

Statistical Machine Learning

Christian Walder

Machine Learning Research Group
CSIRO Data61

and

College of Engineering and Computer Science
The Australian National University

Canberra
Semester One, 2020.

(Many figures from C. M. Bishop, "Pattern Recognition and Machine Learning")



Outlines

- Overview
- Introduction
- Linear Algebra
- Probability
- Linear Regression 1
- Linear Regression 2
- Linear Classification 1
- Linear Classification 2
- Kernel Methods
- Sparse Kernel Methods
- Mixture Models and EM 1
- Mixture Models and EM 2
- Neural Networks 1
- Neural Networks 2
- Principal Component Analysis
- Autoencoders
- Graphical Models 1
- Graphical Models 2
- Graphical Models 3
- Sampling
- Sequential Data 1
- Sequential Data 2



Part XVII

Probabilistic Graphical Models 1

Motivation

Bayesian Network

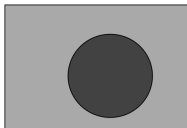
Plate Notation

*Conditional
Independence*



- **Image Segmentation**

Cluster the gray value representation of an image



- **Neighbourhood information lost**

Need to use the structure of the image.



Motivation

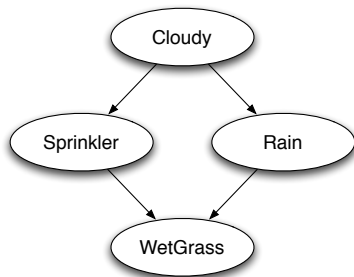
Bayesian Network

Plate Notation

Conditional
Independence



- Why is the grass wet?



- Introduce four Boolean variables :
 $C(loudy), S(prinkler), R(ain), W(etGrass) \in \{F(alse), T(rue)\}$.

Motivation

Bayesian Network

Plate Notation

Conditional
Independence



Motivation via Independence (2/4)

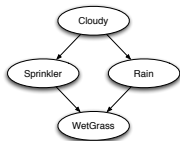
- Model the conditional probabilities

$p(C = F)$	$p(C = T)$
0.2	0.8

C	$p(S = F)$	$p(S = T)$
F	0.5	0.5
T	0.9	0.1

C	$p(R = F)$	$p(R = T)$
F	0.8	0.2
T	0.2	0.8

S R	$p(W = F)$	$p(W = T)$
F F	1.0	0.0
T F	0.1	0.9
F T	0.1	0.9
T T	0.01	0.99



Motivation

Bayesian Network

Plate Notation

Conditional Independence



Motivation

Bayesian Network

Plate Notation

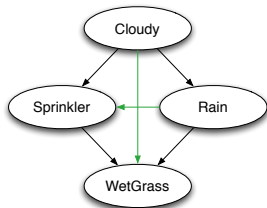
Conditional Independence

Motivation via Independence (3/4)

- If everything depends on everything

C	S	R	W	$p(C, S, R, W)$
F	F	F	F	...
F	F	F	T	...
...
T	T	T	F	...
T	T	T	T	...

$$\begin{aligned}
 p(W, S, R, C) &= p(W | S, R, C) p(S, R, C) \\
 &= p(W | S, R, C) p(S | R, C) p(R, C) \\
 &= p(W | S, R, C) p(S | R, C) p(R | C) p(C)
 \end{aligned}$$

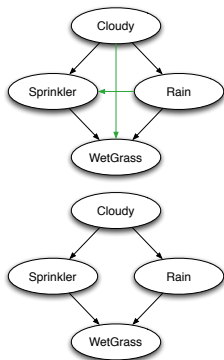


Motivation via Independence (4/4)



$$p(W) = \sum_{S,R,C} p(W | S, R, C) p(S | R, C) p(R | C) p(C)$$

$$p(W) = \sum_{S,R} p(W | S, R) \sum_C p(S | C) p(R | C) p(C)$$



Motivation

Bayesian Network

Plate Notation

*Conditional
Independence*



- Two key observations when dealing with probabilities

- 1 Distributive Law can save operations

$$\underbrace{a(b+c)}_{2 \text{ operations}} = \underbrace{ab+ac}_{3 \text{ operations}}$$

