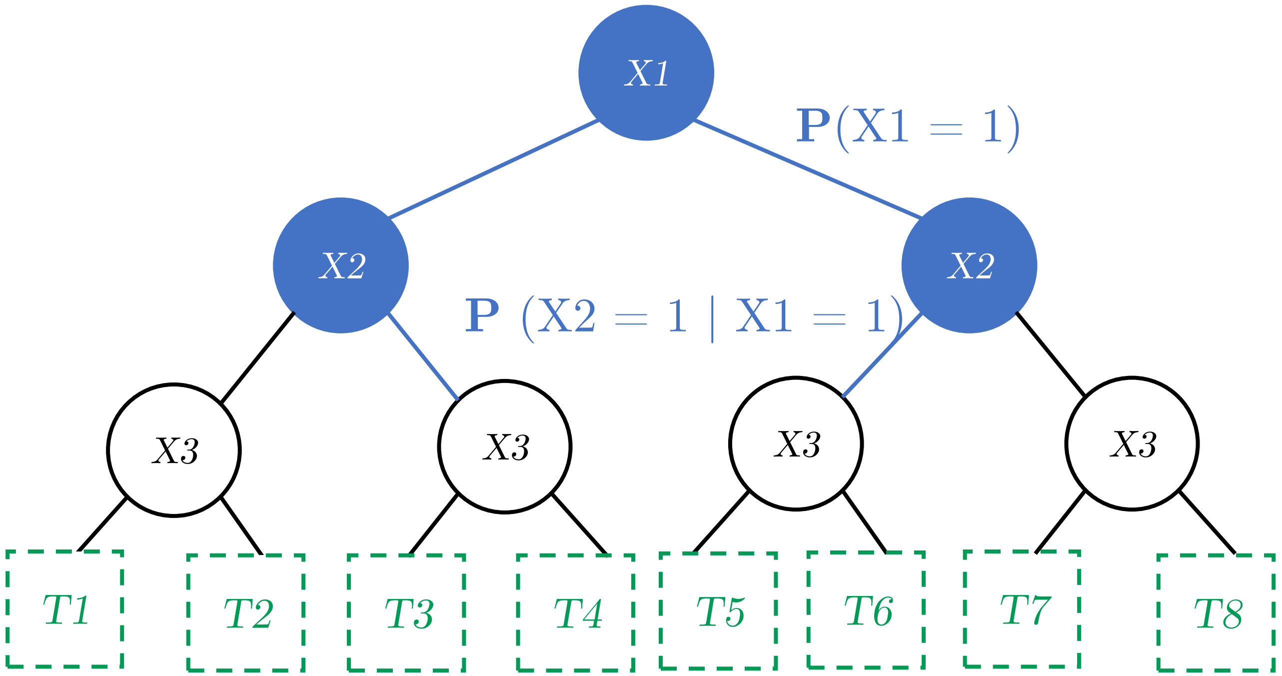
$+$

Answering queries using Cutset Network (MAR)

“What is the probability of High BP?”

$\mathbf{P}(\mathrm{X2} = 1)$

$= \mathbf{P}(\mathrm{X1} = 0) \cdot \mathbf{P}\ (\mathrm{X2} = 1 \mid \mathrm{X1} = 0)$
$+ \mathbf{P}(\mathrm{X1} = 1) \cdot \mathbf{P}\ (\mathrm{X2} = 1 \mid \mathrm{X1} = 1)$

