Bayesian Networks

Probabilistic Graphical Models

Jerónimo Hernández-González

Concepts

- Joint/Marginal/Conditional probability distribution
- Factorization
- Chain rule
- Bayes rule
- (Conditional) independence

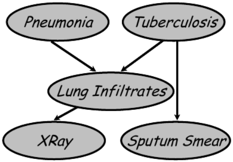
Example: conditional independence

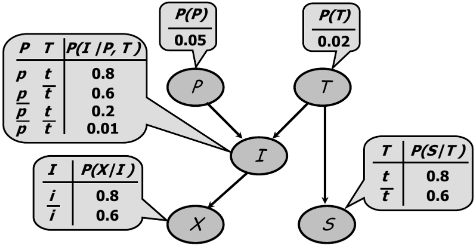
Given a set of variables $\boldsymbol{X} = X_1, ..., X_5$, and $3 \perp\!\!\!\perp 4|1, 5$

$$\begin{aligned} p(\boldsymbol{X}) &= p(X_2|\boldsymbol{X}_{1,3,4,5})\textcolor{red}{p(X_3|\boldsymbol{X}_{1,4,5})}p(\boldsymbol{X}_{1,4,5}) \\ &= p(X_2|\boldsymbol{X}_{1,3,4,5})\textcolor{green}{p(X_3|\boldsymbol{X}_{1,5})}p(\boldsymbol{X}_{1,4,5}) \end{aligned}$$

Bayesian networks

- What is a Bayesian network?
- Chain rule.
- What does it mean that a probability distribution factorizes over a graph $\mathcal{G}$?





Bayesian network

Components

Directed acyclic graph $G = (V, E)$ + parameters $\mathbf{\Theta} = (\Theta_1, ..., \Theta_n)$

- Represents a joint probability distribution:

Vertices: related to random variables

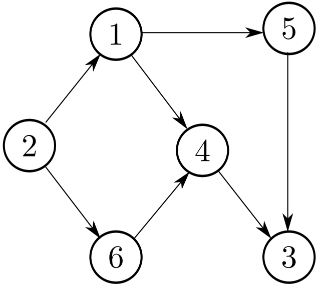
Edges: related to the simplification of the chain rule

Parameters: tables of the conditional probability distributions

Directed acyclic graph (DAG)

Formalism

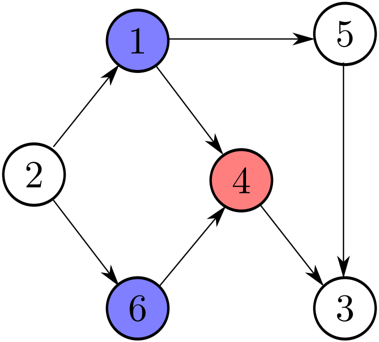
- A DAG G is a pair (V, E)
- $V = \{1, ..., n\}$ represent the set of vertices
- $E = \{(u, v) : u, v \in V, u \neq v\}$ represent the set of arcs
- There are no directed cycles in G



DAG concepts

Definitions

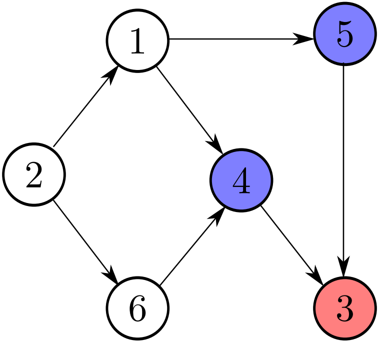
Parents, children, ancestral set, ancestral ordering, moral



DAG concepts

Definitions

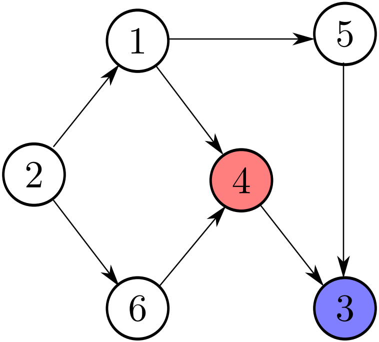
Parents, children, ancestral set, ancestral ordering, moral



DAG concepts

Definitions

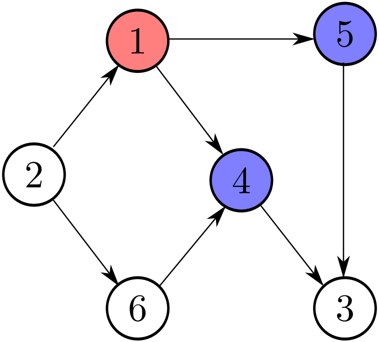
Parents, children, ancestral set, ancestral ordering, moral



DAG concepts

Definitions

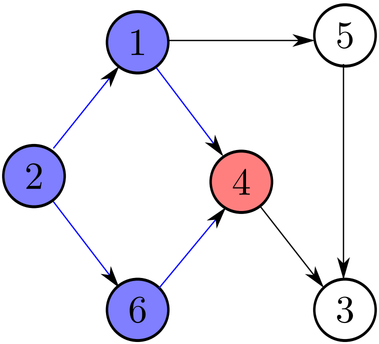
Parents, children, ancestral set, ancestral ordering, moral



DAG concepts

Definitions

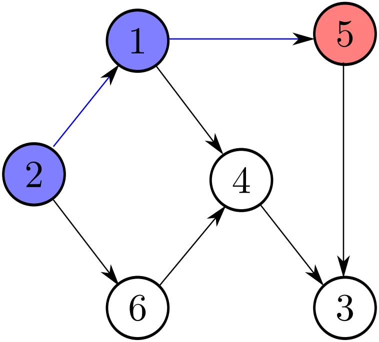
Parents, children, ancestral set, ancestral ordering, moral



DAG concepts

Definitions

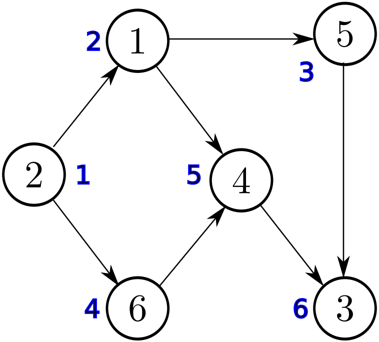
Parents, children, ancestral set, ancestral ordering, moral



DAG concepts

Definitions

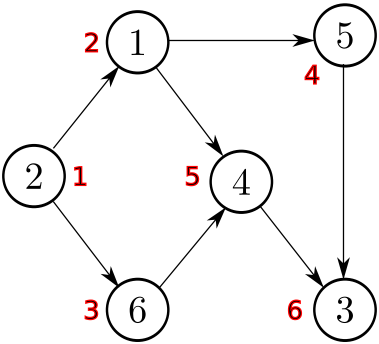
Parents, children, ancestral set, ancestral ordering, moral