- 2 If some probabilities do not depend on all random variables, we might be able to factor them out. For example, assume

$$p(x_1, x_3 | x_2) = p(x_1 | x_2) p(x_3 | x_2),$$

then (using $\sum_{x_3} p(x_3 | x_2) = 1$)

$$\begin{aligned} p(x_1) &= \sum_{x_2, x_3} p(x_1, x_2, x_3) = \sum_{x_2, x_3} p(x_1, x_3 | x_2) p(x_2) \\ &= \underbrace{\sum_{x_2, x_3} p(x_1 | x_2) p(x_3 | x_2) p(x_2)}_{O(|\mathcal{X}_1||\mathcal{X}_2||\mathcal{X}_3|)} = \underbrace{\sum_{x_2} p(x_1 | x_2) p(x_2)}_{O(|\mathcal{X}_1||\mathcal{X}_2|)} \end{aligned}$$

Motivation

Bayesian Network

Plate Notation

Conditional
Independence

Motivation via Complexity Reduction



- How to deal with more complex expression?

$$p(x_1) p(x_2) p(x_3) p(x_4 | x_1, x_2, x_3) p(x_5 | x_1, x_3) p(x_6 | x_4) p(x_7 | x_4, x_5)$$

Motivation

Bayesian Network

Plate Notation

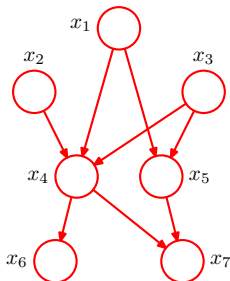
Conditional
Independence



- How to deal with more complex expression?

$$p(x_1) p(x_2) p(x_3) p(x_4|x_1, x_2, x_3) p(x_5|x_1, x_3) p(x_6|x_4) p(x_7|x_4, x_5)$$

- Graphical models



Motivation

Bayesian Network

Plate Notation

Conditional
Independence



Graphical models

- Visualise the structure of a probabilistic model
- Complex computations with formulas \rightarrow manipulations with graphs
- Obtain insights into model properties by inspection
- Develop and motivate new models

Motivation

Bayesian Network

Plate Notation

*Conditional
Independence*

Probabilistic Graphical Models: Main Flavours



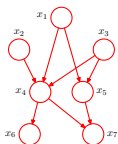
- Graph
 - 1 Nodes (vertices) : a random variable
 - 2 Edges (links, arcs; directed or undirected) : probabilistic relationship
- Directed Graph : **Bayesian Network** (also called Directed Graphical Model) expressing **causal** relationship between variables
- Undirected Graph : **Markov Random Field** expressing **soft constraints** between variables
- Factor Graph : convenient for solving **inference** problems (derived from Bayesian Networks or Markov Random Fields).

Motivation

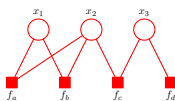
Bayesian Network

Plate Notation

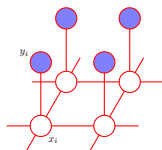
Conditional Independence



Bayesian Network



Factor Graph



Markov Random Field

Bayesian Network: Arbitrary Joint Distribution

$$p(a, b, c) = p(c | a, b) p(a, b) = p(c | a, b) p(b | a) p(a)$$



Motivation

Bayesian Network

Plate Notation

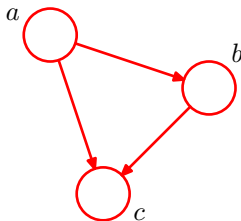
*Conditional
Independence*

Bayesian Network: Arbitrary Joint Distribution



$$p(a, b, c) = p(c | a, b) p(a, b) = p(c | a, b) p(b | a) p(a)$$

- 1 Draw a node for each conditional distribution associated with a random variable.
- 2 Draw an edge **from** each conditional distribution associated with a random variable **to** all other conditional distributions which are conditioned on this variable.