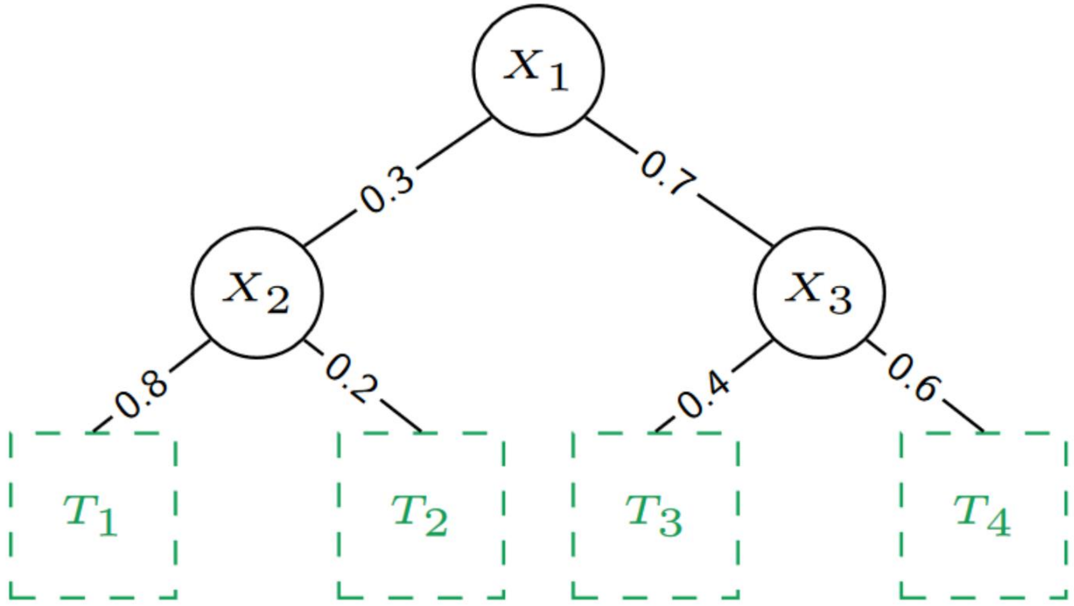


EVI/MAR inference

Given: query Q; $X_Q \subseteq \mathbf{X}$

To Do: Find $\mathbf{P}(X_Q = x_Q)$

1. Start at root.
2. Traverse the network depth-first.
3. If current node is query variable,
   Select child based on value.
4. Otherwise,
   Take sum of query over all children.

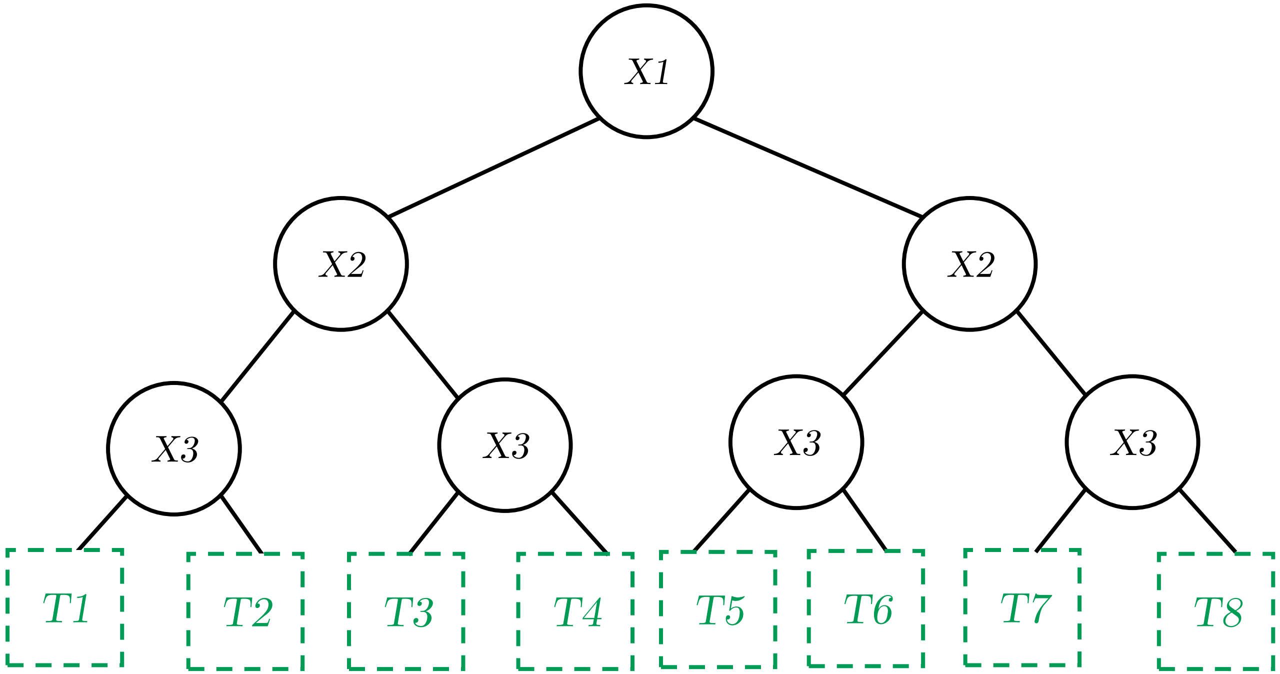


Answering queries using Cutset Network (MAP)

"What is the most likely diagnosis for a subject with all risk factors?"

