Fundamentals of Probabilistic Data Mining

Chapter II - Probabilistic Graphical Models

Xavier Alameda-Pineda

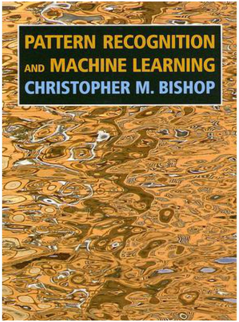
Ensimag/Inria

Table of Today's Contents

References

There is **a lot** of bibliography on probabilistic graphical models.

I strongly suggest the following book:



Pattern Recognition and Machine Learning, from Christopher M. Bishop (Springer)

The concepts discussed in FPDM correspond to different parts of Ch. 2, 8, 9, 10, 12, 13.

Conditional Dependency

Conditional Independence: reminder

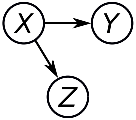
Definition

Let X, Y, and Z be random variables, we say that X and Y are **conditionally independent** given Z, and write $X \perp\!\!\!\perp Y \mid Z$, iff one of the following equivalent expressions holds:

- $p(x, y|z) = p(x|z)p(y|z)$
- $p(x|y, z) = p(x|z)$
- $p(y|x, z) = p(y|z)$

Proof of the equivalence: Any questions?

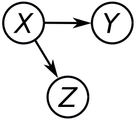
Conditional Independence: reminder (2)



Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x)$:

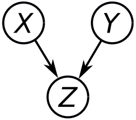
$$p(y, z|x) = p(y|x)p(z|x) \quad \text{and} \quad p(y, z) \neq p(y)p(z).$$

Conditional Independence: reminder (2)



Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x)$:

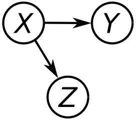
$$p(y, z|x) = p(y|x)p(z|x) \quad \text{and} \quad p(y, z) \neq p(y)p(z).$$



Two-parents $p(x, y, z) = p(z|x, y)p(y)p(x)$:

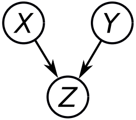
$$p(x, y|z) \neq p(x|z)p(y|z) \quad \text{and} \quad p(x, y) = p(x)p(y).$$

Conditional Independence: reminder (2)



Two-kids $p(x, y, z) = p(z|x)p(y|x)p(x)$:

$$p(y, z|x) = p(y|x)p(z|x) \quad \text{and} \quad p(y, z) \neq p(y)p(z).$$



Two-parents $p(x, y, z) = p(z|x, y)p(y)p(x)$:

$$p(x, y|z) \neq p(x|z)p(y|z) \quad \text{and} \quad p(x, y) = p(x)p(y).$$

Never forget that...

"Independence" and "Conditional Independence" are **two different** things.

Conditional Independence: 3 Gaussian variables

Let X, Y and Z three random variables following:

$$\begin{aligned} X &= \varepsilon_x, \\ Y &= \varepsilon_y + pX, \\ Z &= \varepsilon_z + cX + kY, \end{aligned} \qquad \text{with} \qquad \begin{aligned} &\epsilon_*, \sim \mathcal{N}(\mu_*, \nu_*), \\ &(* \text{ is } x,\ y, \text{ or } z.) \\ &p, c, k \in \mathbb{R}. \end{aligned}$$

Conditional Independence: 3 Gaussian variables

Let X, Y and Z three random variables following:

$$\begin{aligned} X &= \varepsilon_x, \\ Y &= \varepsilon_y + pX, \\ Z &= \varepsilon_z + cX + kY, \end{aligned} \qquad \text{with} \qquad \begin{aligned} &\epsilon_*, \sim \mathcal{N}(\mu_*, \nu_*), \\ &(* \text{ is } x,\ y, \text{ or } z.) \\ &p, c, k \in \mathbb{R}. \end{aligned}$$

Then, the joint probability $p(x, y, z)$ is a **multivariate Gaussian**:

$$p(x, y, z) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\Big(-\frac{1}{2}(v-\mu)^\top \Sigma^{-1}(v-\mu)\Big),$$

where $v = (x, y, z)^\top$ is the joint vector and Σ and μ are the so-called **covariance matrix** and **mean vector**.
($\rightarrow$ check Bishop's book, section 2.3 until 2.3.4).