DAG concepts

Definitions

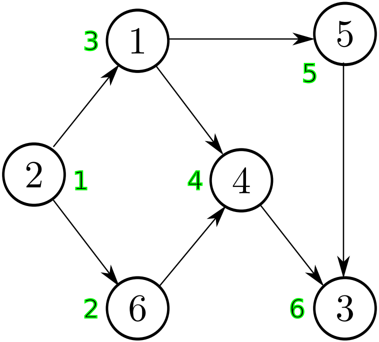
Parents, children, ancestral set, ancestral ordering, moral



DAG concepts

Definitions

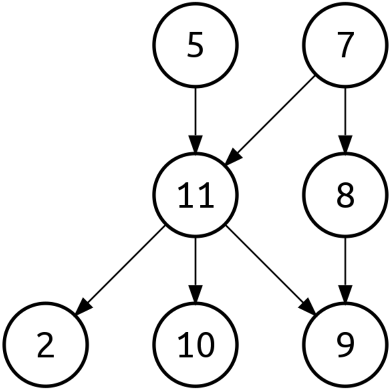
Parents, children, ancestral set, ancestral ordering, moral



DAG concepts

Definitions

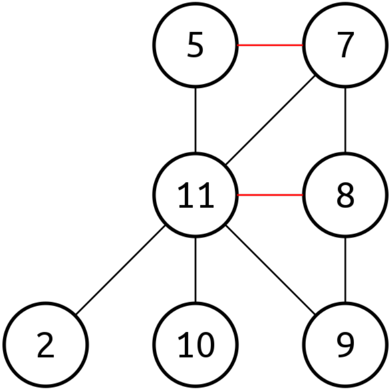
Parents, children, ancestral set, ancestral ordering, moral



DAG concepts

Definitions

Parents, children, ancestral set, ancestral ordering, moral



Exercise

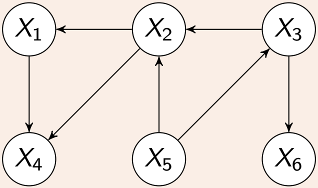
Let be $\boldsymbol{X} = (X_1, X_2, X_3, X_4, X_5, X_6)$

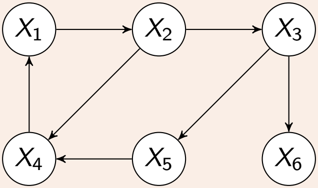
Let $G_1 = (\boldsymbol{X}, E_1)$, where
$E_1 = \{(X_3, X_2), (X_5, X_3), (X_2, X_1), (X_5, X_2), (X_3, X_6), (X_2, X_4), (X_1, X_4)\}$

Let $G_2 = (\boldsymbol{X}, E_2)$, where
$E_2 = \{(X_2, X_3), (X_3, X_5), (X_3, X_6), (X_2, X_4), (X_5, X_4), (X_4, X_1), (X_1, X_2)\}$

1. Are they DAGs?
2. Obtain the parents and children of X_3 and X_5
3. Find an ancestral ordering for G_1 and G_2

Exercise





Bayesian networks

Factorization

A Bayesian network $M = (G, \boldsymbol{\Theta})$ can be expressed as a product of conditional probability distribution:

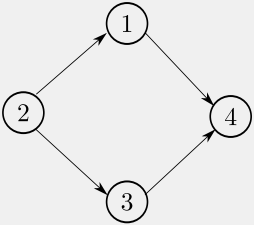
$$p_M(\boldsymbol{x}) = \prod_{i=1}^{n} p(x_i | \boldsymbol{pa}_i; \boldsymbol{\Theta_i})$$

where $\boldsymbol{pa}_i$ relates to the set of variables with an edge towards X_i.

- Qualitative: G describes the skeleton of the factorization
- Quantitative: $\boldsymbol{\Theta}$ determines the shape of the conditional distributions

Bayesian networks

Factorization



Bayesian network's factorization:

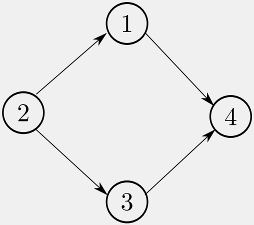
$$p(\boldsymbol{X}) = p(X_4|X_1, X_3) \cdot p(X_1|X_2) \cdot p(X_3|X_2) \cdot p(X_2)$$

Chain rule:

$$p(\boldsymbol{X}) = p(X_4|X_1, {\color{red}X_2}, X_3) \cdot p(X_1|X_2, {\color{red}X_3}) \cdot p(X_3|X_2) \cdot p(X_2)$$

Bayesian networks

Factorization



Conditional independencies

- $X_4 \perp\!\!\!\perp X_2 \mid X_1, X_3$
- $X_1 \perp\!\!\!\perp X_3 \mid X_2$

Bayesian network's factorization:

$$p(\boldsymbol{X}) = p(X_4|X_1, X_3) \cdot p(X_1|X_2) \cdot p(X_3|X_2) \cdot p(X_2)$$

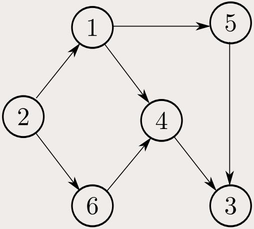
Chain rule:

$$p(\boldsymbol{X}) = p(X_4|X_1, {\color{red}X_2}, X_3) \cdot p(X_1|X_2, {\color{red}X_3}) \cdot p(X_3|X_2) \cdot p(X_2)$$

Exercise

Questions:

1. Find an ancestral ordering
2. Determine the factorized chain rule

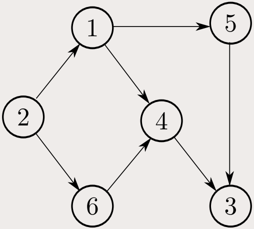


Exercise:

A possible answer

1. Find an ancestral ordering: $2, 1, 6, 4, 5, 3$
2. Determine the factorized chain rule

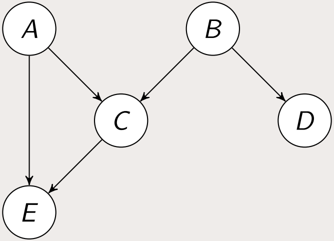
$$p_M(\boldsymbol{x}) = p(x_1|x_2) \cdot p(x_2) \cdot p(x_3|x_4, x_5) \cdot p(x_4|x_1, x_6) \cdot p(x_5|x_1) \cdot p(x_6|x_2)$$



Exercise:

Factorization

Given the model, which of these is an appropriate decomposition of the joint distribution?



a) $P(A, B, C, D, E) = P(A)P(B)P(A, B|C)P(B|D)P(A, C|E)$

b) $P(A, B, C, D, E) = P(A)P(B)P(C|A)P(C|B)P(D|B)P(E|A)P(E|C)$

c) $P(A, B, C, D, E) = P(A)P(B)P(C|A, B)P(D|B)P(E|A, C)$

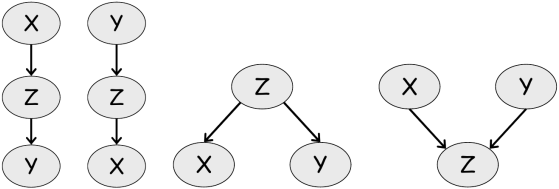
d) $P(A, B, C, D, E) = P(A)P(B)P(C)P(D)P(E)$