- We have chosen a particular ordering of the variables !

Motivation

Bayesian Network

Plate Notation

Conditional
Independence



- General case for K variables

$$p(x_1, \dots, x_K) = p(x_K | x_1, \dots, x_{K-1}) \dots p(x_2 | x_1) p(x_1)$$

- The graph of this distribution is **fully** connected. (Prove it.)
- What happens if we deal with a distribution represented by a graph which is **not** fully connected?
- Can not be the most general distribution anymore.
- The **absence** of edges carries important information.

Motivation

Bayesian Network

Plate Notation

*Conditional
Independence*

Joint Factorisation \rightarrow Graph

$$p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$



Motivation

Bayesian Network

Plate Notation

*Conditional
Independence*



Joint Factorisation \rightarrow Graph

$$p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$

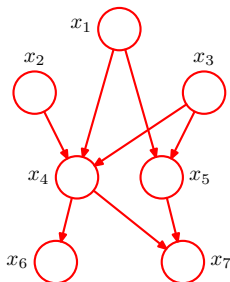
- 1 Draw a node for each conditional distribution associated with a random variable.
- 2 Draw an edge **from** each conditional distribution associated with a random variable **to** all other conditional distribution which are conditioned on this variable.



Joint Factorisation \rightarrow Graph

$$p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$

- 1 Draw a node for each conditional distribution associated with a random variable.
- 2 Draw an edge **from** each conditional distribution associated with a random variable **to** all other conditional distribution which are conditioned on this variable.



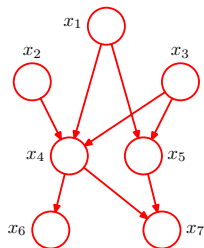
Motivation

Bayesian Network

Plate Notation

Conditional Independence

Graph \rightarrow Joint Factorisation



Can we get the expression from the graph?



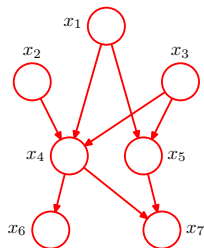
Motivation

Bayesian Network

Plate Notation

*Conditional
Independence*

Graph \rightarrow Joint Factorisation



Can we get the expression from the graph?

- 1 Write a product of probability distributions, one for each associated random variable. \leftrightarrow Draw a node for each conditional distribution associated with a random variable.
- 2 Add all random variables associated with parent nodes to the list of conditioning variables. \leftrightarrow Draw an edge **from** each conditional distribution associated with a random variable **to** all other conditional distribution which are conditioned on this variable.

$$p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)$$





- The joint distribution defined by a graph is given by the product, over all of the nodes of the graph, of a conditional distribution for each node conditioned on the variables corresponding to the parents of the node in the graph.

$$p(\mathbf{x}) = \prod_{k=1}^K p(x_k \mid \text{pa}(x_k))$$

where $\text{pa}(x_k)$ denotes the set of parents of x_k and $\mathbf{x} = (x_1, \dots, x_K)$.

- The **absence** of edges (compared to the fully connected case) is the point.

Motivation

Bayesian Network

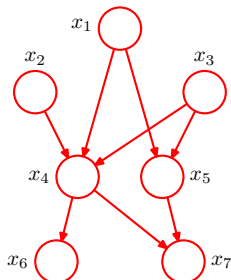
Plate Notation

*Conditional
Independence*

Joint Factorisation \leftrightarrow Graph



- Restriction : Graph must be a **directed acyclic graph** (DAG).
- There are no closed paths in the graph when moving along the directed edges.
- Or equivalently: There exists an ordering of the nodes such that there are no edges that go from any node to any lower numbered node.



- Extension: Can also have **sets** of variables, or **vectors** at a node.

Motivation

Bayesian Network

Plate Notation

*Conditional
Independence*

Bayesian Network: Normalisation



- Given

$$p(\mathbf{x}) = \prod_{k=1}^K p(x_k \mid \text{pa}(x_k)).$$

- Is $p(\mathbf{x})$ normalised, $\sum_{\mathbf{x}} p(\mathbf{x}) = 1$?