StaRLinG Agent

$\arg\max \mathbf{P}(X1{=}x1 \mid X2{=}1, \ldots, X5{=}1)$



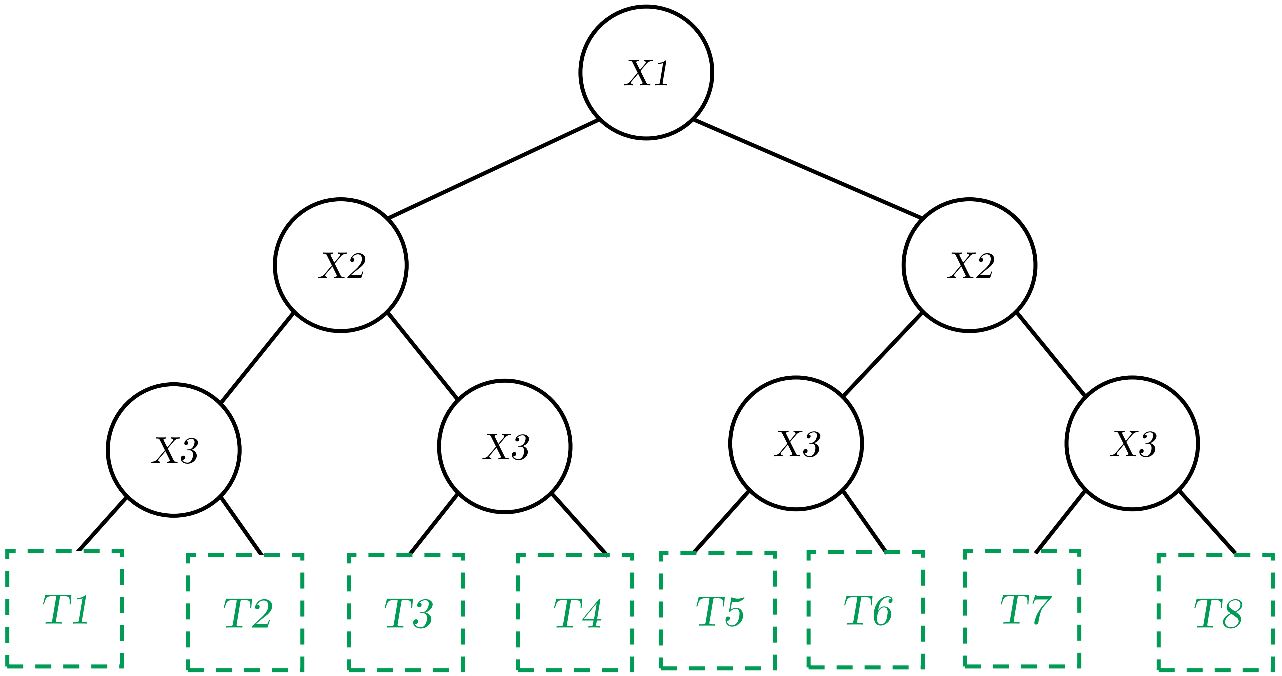
Answering queries using Cutset Network (MAP)

“What is the most likely diagnosis for a subject with all risk factors?”

StaRLinG Agent

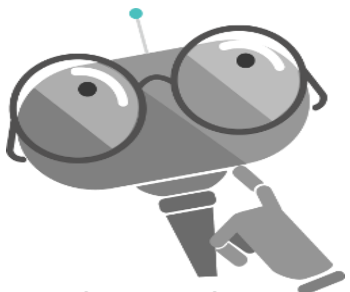
$$\boxed{\arg\max \mathbf{P}(X1{=}x1 \mid X2{=}1, \ldots, X5{=}1)}$$

$$= \arg\max \mathbf{P}(X1{=}x1, X2{=}1, \ldots, X5{=}1)$$



Answering queries using Cutset Network (MAP)

"What is the most likely diagnosis for a subject with all risk factors?"