Bayesian Networks

Probabilistic Graphical Models

Jerónimo Hernández-González

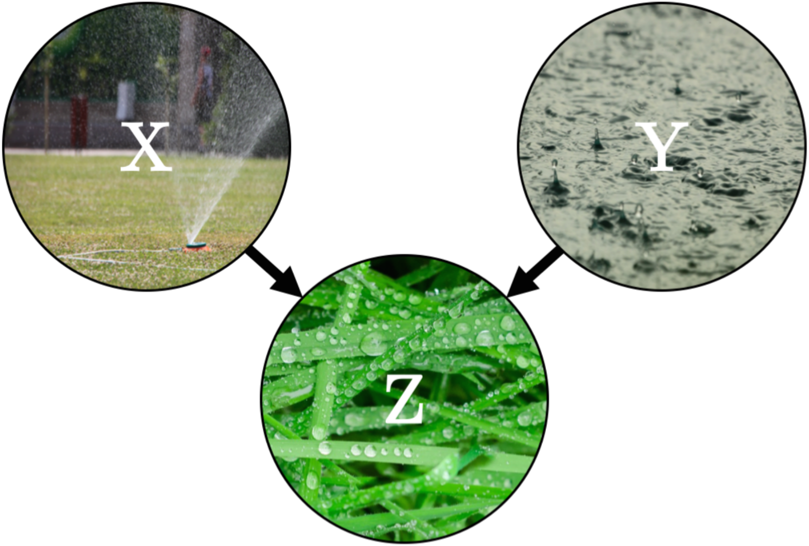
Flow of probabilistic influence



In Bayesian networks, influence flow is stopped by observed nodes and non-observed v-structures.
A v-structure is observed if Z or any of its descendants is observed.

Flow of probabilistic influence

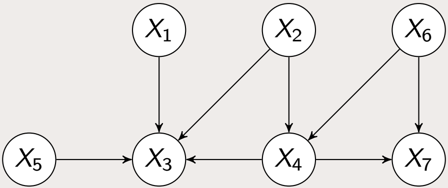
The wet grass example



Z indicates whether our garden is wet; Possible explanations are: X, we turned on the sprinklers; Y, rain.
If grass is wet (Z =True) and we didn't turn on the sprinklers (X =False), then the probability of rain (A =True) increases!

Exercise

Flow of influence



True or false:

- $X_5 \perp\!\!\!\perp X_3$
- $X_6 \perp\!\!\!\perp X_3$
- $X_1 \perp\!\!\!\perp X_5$
- $X_3 \perp\!\!\!\perp X_7$
- $X_1 \perp\!\!\!\perp X_5 \mid X_3$
- $X_6 \perp\!\!\!\perp X_2 \mid X_3$

Flow of probabilistic influence

Active trail

- Let $\mathcal{G}$ be a DAG
- Let $X_1 \rightleftharpoons \ldots \rightleftharpoons X_m$ be a trail in $\mathcal{G}$
- A trail is active given a set of observed variables $\boldsymbol{Z}$ if:
  1. Whenever there is a v-structure $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, X_i or one of its descendants is in $\boldsymbol{Z}$
  2. no other node along the trail is in $\boldsymbol{Z}$

d-separation

- Let $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ be three disjoint sets of variables in $\mathcal{G}$.
- $\boldsymbol{Z}$ *d*-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$ in $\mathcal{G}$ if $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{Z}$ holds in $\mathcal{G}$
- $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{Z}$ holds in $\mathcal{G}$ if there is no active trail between any variable in $\boldsymbol{X}$ and any variable in $\boldsymbol{Y}$ given $\boldsymbol{Z}$.

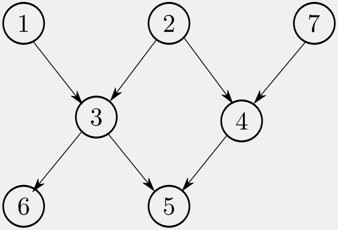
*Graphical criteria of **d**-separation*

$\boldsymbol{Z}$ *d*-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



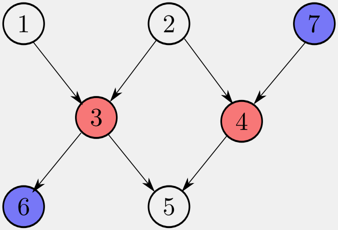
Graphical criteria of d-separation

$\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_6 \perp\!\!\!\perp X_7 \mid X_3, X_4?$$

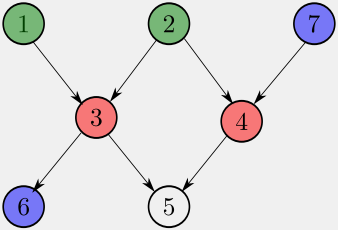
Graphical criteria of d-separation

$\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_6 \perp\!\!\!\perp X_7 \mid X_3, X_4?$$

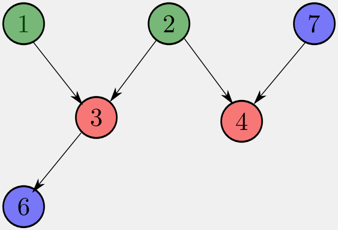
Graphical criteria of d-separation

Z d-separates X from Y?

Three steps:

1. Identify the ancestors of X, Y and Z
2. Remove the rest of variables and moralize the left subgraph
3. Does Z block all the paths from X to Y?

Example



$$X_6 \perp\!\!\!\perp X_7 \mid X_3, X_4?$$

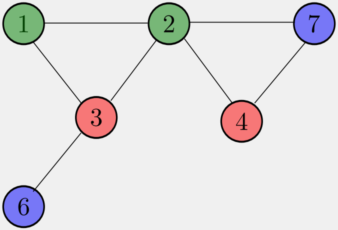
Graphical criteria of d-separation

Z d-separates X from Y?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$X_6 \perp\!\!\!\perp X_7 \mid X_3, X_4$? true

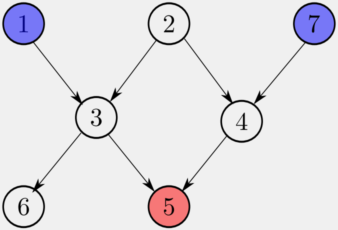
Graphical criteria of $\boldsymbol{d}$*-separation*

$\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_1 \perp\!\!\!\perp X_7 \mid X_5?$$

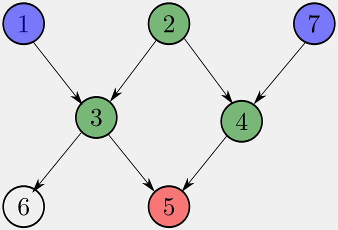
Graphical criteria of d-separation

Z d-separates X from Y?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_1 \perp\!\!\!\perp X_7 \mid X_5?$$