Motivation

Bayesian Network

Plate Notation

*Conditional
Independence*

Bayesian Network: Normalisation



- Given

$$p(\mathbf{x}) = \prod_{k=1}^K p(x_k \mid \text{pa}(x_k)).$$

- Is $p(\mathbf{x})$ normalised, $\sum_{\mathbf{x}} p(\mathbf{x}) = 1$?
- As graph is DAG, there always exists a node with no outgoing edges, say x_i .

Motivation

Bayesian Network

Plate Notation

Conditional Independence

$$\sum_{\mathbf{x}} p(\mathbf{x}) = \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_K} \prod_{\substack{k=1 \\ k \neq i}}^K p(x_k \mid \text{pa}(x_k)) \underbrace{\sum_{x_i} p(x_i \mid \text{pa}(x_i))}_{=1}$$

because $\sum_{x_i} p(x_i \mid \text{pa}(x_i)) = \sum_{x_i} \frac{p(x_i, \text{pa}(x_i))}{p(\text{pa}(x_i))} = \frac{p(\text{pa}(x_i))}{p(\text{pa}(x_i))} = 1$

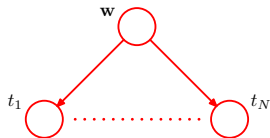
- Repeat, until no node left.

Bayesian Network - Plate Notation



- Bayesian polynomial regression : observed inputs \mathbf{x} , observed targets \mathbf{t} , noise variance σ^2 , hyperparameter α controlling the priors for \mathbf{w} .
- Focusing on \mathbf{t} and \mathbf{w} only

$$p(\mathbf{t}, \mathbf{w}) = p(\mathbf{w}) \prod_{n=1}^N p(t_n | \mathbf{w})$$



Motivation

Bayesian Network

Plate Notation

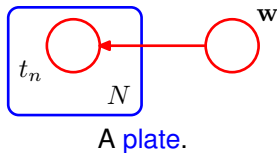
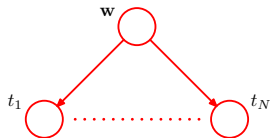
Conditional
Independence

Bayesian Network - Plate Notation



- Bayesian polynomial regression : observed inputs \mathbf{x} , observed targets \mathbf{t} , noise variance σ^2 , hyperparameter α controlling the priors for \mathbf{w} .
- Focusing on \mathbf{t} and \mathbf{w} only

$$p(\mathbf{t}, \mathbf{w}) = p(\mathbf{w}) \prod_{n=1}^N p(t_n | \mathbf{w})$$



Motivation

Bayesian Network

Plate Notation

Conditional Independence

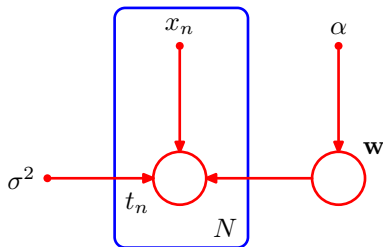


- Include also the parameters into the graphical model

$$p(\mathbf{t}, \mathbf{w} \mid \mathbf{x}, \alpha, \sigma^2) = p(\mathbf{w} \mid \alpha) \prod_{n=1}^N p(t_n \mid \mathbf{w}, x_n, \sigma^2)$$

Random variables = open circles

Deterministic variables = smaller solid circles



Motivation

Bayesian Network

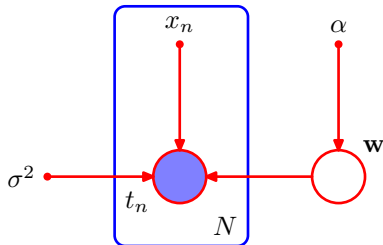
Plate Notation

Conditional
Independence

Bayesian Network - Plate Notation



- Random variables
 - Observed random variables, e.g. \mathbf{t}
 - Unobserved random variables, e.g. \mathbf{w} ,
(latent random variables, hidden random variables)
- Shade the observed random variables in the graphical model.