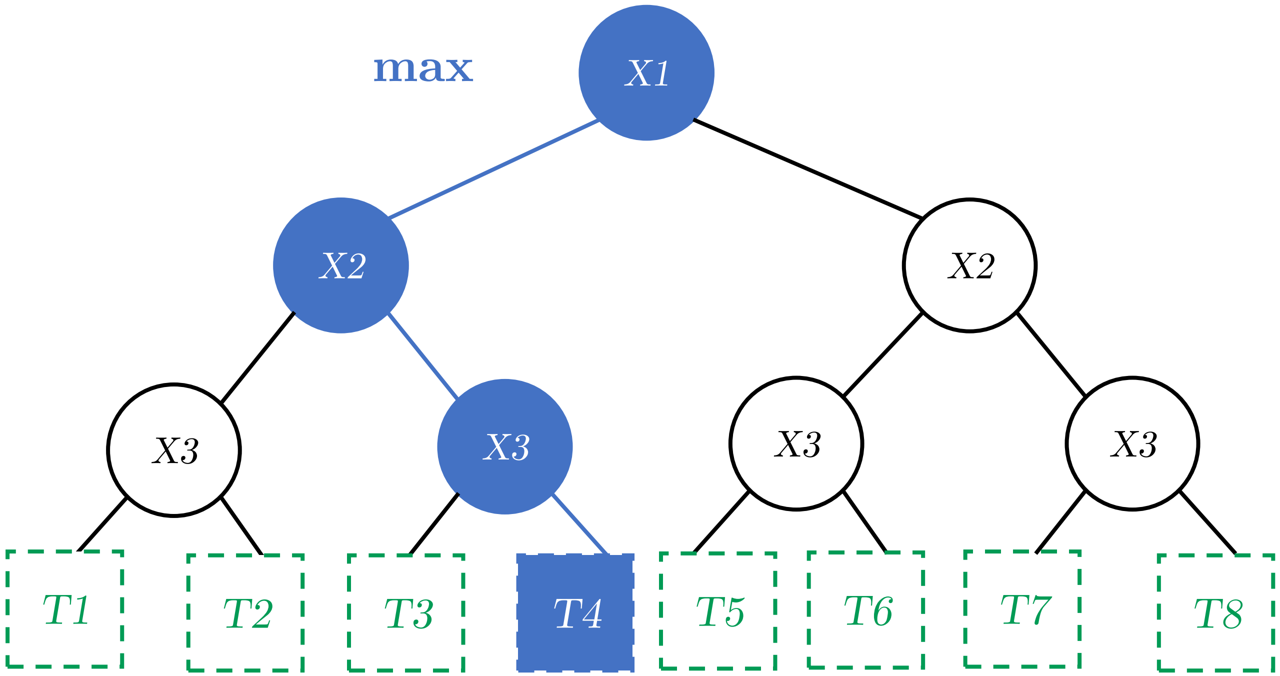


$\arg\max \mathbf{P}(X1{=}x1 \mid X2{=}1, \ldots, X5{=}1)$

$= \arg\max \mathbf{P}(X1{=}x1, X2{=}1, \ldots, X5{=}1)$

$= \arg\max \{\mathbf{P}(X1{=}0, X2{=}1, \ldots, X5{=}1),$

$\}$



Answering queries using Cutset Network (MAP)

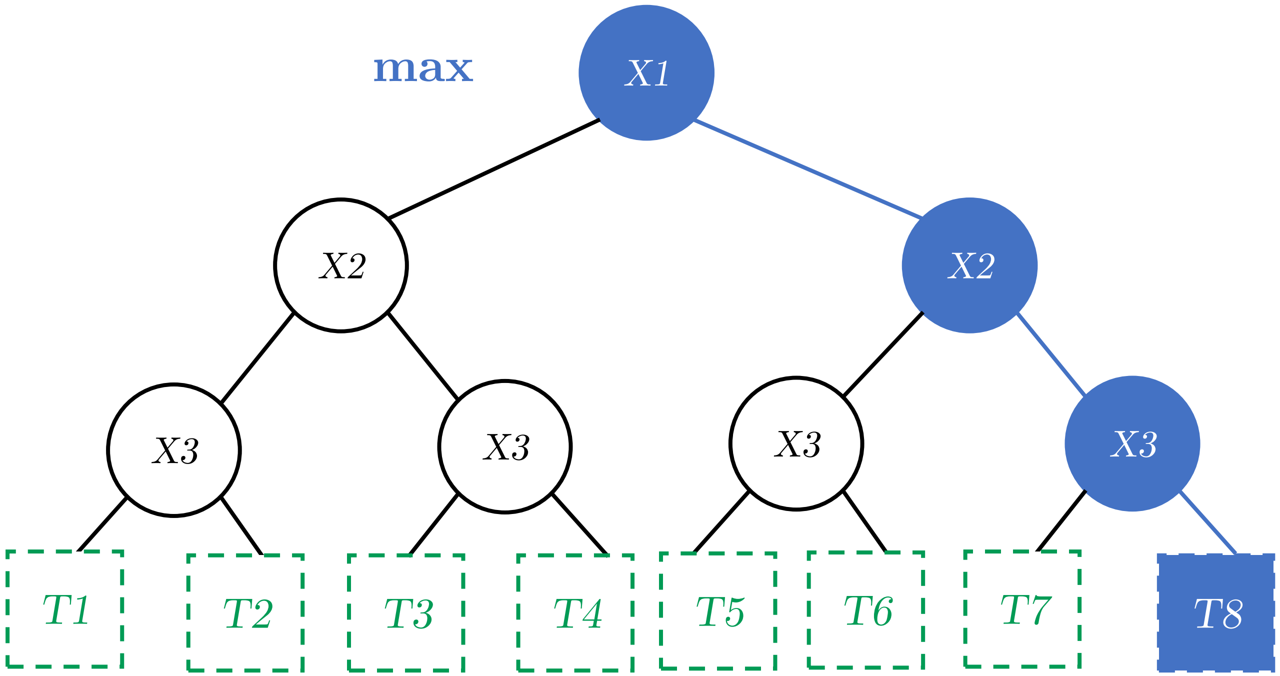
"What is the most likely diagnosis for a subject with all risk factors?"