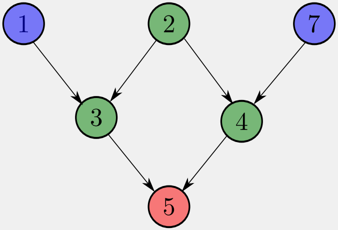
Graphical criteria of $\boldsymbol{d}$*-separation*

$\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_1 \perp\!\!\!\perp X_7 \mid X_5?$$

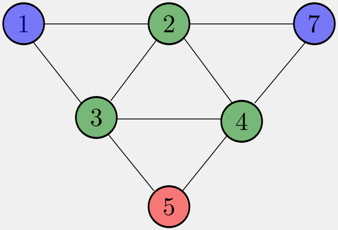
Graphical criteria of $\boldsymbol{d}$*-separation*

$\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$X_1 \perp\!\!\!\perp X_7 \mid X_5$? false

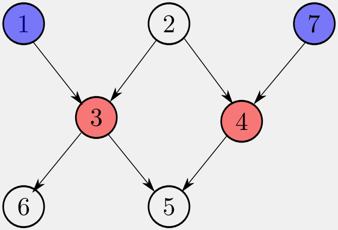
Graphical criteria of $\boldsymbol{d}$*-separation*

$\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_1 \perp\!\!\!\perp X_7 \mid X_3, X_4?$$

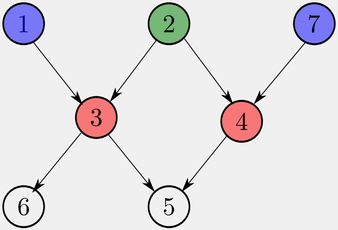
Graphical criteria of **d***-separation*

$\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_1 \perp\!\!\!\perp X_7 \mid X_3, X_4?$$

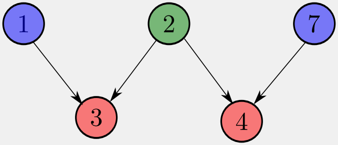
Graphical criteria of d-separation

Z d-separates X from Y?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_1 \perp\!\!\!\perp X_7 \mid X_3, X_4?$$

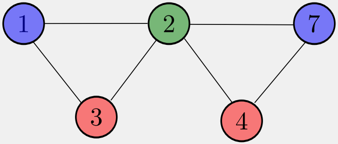
Graphical criteria of d-separation

$\boldsymbol{Z}$ *d*-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_1 \perp\!\!\!\perp X_7 \mid X_3, X_4 ?\ \text{false}$$

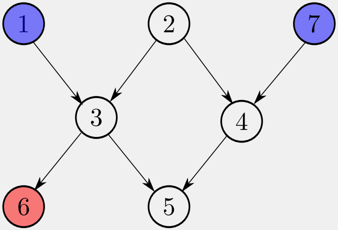
Graphical criteria of $\boldsymbol{d}$*-separation*

$\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_1 \perp\!\!\!\perp X_7 \mid X_6?$$

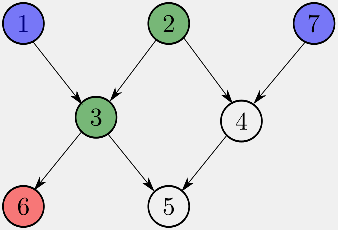
Graphical criteria of d-separation

Z d-separates X from Y?

Three steps:

1. Identify the ancestors of X, Y and Z
2. Remove the rest of variables and moralize the left subgraph
3. Does Z block all the paths from X to Y?

Example



$$X_1 \perp\!\!\!\perp X_7 \mid X_6?$$

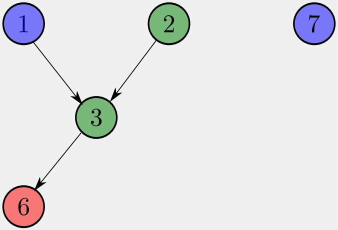
Graphical criteria of d-separation

$\boldsymbol{Z}$ d-separates $\boldsymbol{X}$ from $\boldsymbol{Y}$?

Three steps:

1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$$X_1 \perp\!\!\!\perp X_7 \mid X_6?$$

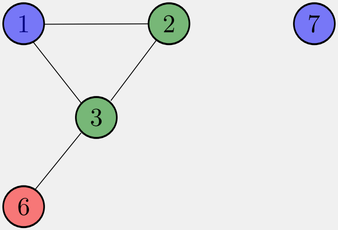
Graphical criteria of d-separation

Z d-separates X from Y?

Three steps:

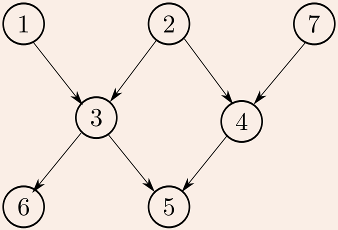
1. Identify the ancestors of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$
2. Remove the rest of variables and moralize the left subgraph
3. Does $\boldsymbol{Z}$ block all the paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$?

Example



$X_1 \perp\!\!\!\perp X_7 \mid X_6$? true

Exercise

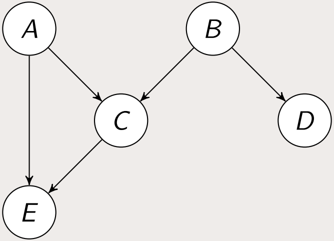


- Check if these terms are verified:
  - $X_1 \perp\!\!\!\perp X_4 \mid X_5$
  - $X_1 \perp\!\!\!\perp X_5 \mid X_3, X_4$
  - $X_6 \perp\!\!\!\perp X_7 \mid X_3$
  - $X_1 \perp\!\!\!\perp X_7 \mid X_5$
  - $X_1 \perp\!\!\!\perp X_4 \mid X_6$
  - $X_1 \perp\!\!\!\perp X_4 \mid X_2, X_3$
- Identify the subsets of variables that verify:
  - $X_1 \perp\!\!\!\perp X_7 \mid ?$
  - $X_1 \perp\!\!\!\perp X_5 \mid ?$

Exercise

Independencies in a graph

Which pairs of variables are independent in this BN, given that none of them have been observed?



a) $C \perp\!\!\!\perp D$

b) $A \perp\!\!\!\perp B$

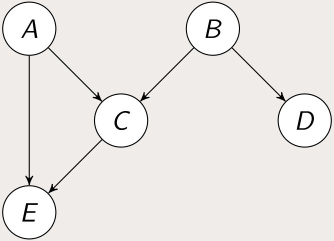
c) $B \perp\!\!\!\perp E$

d) $A \perp\!\!\!\perp C$

Exercise

Independencies in a graph

Now assume that the value of E is observed. Which pairs of variables are independent given E?