Motivation

Bayesian Network

Plate Notation

Conditional Independence



Motivation

Bayesian Network

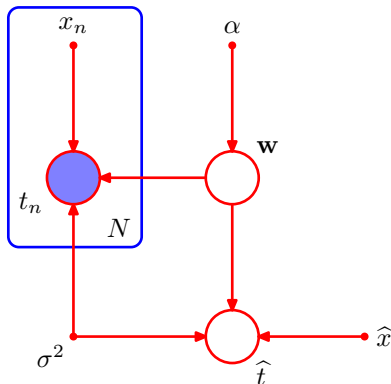
Plate Notation

Conditional Independence

Bayesian Network - Plate Notation

- Prediction : new data point \hat{x} . Want to predict \hat{t} .

$$p(\hat{t}, \mathbf{t}, \mathbf{w} | \hat{x}, \mathbf{x}, \alpha, \sigma^2) = \left[\prod_{n=1}^N p(t_n | x_n, \mathbf{w}, \sigma^2) \right] p(\mathbf{w} | \alpha) p(\hat{t} | \hat{x}, \mathbf{w}, \sigma^2)$$



Polynomial regression model including prediction.



- Any $p(\mathbf{x})$ may be written

$$\begin{aligned} p(\mathbf{x}) &= \prod_{k=1}^K p(x_k | x_1, x_2, \dots, x_{k-1}) \\ &= \prod_{k=1}^K p(x_k | \text{pa}(x_k)) \end{aligned}$$

- What does it mean if $\text{pa}(x_k) \neq x_1, x_2, \dots, x_{k-1}$ in the above equation?
 - Sparse graphical model.
 - $p(\mathbf{x})$ no longer general.
 - Assumption, e.g. $p(W|C, S, R) = p(W|S, R)$.
 - **Conditional independence.**

Motivation

Bayesian Network

Plate Notation

Conditional
Independence



Definition (Conditional Independence)

If for three random variables a , b , and c the following holds

$$p(a | b, c) = p(a | c)$$

then a is **conditionally independent** of b given c .

Notation : $a \perp\!\!\!\perp b | c$.

- The above equation must hold for all possible values of c .
- Consequence :

$$\begin{aligned} p(a, b | c) &= p(a | b, c) p(b | c) \\ &= p(a | c) p(b | c) \end{aligned}$$

- Conditional independence simplifies
 - the structure of the model
 - the computations needed to perform inference/learning.
- NB: **conditional dependence** is a related but different concept we are not concerned with in this course.

Motivation

Bayesian Network

Plate Notation

Conditional
Independence



Rules for Conditional Independence

Symmetry : $X \perp\!\!\!\perp Y | Z \implies Y \perp\!\!\!\perp X | Z$

Decomposition : $Y, W \perp\!\!\!\perp X | Z \implies Y \perp\!\!\!\perp X | Z$ and $W \perp\!\!\!\perp X | Z$

Weak Union : $X \perp\!\!\!\perp Y, W | Z \implies X \perp\!\!\!\perp Y | Z, W$

Contraction : $X \perp\!\!\!\perp W | Z, Y$
and $X \perp\!\!\!\perp Y | Z \implies X \perp\!\!\!\perp W, Y | Z$

Intersection : $X \perp\!\!\!\perp Y | Z, W$
and $X \perp\!\!\!\perp W | Z, Y \implies X \perp\!\!\!\perp Y, W | Z$

Note: Intersection is only valid for $p(X), p(Y), p(Z), p(W) > 0$.