$\arg\max \mathbf{P}(X1{=}x1 \mid X2{=}1, \ldots, X5{=}1)$

$= \arg\max \mathbf{P}(X1{=}x1, X2{=}1, \ldots, X5{=}1)$

$= \arg\max \{\mathbf{P}(X1{=}0, X2{=}1, \ldots, X5{=}1),$

$\mathbf{P}(X1{=}1, X2{=}1, \ldots, X5{=}1)\}$



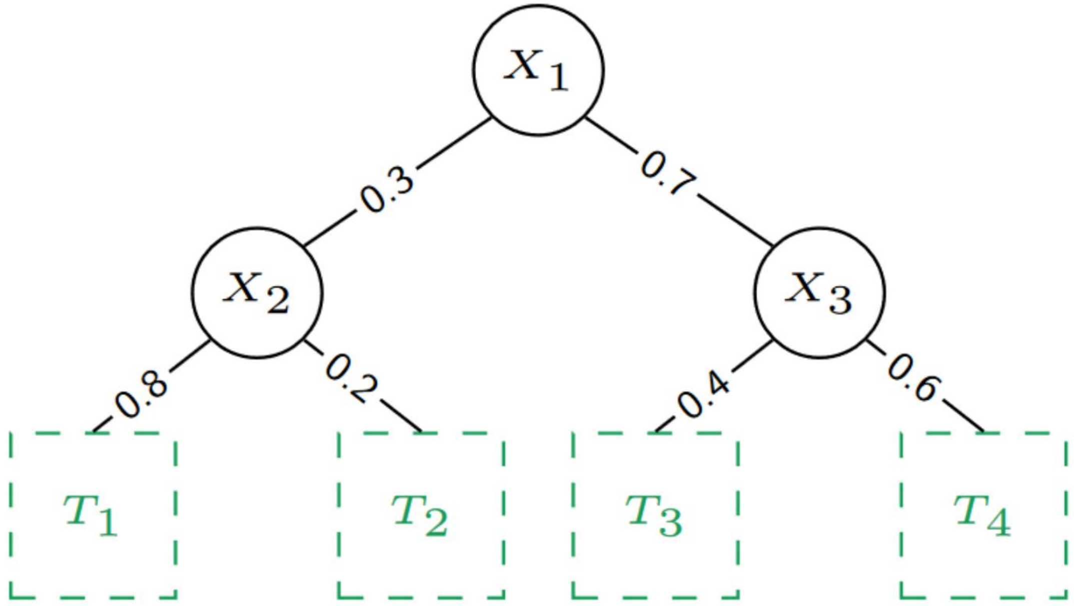
MAP inference

Given: query variables X_Q ,evidence E

$$\text{such that } \mathrm{X} = \mathrm{X_Q} \cup \mathrm{X_E}$$

To Do: arg max $\mathbf{P}(X_Q\,,\, X_E = x_E)$

1. Start at root.
2. Traverse the network depth-first.
3. If current node is evidence variable,
   Select child based on value.
4. Otherwise,
   Take max of query over all children.