a) $A \perp\!\!\!\perp D | E$

b) $A \perp\!\!\!\perp C | E$

c) $A \perp\!\!\!\perp B | E$

d) $D \perp\!\!\!\perp C | E$

e) $B \perp\!\!\!\perp D | E$

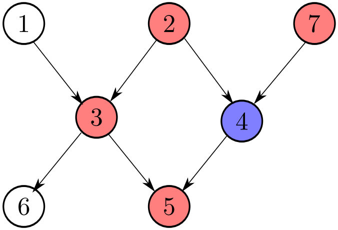
Markov blanket

Definition

- Parents, children and parents of the children
- Given X_j and $\boldsymbol{Mb}_j$, for any set of variables $\boldsymbol{X}_c \subseteq (\boldsymbol{V} \setminus \boldsymbol{Mb}_j)$:

$$\boldsymbol{X}_c \perp\!\!\!\perp X_j \mid \boldsymbol{Mb}_j$$

** Applications for structural learning, feature subset selection,...



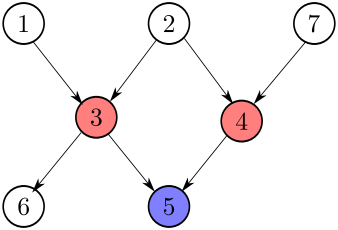
Markov blanket

Definition

- Parents, children and parents of the children
- Given X_j and $\boldsymbol{Mb}_j$, for any set of variables $\boldsymbol{X}_c \subseteq (\boldsymbol{V} \setminus \boldsymbol{Mb}_j)$:

$$\boldsymbol{X}_c \perp\!\!\!\perp X_j \mid \boldsymbol{Mb}_j$$

** Applications for structural learning, feature subset selection,...



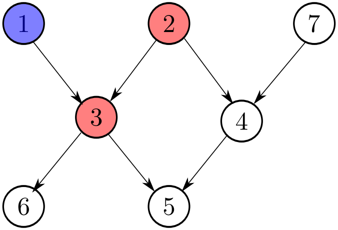
Markov blanket

Definition

- Parents, children and parents of the children
- Given X_j and $\boldsymbol{Mb}_j$, for any set of variables $\boldsymbol{X}_c \subseteq (\boldsymbol{V} \setminus \boldsymbol{Mb}_j)$:

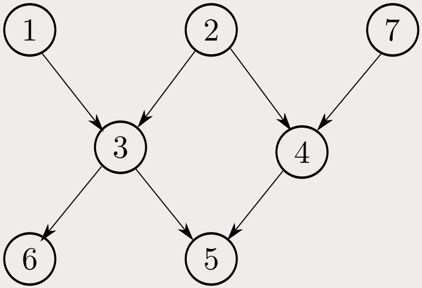
$$\boldsymbol{X}_c \perp\!\!\!\perp X_j \mid \boldsymbol{Mb}_j$$

** Applications for structural learning, feature subset selection,...

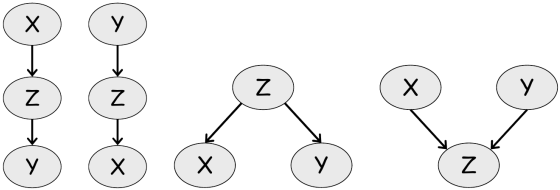


Exercise

Determine the Markov Blanket of X_3



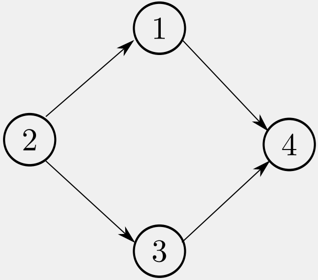
Equivalence classes



Equivalence classes

Redundancy

Two different DAGs can determine the same dependence model



The same undirected edges with the same V-structures

** Applications in structural learning

Independency Maps

I-maps

- $I(\mathcal{G})$ is the set of independence statements that hold on $\mathcal{G}$

  Any distribution that factorizes over $\mathcal{G}$ satisfies the ind. statements in $I(\mathcal{G})$
- $I(\mathcal{P})$ is the set of independence statements that hold for P
- $\mathcal{G}$ is an I-map of P if $I(\mathcal{G}) \subseteq I(\mathcal{P})$

  The other way around is not necessarily true ($\sim$ wasted parameters)

D-separation and I-maps

- If P factorizes over $\mathcal{G}$ and X and Y are d-separated in $\mathcal{G}$ given Z then P satisfies $X \perp\!\!\!\perp Y|Z$

Exercise

I-maps

Which of the following statements about I-maps are true? You may select 1 or more options, or none of them.

a) The graph K that is the same as the graph G, except that all of the edges are oriented in the opposite direction as the corresponding edges in G, is always an I-map for G, regardless of the structure of G.

b) A graph K is an I-map for a graph G if and only if K encodes all of the independences that G has, and more.

c) An I-map is a function that maps a graph G to itself, i.e., $f(G) = G$.

d) A graph K is an I-map for a graph G if and only if all of the independencies encoded by K are also encoded by G.

Connecting the three views

- Given a Bayesian network structure $\mathcal{G}$, and a probability distribution P that factorizes according to $\mathcal{G}$, then P satisfies the independencies that hold in $\mathcal{G}$.

$$P \text{ factorizes} \implies P \text{ satisfies independencies}$$

- Given a Bayesian network structure $\mathcal{G}$, and a probability distribution P that satisfies the independencies that hold in $\mathcal{G}$, then P factorizes over $\mathcal{G}$.

$$P \text{ satisfies independencies} \implies P \text{ factorizes}$$

Exercise

Model

Let $X_1, X_2, ..., X_5$ be binary random variables. Given p such that it verifies the set of conditional independences $I = \{X_2 \perp\!\!\!\perp X_1;$ $X_3 \perp\!\!\!\perp X_2|X_1; \quad X_4 \perp\!\!\!\perp X_1, X_2|X_3; \quad X_5 \perp\!\!\!\perp X_2, X_3|X_1, X_4\}$ and the ancestral ordering $1, 2, 3, 4, 5$:

- Obtain the simplified chain rule
- Compute the number of free parameters of the model

Exercise

Ancestral ordering that minimizes the number of parameters

Given a BN over a set of variables $X_1, ..., X_5$,

1. $I = \{X_5 \perp\!\!\!\perp X_1 | X_3; \quad X_2 \perp\!\!\!\perp X_1, X_5 | X_3; \quad X_3 \perp\!\!\!\perp X_1;$
   $X_4 \perp\!\!\!\perp X_3, X_5 | X_1, X_2\}$
2. $I = \{X_4 \perp\!\!\!\perp X_2, X_1 | X_3; \quad X_2 \perp\!\!\!\perp X_1; \quad X_5 \perp\!\!\!\perp X_1, X_4 | X_2, X_3\}$
3. $I = \{X_3 \perp\!\!\!\perp X_1 | X_2; \quad X_4 \perp\!\!\!\perp X_3 | X_2, X_1; \quad X_5 \perp\!\!\!\perp X_1, X_4 | X_2, X_3\}$

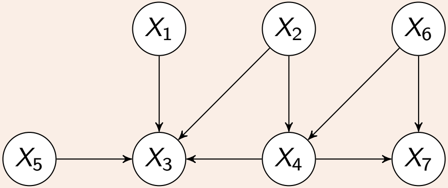
Exercise

1. Let $P(X)$ be a distribution such that

$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2)P(X_3|X_4, X_1)P(X_4|X_2)$$