Motivation

Bayesian Network

Plate Notation

Conditional
Independence

*Motivation**Bayesian Network**Plate Notation**Conditional
Independence*

Conditional Independence: TT (1/3)

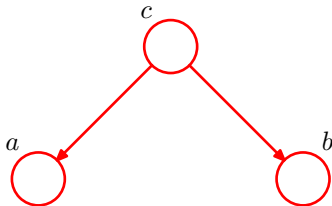
- Can we work with the graphical model directly?
- Check the simplest examples containing only three nodes.
- First example has joint distribution

$$p(a, b, c) = p(a | c) p(b | c) p(c)$$

- Marginalise both sides over c

$$p(a, b) = \sum_c p(a | c) p(b | c) p(c) \neq p(a) p(b).$$

- Does not hold : $a \perp\!\!\!\perp b | \emptyset$ (where \emptyset is the empty set).



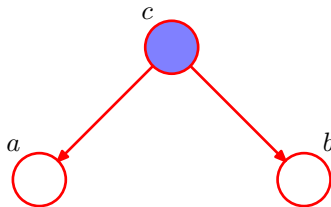
Conditional Independence: TT (2/3)



- Now condition on c .

$$p(a, b | c) = \frac{p(a, b, c)}{p(c)} = p(a | c) p(b | c)$$

- Therefore $a \perp\!\!\!\perp b | c$.



Motivation

Bayesian Network

Plate Notation

Conditional
Independence



Motivation

Bayesian Network

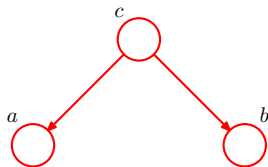
Plate Notation

Conditional Independence

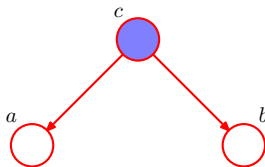
Conditional Independence: TT (3/3)

Graphical interpretation

- In both graphical models there is a **path** from a to b .
- The node c is called **tail-to-tail** (TT) with respect to this path because the node c is connected to the tails of the arrows in the path.
- The presence of the TT-node c in the path left renders a dependent on b (and b dependent on a).
- Conditioning on c **blocks** the path from a to b and causes a and b to become conditionally independent on c .



Not $a \perp\!\!\!\perp b \mid \emptyset$



$a \perp\!\!\!\perp b \mid c$



Motivation

Bayesian Network

Plate Notation

Conditional
Independence

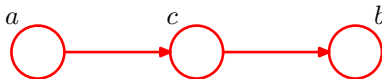
- Second example.

$$p(a, b, c) = p(a) p(c | a) p(b | c)$$

- Marginalise over c to test for independence.

$$p(a, b) = p(a) \sum_c p(c | a) p(b | c) = p(a) p(b | a) \neq p(a) p(b)$$

- Does not hold : $a \perp\!\!\!\perp b | \emptyset$.



Conditional Independence: HT (2/3)

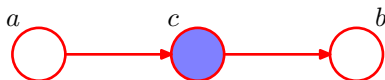


- Now condition on c .

$$p(a, b | c) = \frac{p(a, b, c)}{p(c)} = \frac{p(a)p(c|a)p(b|c)}{p(c)} = p(a|c)p(b|c)$$

where we used Bayes' theorem $p(c|a) = p(a|c)p(c)/p(a)$.

- Therefore $a \perp\!\!\!\perp b | c$.



Motivation

Bayesian Network

Plate Notation

Conditional
Independence

Conditional Independence: HT (3/3)

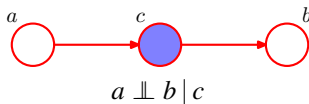
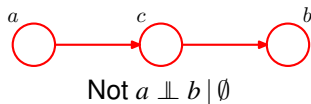


Motivation

Bayesian Network

Plate Notation

Conditional
Independence



*Motivation**Bayesian Network**Plate Notation**Conditional
Independence*

Conditional Independence: HH (1/4)

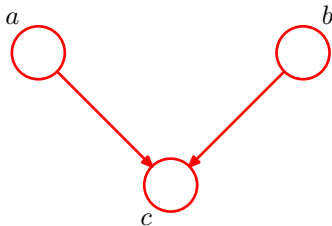
- Third example. (A little bit more subtle.)

$$p(a, b, c) = p(a) p(b) p(c | a, b)$$

- Marginalise over c to test for independence.

$$\begin{aligned} p(a, b) &= \sum_c p(a) p(b) p(c | a, b) = p(a) p(b) \sum_c p(c | a, b) \\ &= p(a) p(b) \end{aligned}$$

- a and b are independent if NO variable is observed:
 $a \perp\!\!\!\perp b \mid \emptyset$.