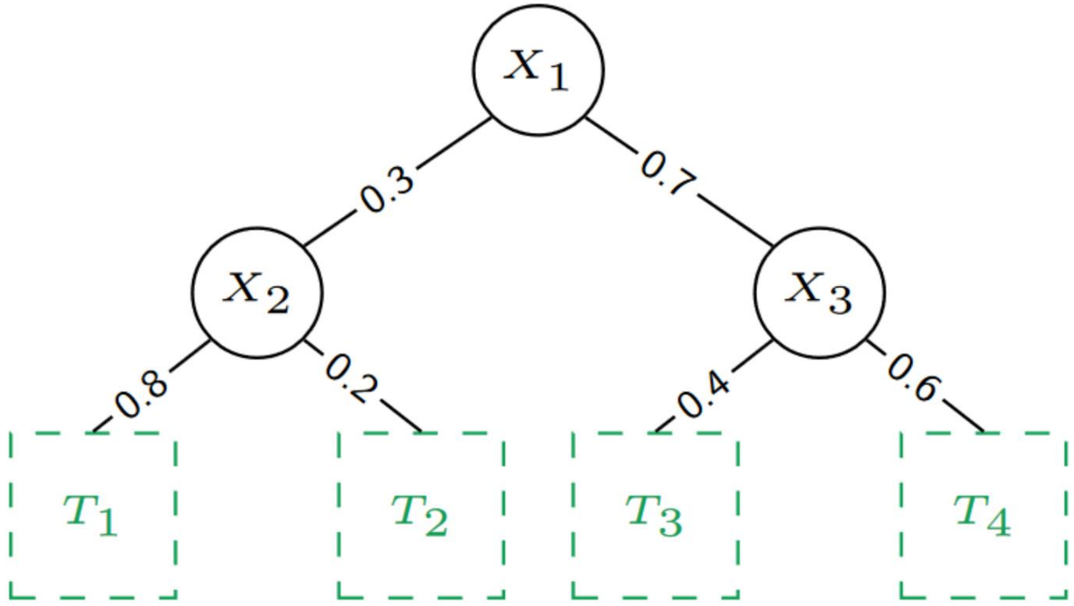
Time complexity

n, number of variables

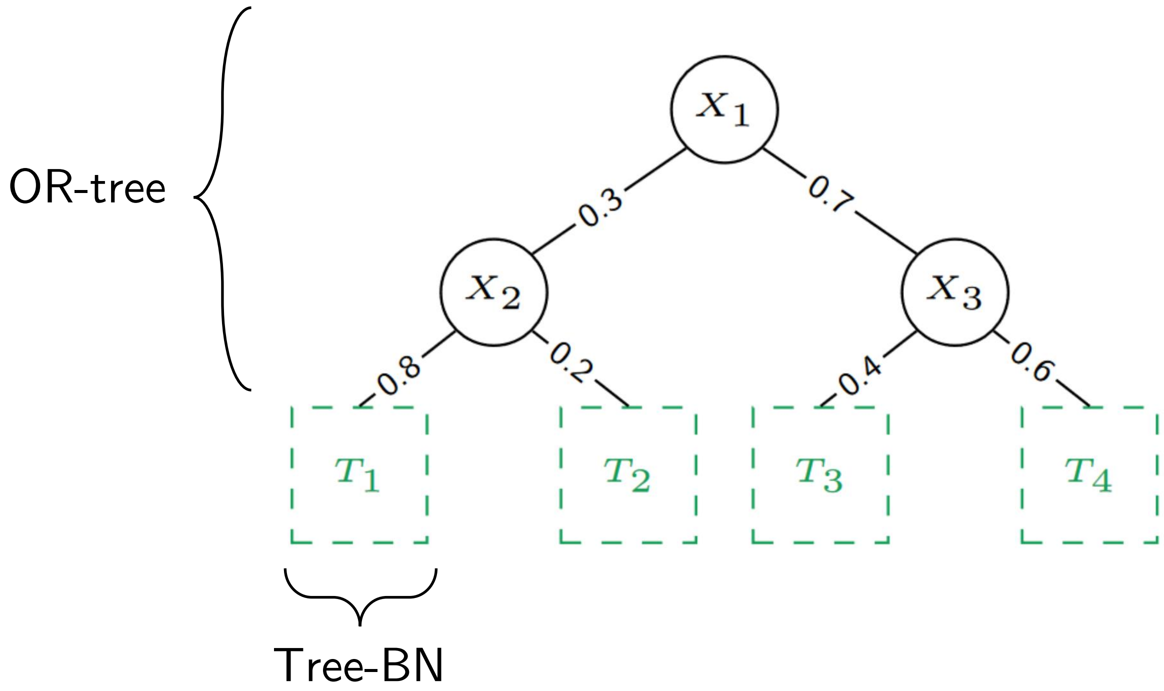
d, max. depth

Cost of traversing OR-tree $= \mathrm{O}(d)$

Cost of inference in tree-BN $= \mathrm{O}(n\text{-}d)$



Cutset Networks



Combination of OR-trees & Tree-BNs

Tractable EVI, MAR and MAP queries

No latent variables

Naturally encodes variety of knowledge

Rahman, Tahrima, Prasanna Kothalkar, and Vibhav Gogate. "Cutset networks: A simple, tractable, and scalable approach for improving the accuracy of chow-liu trees." ECML PKDD 2014,

Cutset Networks: Outline

1. Representation
2. Inference
3. **Learning**

Learning Cutset Networks

1. Build OR-tree
2. Learn Tree-BN at the leaves of the OR-tree (Chow-liu algorithm)

LearnCNet: learning by greedy search [Rahman et al. 14]

Given: Data set $\mathcal{D}$ over n variables $\mathbf{X}$

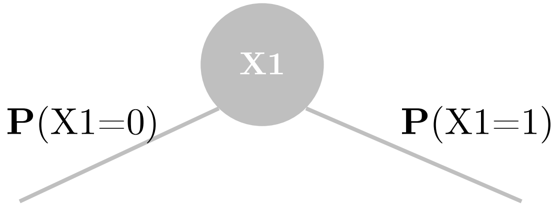
To Do: Learn a Cutset network $\mathcal{M}$

Rahman, Tahrima, Prasanna Kothalkar, and Vibhav Gogate. "Cutset networks: A simple, tractable, and scalable approach for improving the accuracy of chow-liu trees." ECML PKDD 2014,

LearnCNet: learning by greedy search [Rahman et al. 14]

Given: Data set $\mathcal{D}$ over n variables $\mathbf{X}$

To Do: Learn a Cutset network $\mathcal{M}$

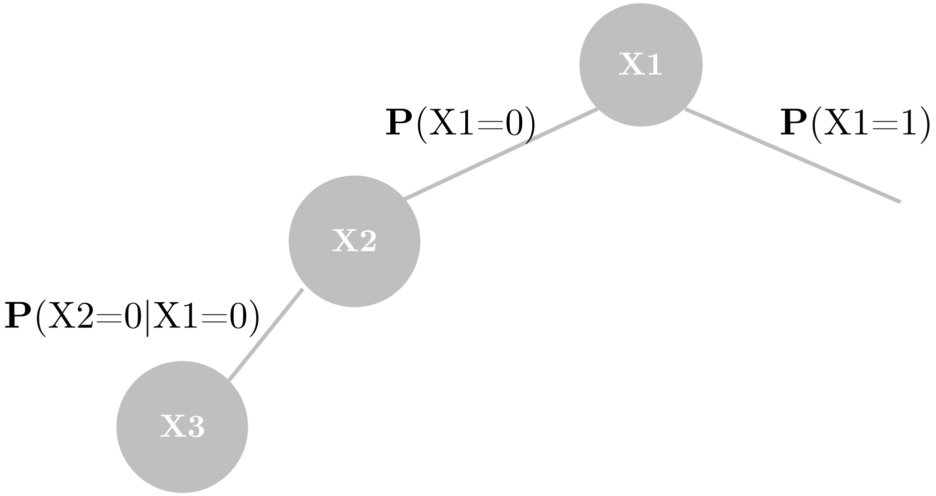


1. Select variable X_i using heuristic
2. Split on X_i, estimate edge weights

LearnCNet: learning by greedy search [Rahman et al. 14]