   Are these (conditional) independencies true?
   - $X_1 \perp\!\!\!\perp X_4$
   - $X_2 \perp\!\!\!\perp X_1$
   - $X_2 \perp\!\!\!\perp X_1 \mid X_3$
   - $X_1 \perp\!\!\!\perp X_2 \mid X_3, X_4$

2. Given that P satisfies the independencies in



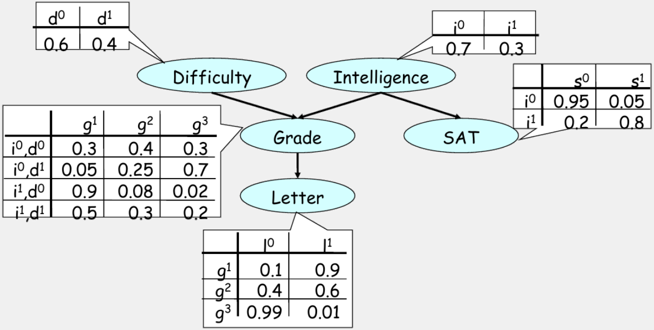
   Provide a factorization for P.

Flow of probabilistic influence

Reasoning Patterns

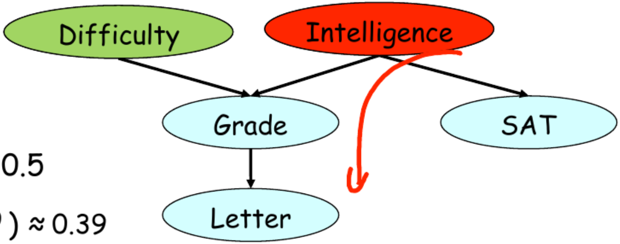
- Causal reasoning
- Evidential reasoning
- Intercausal reasoning

The student network



Flow of probabilistic influence

Reasoning Patterns



$P(l^1) \approx 0.5$

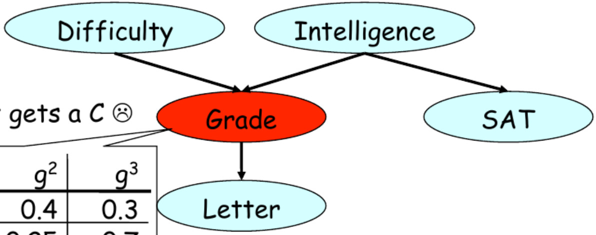
$P(l^1 \mid i^0) \approx 0.39$

$P(l^1 \mid i^0, d^0) \approx 0.51$

Causal reasoning

$P(d^1) = 0.4$
$P(d^1 \mid g^3) \approx 0.63$

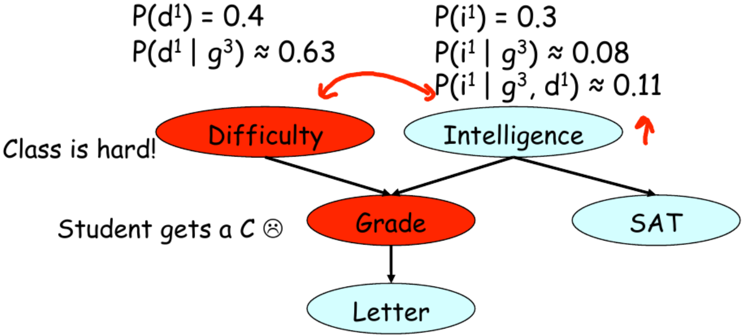
$P(i^1) = 0.3$
$P(i^1 \mid g^3) \approx 0.08$



Student gets a C ☹

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

Evidential reasoning

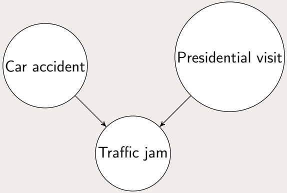


Intercausal reasoning

Exercise

Inter-causal reasoning

In this model for traffic jams in a small town, which we assume can be caused by a car accident, or by a visit from the president.



P(PV = True) = 0.01
P(CA = True) = 0.1
P(TJ = True | PV = False, CA = False) = 0.1
P(TJ = True | PV = False, CA = True) = 0.5
P(TJ = True | PV = True, CA = False) = 0.6
P(TJ = True | PV = True, CA = True) = 0.9

- Calculate $P(CA = True|TJ = True)$
- Calculate $P(CA = True|TJ = True, PV = True)$

  where TJ, CA and PV stand for Traffic Jam, Car Accident and Presidential Visit respectively.

Bayesian Networks

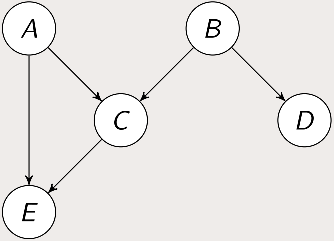
Probabilistic Graphical Models

Jerónimo Hernández-González

Exercise

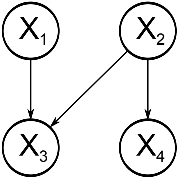
Independent parameters

How many independent parameters are required to uniquely define the CPD of E if A, B, and D are binary, and C and E have three values each?



- 8
- 11
- 17
- 3
- 18
- 12
- 6

Graph



Set of probability distributions, $p(X_i|\boldsymbol{PA}_i)$

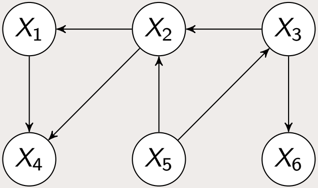
X_i	$\lvert\Omega_{X_i}\rvert$	$\boldsymbol{PA}_i$	$\lvert\Omega_{\boldsymbol{PA}_i}\rvert$
X_1	3	$\emptyset$	-
X_2	2	$\emptyset$	-
X_3	2	(X_1, X_2)	6
X_4	2	X_2	2

$$\text{No. params} = \sum_i (|\Omega_{X_i}| - 1) \times |\Omega_{\boldsymbol{PA}_i}|$$

Exercise

Storing a Bayesian network

Let $P(X)$ be a Bayesian network over $\mathcal{G}$. How much memory do we need to store it?