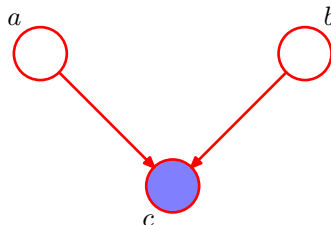
Conditional Independence: HH (2/4)



- Now condition on c .

$$p(a, b | c) = \frac{p(a, b, c)}{p(c)} = \frac{p(a)p(b)p(c | a, b)}{p(c)} \neq p(a | c)p(b | c).$$

- Does not hold : $a \perp\!\!\!\perp b | c$.



Motivation

Bayesian Network

Plate Notation

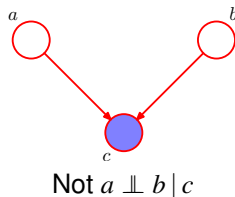
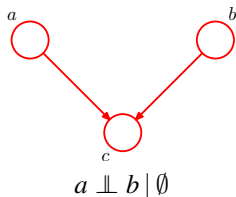
Conditional
Independence

Conditional Independence: HH (3/4)



Graphical interpretation

- In both graphical models there is a **path** from a to b .
- The node c is called **head-to-head** (HH) with respect to this path because the node c is connected to the heads of the arrows in the path.
- The presence of the HH-node c in the path left makes a independent of b (and b independent of a). The unobserved c **blocks** the path from a to b .



Motivation

Bayesian Network

Plate Notation

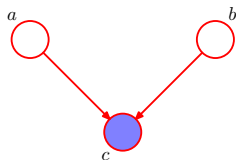
Conditional
Independence

Conditional Independence: HH (4/4)

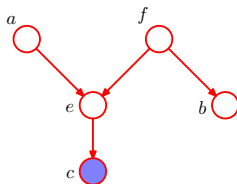


Graphical interpretation

- Conditioning on c **unblocks** the path from a to b , and renders a conditionally dependent on b given c .
- Some more terminology: Node y is a **descendant** of node x if there is a path from x to y in which each step follows the directions of the arrows.
- A HH-path will become unblocked if either the node, **or any of its descendants**, is observed.



Not $a \perp\!\!\!\perp b \mid c$



Not $a \perp\!\!\!\perp f \mid c$

Motivation

Bayesian Network

Plate Notation

Conditional Independence



- Conditional independence and has been established for
 - all configurations of three variables: (HH, HT, TT) \times (observed, unobserved)
 - the subtle case of HH junction with observed descendent.
- We can generalise these results to arbitrary Bayesian Networks.
- Roughly: the graph connectivity implies conditional independence for those sets of nodes which are **directionally separated**.

Motivation

Bayesian Network

Plate Notation

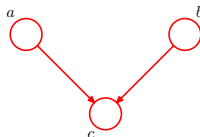
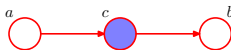
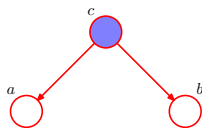
*Conditional
Independence*



Definition (Blocked Path)

A blocked path is a path which contains

- an observed TT- or HT-node, or
- an unobserved HH-node whose descendants are all unobserved.