Given: Data set $\mathcal{D}$ over n variables $\mathbf{X}$

To Do: Learn a Cutset network $\mathcal{M}$



1. Select variable X_i using heuristic
2. Split on X_i, estimate edge weights
3. If stopping condition not met,

   Recurse into children

LearnCNet: learning by greedy search [Rahman et al. 14]

Given: Data set $\mathcal{D}$ over n variables $\mathbf{X}$

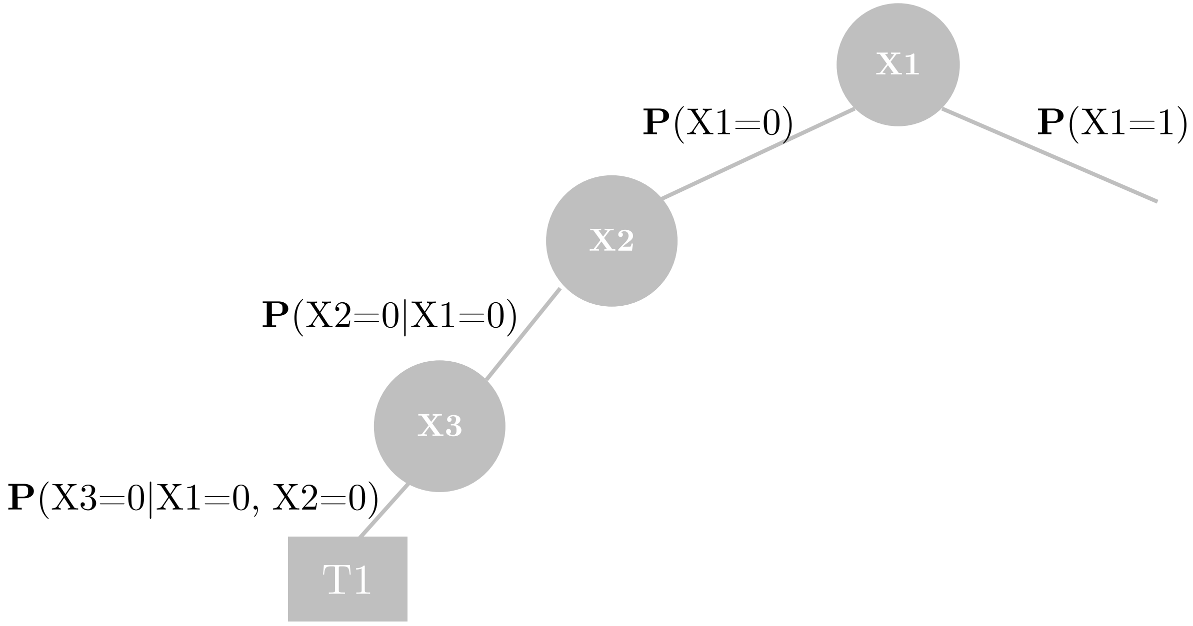
To Do: Learn a Cutset network $\mathcal{M}$

1. Select variable X_i using heuristic
2. Split on X_i, estimate edge weights
3. If stopping condition not met,

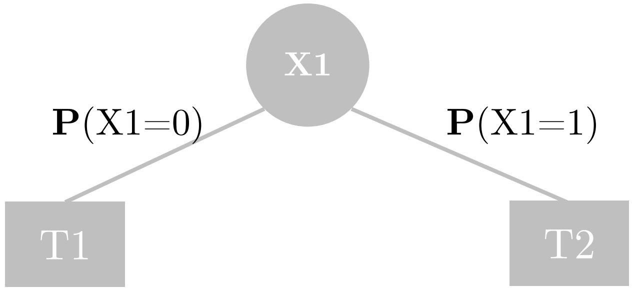
   Recurse into children

4. Else,

   Learn tree-BN over remaining variables



dCSN: learning by direct ML[Mauro et al. 15]



Given: Data set $\mathcal{D}$, tree-BN T over n variables $\mathbf{X}$

To Do: Decompose T into Cutset network $\mathcal{M}$

1. Select a variable, X_i using the BIC (LL regularized by #bits)
2. Replace T with a CN of depth 1, rooted at X_i if it improves BIC
3. If stopping condition not met,

   Recurse into children

Di Mauro, Nicola, Antonio Vergari, and Floriana Esposito. "Learning accurate cutset networks by exploiting decomposability." AI*IA 2015

Recap: PCs

Sum-product network (SPN)

- Nodes { $+$, $\times$, x }
- Smooth, Decomposable
- $+$ introduces latent variables
- Universal density approximator
- EVI, MAR tractable
- Mixture is SPN

Cutset network (CN)

- Nodes { X , T }
- Smooth, Decomposable, Deterministic
- No latent variables; Interpretable
- Not a universal density approximator
- EVI, MAR, MAP tractable
- Mixture is not CN