Bayesian networks

Number of parameters

X_j	Model parameters
X_1	$\theta_{1\emptyset 1} = p(x_{11}), \theta_{1\emptyset 2} = p(x_{12}),\ \theta_{1\emptyset 3} = p(x_{13})$
X_2	$\theta_{2\emptyset 1} = p(x_{21}),\ \theta_{2\emptyset 2} = p(x_{22})$
X_3	$\theta_{311} = p(x_{31}\|x_{11}, x_{21}), \theta_{321} = p(x_{31}\|x_{11}, x_{22}), \theta_{331} = p(x_{31}\|x_{12}, x_{21}),$
	$\theta_{341} = p(x_{31}\|x_{12}, x_{22}), \theta_{351} = p(x_{31}\|x_{13}, x_{21}), \theta_{361} = p(x_{31}\|x_{13}, x_{22}),$
	$\theta_{312} = p(x_{32}\|x_{11}, x_{21}), \theta_{322} = p(x_{32}\|x_{11}, x_{22}), \theta_{332} = p(x_{32}\|x_{12}, x_{21}),$
	$\theta_{342} = p(x_{32}\|x_{12}, x_{22}), \theta_{352} = p(x_{32}\|x_{13}, x_{21}), \theta_{362} = p(x_{32}\|x_{13}, x_{22})$
X_4	$\theta_{411} = p(x_{41}\|x_{21}), \theta_{421} = p(x_{41}\|x_{22}),$
	$\theta_{412} = p(x_{42}\|x_{21}), \theta_{422} = p(x_{42}\|x_{22})$

Bayesian networks

Number of parameters

X_j	Model parameters
X_1	$\theta_{1\emptyset 1} = p(x_{11}), \theta_{1\emptyset 2} = p(x_{12}), \theta_{1\emptyset 3} = p(x_{13})$
X_2	$\theta_{2\emptyset 1} = p(x_{21}), \theta_{2\emptyset 2} = p(x_{22})$
X_3	$\theta_{311} = p(x_{31}\|x_{11}, x_{21}), \theta_{321} = p(x_{31}\|x_{11}, x_{22}), \theta_{331} = p(x_{31}\|x_{12}, x_{21}),$
	$\theta_{341} = p(x_{31}\|x_{12}, x_{22}), \theta_{351} = p(x_{31}\|x_{13}, x_{21}), \theta_{361} = p(x_{31}\|x_{13}, x_{22}),$
	$\theta_{312} = p(x_{32}\|x_{11}, x_{21}), \theta_{322} = p(x_{32}\|x_{11}, x_{22}), \theta_{332} = p(x_{32}\|x_{12}, x_{21}),$
	$\theta_{342} = p(x_{32}\|x_{12}, x_{22}), \theta_{352} = p(x_{32}\|x_{13}, x_{21}), \theta_{362} = p(x_{32}\|x_{13}, x_{22})$
X_4	$\theta_{411} = p(x_{41}\|x_{21}), \theta_{421} = p(x_{41}\|x_{22}),$
	$\theta_{412} = p(x_{42}\|x_{21}), \theta_{422} = p(x_{42}\|x_{22})$

Bayesian networks

Number of parameters

X_j	Model parameters
X_1	$\theta_{1\emptyset 1} = p(x_{11}), \theta_{1\emptyset 2} = p(x_{12}), \theta_{1\emptyset 3} = p(x_{13})$
X_2	$\theta_{2\emptyset 1} = p(x_{21}), \theta_{2\emptyset 2} = p(x_{22})$
X_3	$\theta_{311} = p(x_{31}\|x_{11}, x_{21}), \theta_{321} = p(x_{31}\|x_{11}, x_{22}), \theta_{331} = p(x_{31}\|x_{12}, x_{21}),$
	$\theta_{341} = p(x_{31}\|x_{12}, x_{22}), \theta_{351} = p(x_{31}\|x_{13}, x_{21}), \theta_{361} = p(x_{31}\|x_{13}, x_{22}),$
	$\theta_{312} = p(x_{32}\|x_{11}, x_{21}), \theta_{322} = p(x_{32}\|x_{11}, x_{22}), \theta_{332} = p(x_{32}\|x_{12}, x_{21}),$
	$\theta_{342} = p(x_{32}\|x_{12}, x_{22}), \theta_{352} = p(x_{32}\|x_{13}, x_{21}), \theta_{362} = p(x_{32}\|x_{13}, x_{22})$
X_4	$\theta_{411} = p(x_{41}\|x_{21}), \theta_{421} = p(x_{41}\|x_{22}),$
	$\theta_{412} = p(x_{42}\|x_{21}), \theta_{422} = p(x_{42}\|x_{22})$

X_1	$p(X_1)$
a	0,4
b	0,3
c	0,3

X_3	X_1	X_2	$p(X_3\|X_1, X_2)$	X_3	X_1	X_2	$p(X_3\|X_1, X_2)$
a	a	a	0,55	a	b	b	0,30
b	a	a	0,45	b	b	b	0,70
a	a	b	0,40	a	c	a	0,80
b	a	b	0,60	b	c	a	0,20
a	b	a	0,35	a	c	b	0,25
b	b	a	0,65	b	c	b	0,75

X_4	X_2	$p(X_4\|X_2)$
a	a	0,25
b	a	0,75
a	b	0,66
b	b	0,33

Bayesian networks

Number of parameters

X_j	Model parameters
X_1	$\theta_{1\emptyset 1} = p(x_{11}), \theta_{1\emptyset 2} = p(x_{12}), \theta_{1\emptyset 3} = p(x_{13})$
X_2	$\theta_{2\emptyset 1} = p(x_{21}), \theta_{2\emptyset 2} = p(x_{22})$
X_3	$\theta_{311} = p(x_{31}\|x_{11}, x_{21}), \theta_{321} = p(x_{31}\|x_{11}, x_{22}), \theta_{331} = p(x_{31}\|x_{12}, x_{21}),$
	$\theta_{341} = p(x_{31}\|x_{12}, x_{22}), \theta_{351} = p(x_{31}\|x_{13}, x_{21}), \theta_{361} = p(x_{31}\|x_{13}, x_{22}),$
	$\theta_{312} = p(x_{32}\|x_{11}, x_{21}), \theta_{322} = p(x_{32}\|x_{11}, x_{22}), \theta_{332} = p(x_{32}\|x_{12}, x_{21}),$
	$\theta_{342} = p(x_{32}\|x_{12}, x_{22}), \theta_{352} = p(x_{32}\|x_{13}, x_{21}), \theta_{362} = p(x_{32}\|x_{13}, x_{22})$
X_4	$\theta_{411} = p(x_{41}\|x_{21}), \theta_{421} = p(x_{41}\|x_{22}),$
	$\theta_{412} = p(x_{42}\|x_{21}), \theta_{422} = p(x_{42}\|x_{22})$

X_1	$p(X_1)$
a	0,4$= \theta_{1\emptyset 1}$
b	0,3$= \theta_{1\emptyset 2}$
c	0,3

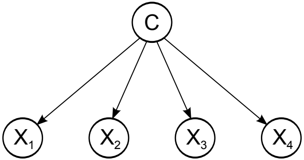
X_3	X_1	X_2	$p(X_3\|X_1, X_2)$	X_3	X_1	X_2	$p(X_3\|X_1, X_2)$
a	a	a	0,55$= \theta_{311}$	a	b	b	0,30$= \theta_{341}$
b	a	a	0,45	b	b	b	0,70
a	a	b	0,40$= \theta_{321}$	a	c	a	0,80$= \theta_{351}$
b	a	b	0,60	b	c	a	0,20
a	b	a	0,35$= \theta_{331}$	a	c	b	0,25$= \theta_{361}$
b	b	a	0,65	b	c	b	0,75