Motivation

Bayesian Network

Plate Notation

Conditional
Independence



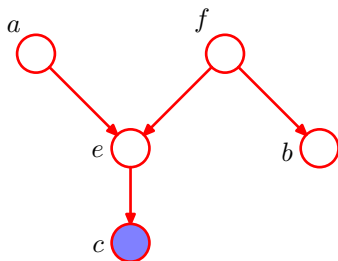
- Consider a general directed graph in which A , B , and C are arbitrary non-intersecting sets of nodes. (There may be other nodes in the graph which are not contained in the union of A , B , and C .)
- Consider all possible paths from any node in A to any node in B .
- Any such path is blocked, if it includes a node such that either
 - the node is HT or TT, and the node is in set C , or
 - the node is HH, and neither the node, nor any of the descendants, is in set C .
- If all paths are blocked, then A is d -separated from B by C , and the joint distribution over all the variables in the graph will satisfy $A \perp\!\!\!\perp B \mid C$.

(Note: D -separation stands for 'directional' separation.)



Example

- The path from a to b is not blocked by f because f is a TT-node and unobserved.
- The path from a to b is not blocked by e because e is a HH-node, and although unobserved itself, one of its descendants (node c) is observed.



Motivation

Bayesian Network

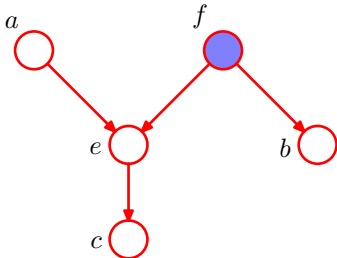
Plate Notation

Conditional
Independence



Another example

- The path from a to b is blocked by f because f is a TT-node and observed. Therefore, $a \perp\!\!\!\perp b \mid f$.
- Furthermore, the path from a to b is also blocked by e because e is a HH-node, and neither it nor its descendants are observed. Therefore $a \perp\!\!\!\perp b \mid f$.



Motivation

Bayesian Network

Plate Notation

Conditional Independence

Conditional Independence \Leftrightarrow Factorisation (1/4)



Theorem (Factorisation \Rightarrow Conditional Independence)

If a probability distribution factorises according to a directed acyclic graph, and if A , B and C are disjoint subsets of nodes such that A is d -separated from B by C in the graph, then the distribution satisfies $A \perp\!\!\!\perp B \mid C$.

Theorem (Conditional Independence \Rightarrow Factorisation)

If a probability distribution satisfies the conditional independence statements implied by d -separation over a particular directed graph, then it also factorises according to the graph.

Motivation

Bayesian Network

Plate Notation

Conditional
Independence

Conditional Independence \Leftrightarrow Factorisation (2/4)



Why is Conditional Independence \Leftrightarrow Factorisation relevant?

- Conditional Independence statements are usually what a domain expert knows about the problem at hand.
- Needed is a model $p(\mathbf{x})$ for computation.
- The Conditional Independence \Rightarrow Factorisation provides $p(x)$ from Conditional Independence statements.
- One can build a global model for computation from local conditional independence statements.

Motivation

Bayesian Network

Plate Notation

*Conditional
Independence*

Conditional Independence \Leftrightarrow Factorisation (3/4)



- Given a set of Conditional Independence statements.
- Adding another statement will in general produce other statements.
- All statements can be read as d -separation in a DAG.
- However, there are sets of Conditional Independence statements which **cannot** be satisfied by **any** Bayesian Network.

Motivation

Bayesian Network

Plate Notation

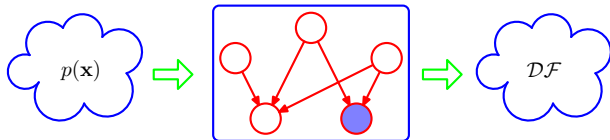
Conditional
Independence

Conditional Independence \Leftrightarrow Factorisation (4/4)



The broader picture

- A directed graphical model can be viewed as a filter accepting probability distributions $p(\mathbf{x})$ and only letting these through which satisfy the factorisation property. The set of all possible distribution $p(\mathbf{x})$ which pass through the filter is denoted as \mathcal{DF} .
- Alternatively, only these probability distributions $p(\mathbf{x})$ pass through the filter (graph), which respect the conditional independencies implied by the d-separation properties of the graph.
- The d-separation theorem says that the resulting set \mathcal{DF} is the same in both cases.



Motivation

Bayesian Network

Plate Notation

Conditional
Independence