Recap: Queries

Question	Query	Expression	Model
Generate a data point	Sampling	$x \sim \boldsymbol{P}_M(\boldsymbol{X})$	Variational auto-encoder
How likely is a data point?	Full evidence (EVI)	$\boldsymbol{P}_M(\boldsymbol{X} = x)$	Normalizing Flow
How likely is this partial data point?	Marginal (MAR)	$\boldsymbol{P}_M(\boldsymbol{X}_E = x_E)$	Sum-product network
What is the most likely assignment given remaining values?	Maximum a posteriori (MAP)	$\arg\max_{x_{-E}} \boldsymbol{P}_M(\boldsymbol{X}_{-E} = x_{-E} \mid \boldsymbol{X}_E = x_E)$	Cutset network
What is the most likely assignment given some values?	Marginal MAP (MMAP)	$\arg\max_{x_Q} \boldsymbol{P}_M(\boldsymbol{X}_Q = x_Q \mid \boldsymbol{X}_E = x_E)$	Fully factorized distribution

Recap: Queries

Question	**Query**	**Expression**	**Model**
Generate a data point	Sampling	$x \sim \boldsymbol{P}_M(\boldsymbol{X})$	Variational auto-encoder
How likely is a data point?	Full evidence (EVI)	$\boldsymbol{P}_M(\boldsymbol{X} = x)$	Normalizing Flow
How likely is this partial data point?	Marginal (MAR)	$\boldsymbol{P}_M(\boldsymbol{X}_E = x_E)$	Sum-product network
What is the most likely assignment given remaining values	Maximum a posteriori (MAP)	$\underset{x_{-E}}{\arg\max}\, \boldsymbol{P}_M(\boldsymbol{X}_{-E} = x_{-E} \mid \boldsymbol{X}_E = x_E)$	Cutset network
What is the most likely assignment given some values	Marginal MAP (MMAP)	$\underset{x_Q}{\arg\max}\, \boldsymbol{P}_M(\boldsymbol{X}_Q = x_Q \mid \boldsymbol{X}_E = x_E)$	Fully factorized distribution

Recap: Queries

Question	Query	Expression	Model
Generate a data point	Sampling	$x \sim \boldsymbol{P}_M(\boldsymbol{X})$	Variational auto-encoder
How likely is a data point?	Full evidence (EVI)	$\boldsymbol{P}_M(\boldsymbol{X} = x)$	Normalizing Flow
How likely is this partial data point?	Marginal (MAR)	$\boldsymbol{P}_M(\boldsymbol{X}_E = x_E)$	Sum-product network
What is the most likely assignment given remaining values?	Maximum a posteriori (MAP)	$\arg\max_{x_{-E}} \boldsymbol{P}_M(\boldsymbol{X}_{-E} = x_{-E} \mid \boldsymbol{X}_E = x_E)$	Cutset network
What is the most likely assignment given some values?	Marginal MAP (MMAP)	$\arg\max_{x_Q} \boldsymbol{P}_M(\boldsymbol{X}_Q = x_Q \mid \boldsymbol{X}_E = x_E)$	Fully factorized distribution

Recap: Queries

Question	**Query**	**Expression**	**Model**
Generate a data point	Sampling	$x \sim \boldsymbol{P}_M(\boldsymbol{X})$	Variational auto-encoder
How likely is a data point?	Full evidence (EVI)	$\boldsymbol{P}_M(\boldsymbol{X} = x)$	Normalizing Flow
How likely is this partial data point?	Marginal (MAR)	$\boldsymbol{P}_M(\boldsymbol{X}_E = x_E)$	Sum-product network
What is the most likely assignment given remaining values?	Maximum a posteriori (MAP)	$\arg\max_{x_{-E}} \boldsymbol{P}_M(\boldsymbol{X}_{-E} = x_{-E} \mid \boldsymbol{X}_E = x_E)$	Cutset network
What is the most likely assignment given some values?	Marginal MAP (MMAP)	$\arg\max_{x_Q} \boldsymbol{P}_M(\boldsymbol{X}_Q = x_Q \mid \boldsymbol{X}_E = x_E)$	Fully factorized distribution

Recap: Queries

Question	**Query**	**Expression**	**Model**
Generate a data point	Sampling	$x \sim \boldsymbol{P}_M(\boldsymbol{X})$	Variational auto-encoder
How likely is a data point?	Full evidence (EVI)	$\boldsymbol{P}_M(\boldsymbol{X} = x)$	Normalizing Flow
How likely is this partial data point?	Marginal (MAR)	$\boldsymbol{P}_M(\boldsymbol{X}_E = x_E)$	Sum-product network
What is the most likely assignment given remaining values?	Maximum a posteriori (MAP)	$\arg\max_{x_{-E}} \boldsymbol{P}_M(\boldsymbol{X}_{-E} = x_{-E} \mid \boldsymbol{X}_E = x_E)$	Cutset network
What is the most likely assignment given some values?	Marginal MAP (MMAP)	$\arg\max_{x_Q} \boldsymbol{P}_M(\boldsymbol{X}_Q = x_Q \mid \boldsymbol{X}_E = x_E)$	Fully factorized distribution

Cutset Networks: Demo

Day 2

Hands-On Demo

bit.ly/tpm-day2-cnets