X_4	X_2	$p(X_4\|X_2)$
a	a	0,25$= \theta_{411}$
b	a	0,75
a	b	0,66$= \theta_{421}$
b	b	0,33

Naive Bayes

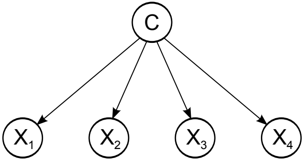
Parameters ($|\Omega_C| = 2$)

$$\theta_C = p(C = 1)$$



Naive Bayes

Parameters ($|\Omega_C| = 2$)

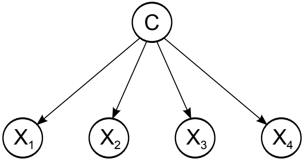


$$\theta_C = p(C = 1)$$

$$\theta_{X_1|C=0} = p(X_1 = 1|C = 0)$$
$$\theta_{X_1|C=1} = p(X_1 = 1|C = 1)$$

Naive Bayes

Parameters ($|\Omega_C| = 2$)



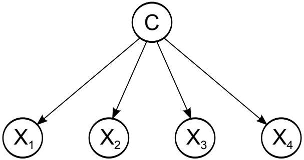
$$\theta_C = p(C = 1)$$

$$\theta_{X_1|C=0} = p(X_1 = 1|C = 0)$$
$$\theta_{X_1|C=1} = p(X_1 = 1|C = 1)$$

$$\theta_{X_2|C=0} = p(X_2 = 1|C = 0)$$
$$\theta_{X_2|C=1} = p(X_2 = 1|C = 1)$$

Naive Bayes

Parameters ($|\Omega_C| = 2$)



$$\theta_C = p(C = 1)$$

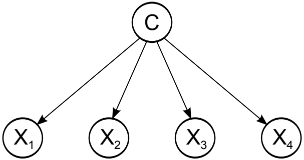
$$\theta_{X_1|C=0} = p(X_1 = 1|C = 0)$$
$$\theta_{X_1|C=1} = p(X_1 = 1|C = 1)$$

$$\theta_{X_2|C=0} = p(X_2 = 1|C = 0)$$
$$\theta_{X_2|C=1} = p(X_2 = 1|C = 1)$$

$$\theta_{X_3|C=0} = p(X_3 = 1|C = 0)$$
$$\theta_{X_3|C=1} = p(X_3 = 1|C = 1)$$

Naive Bayes

Parameters ($|\Omega_C| = 2$)



$$\theta_C = p(C = 1)$$

$$\theta_{X_1|C=0} = p(X_1 = 1|C = 0)$$
$$\theta_{X_1|C=1} = p(X_1 = 1|C = 1)$$

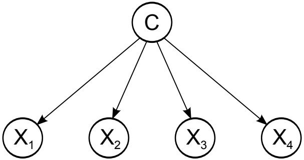
$$\theta_{X_2|C=0} = p(X_2 = 1|C = 0)$$
$$\theta_{X_2|C=1} = p(X_2 = 1|C = 1)$$

$$\theta_{X_3|C=0} = p(X_3 = 1|C = 0)$$
$$\theta_{X_3|C=1} = p(X_3 = 1|C = 1)$$

$$\theta_{X_4|C=0} = p(X_4 = 1|C = 0)$$
$$\theta_{X_4|C=1} = p(X_4 = 1|C = 1)$$

Naive Bayes

Parameters ($|\Omega_C| = 2$)



$$\theta_C = p(C = 1)$$

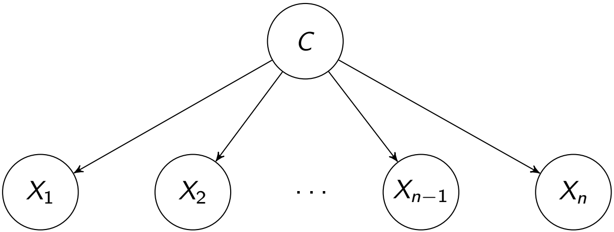
$$\theta_{X_1|C=0} = p(X_1 = 1|C = 0)$$
$$\theta_{X_1|C=1} = p(X_1 = 1|C = 1)$$

$$\theta_{X_2|C=0} = p(X_2 = 1|C = 0)$$
$$\theta_{X_2|C=1} = p(X_2 = 1|C = 1)$$

$$\theta_{X_3|C=0} = p(X_3 = 1|C = 0)$$
$$\theta_{X_3|C=1} = p(X_3 = 1|C = 1)$$

$$\theta_{X_4=1|C=0} = p(X_4 = 1|C = 0)$$
$$\theta_{X_4=2|C=0} = p(X_4 = 2|C = 0)$$
$$\theta_{X_4=1|C=1} = p(X_4 = 1|C = 1)$$
$$\theta_{X_4=2|C=1} = p(X_4 = 2|C = 1)$$

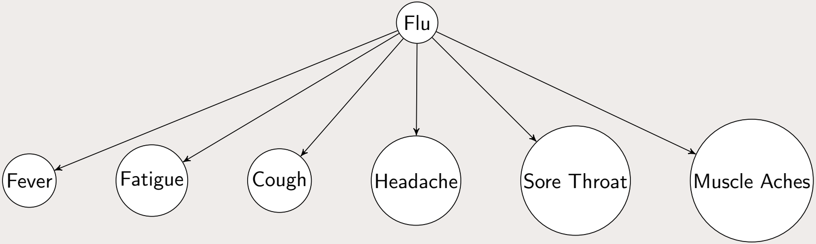
Naive Bayes



Exercise

Naive Bayes model for Flue diagnosis

Which of the following statements is/are true in this model?



Exercise

Naive Bayes model for Flue diagnosis

Which of the following statements is/are true in this model?

a) Say we observe that 1000 people have a headache (and possibly other symptoms), out of which 500 people have the flu (and possibly other symptoms), and 500 people have a fever (and possibly other symptoms). We would expect that approximately 250 people with a headache also have both the flu and a fever.

b) Say we observe that 500 people have a headache (and possibly other symptoms) and 500 people have a fever (and possibly other symptoms). We would expect that approximately 250 people have both a headache and fever.

c) Say we observe that 1000 people have the flu, out of which 500 people have a headache (and possibly other symptoms) and 500 have a fever (and possibly other symptoms). We would expect that approximately 250 people with the flu also have both a headache and fever.

d) Say we observe that 1000 people have the flu, out of which 500 people have a headache (and possibly other symptoms) and 500 have a fever (and possibly other symptoms). We can conclude that exactly 250 people with the flu also have both a headache and fever.

Bayesian networks summary

Model name	Bayesian network
Graph type	Directed Acyclic
Factorization	factor per variable (variable and its parents)
Independence	determined by d-separation in the graph
$F \implies I$	Always
$I \implies F$	Always
Markov blanket	Parents, children and parents of children

Bayesian Networks

Probabilistic Graphical Models

Jerónimo Hernández-González