Homework: (i) prove that you obtain such a Gaussian, (ii) compute μ.

Conditional Independence: 3 Gaussian variables (II)

Σ in the previous equation takes the following form:

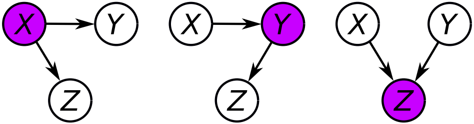
$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\nu_x} + \frac{p^2}{\nu_y} + \frac{c^2}{\nu_z} & \frac{p}{\nu_y} & \frac{c}{\nu_z} \\ \frac{p}{\nu_y} & \frac{1}{\nu_y} + \frac{k^2}{\nu_z} & \frac{k}{\nu_z} \\ \frac{c}{\nu_z} & \frac{k}{\nu_z} & \frac{1}{\nu_z} \end{pmatrix}$$

Conditional Independence: 3 Gaussian variables (II)

Σ in the previous equation takes the following form:

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\nu_x} + \frac{p^2}{\nu_y} + \frac{c^2}{\nu_z} & \frac{p}{\nu_y} & \frac{c}{\nu_z} \\ \frac{p}{\nu_y} & \frac{1}{\nu_y} + \frac{k^2}{\nu_z} & \frac{k}{\nu_z} \\ \frac{c}{\nu_z} & \frac{k}{\nu_z} & \frac{1}{\nu_z} \end{pmatrix}$$

For which values of p, k, c you obtain the models:
two-parents, two-kids and cascaded ? You've got 5 minutes.

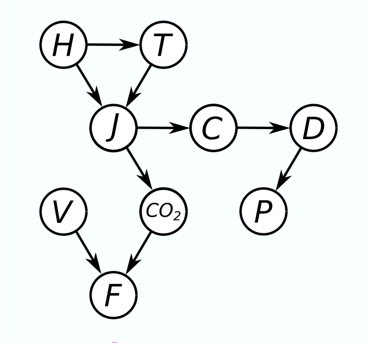


Homework: prove the expression for Σ.

D-separation

D-separation: motivation

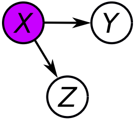
Recall...



Is $P \perp\!\!\!\perp V \mid T$? How would you do it ? Is this strategy scalable ?

D-separation: basics

Let us recall the 3-var models:

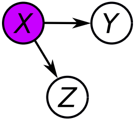


Two kids The path from Z to Y is called "tail-to-tail."

$$p(z, y|x) = p(z|x)p(y|x)$$

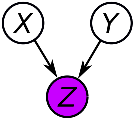
D-separation: basics

Let us recall the 3-var models:



Two kids The path from Z to Y is called "tail-to-tail."

$$p(z, y|x) = p(z|x)p(y|x)$$

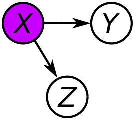


Two parents The path from X to Y is called "head-to-head."

$$p(x, y|z) \neq p(x|z)p(y|z)$$

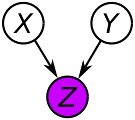
D-separation: basics

Let us recall the 3-var models:



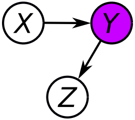
Two kids The path from Z to Y is called “tail-to-tail.”

$$p(z, y|{\color{purple}x}) = p(z|{\color{purple}x})p(y|{\color{purple}x})$$



Two parents The path from X to Y is called “head-to-head.”

$$p(x, y|{\color{purple}z}) \neq p(x|{\color{purple}z})p(y|{\color{purple}z})$$

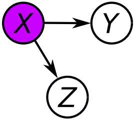


Cascaded The path from X to Z is called “head-to-tail.”

$$p(x, z|{\color{purple}y}) = p(x|{\color{purple}y})p(z|{\color{purple}y})$$

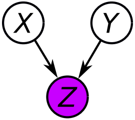
D-separation: basics

Let us recall the 3-var models:



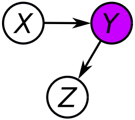
Two kids The path from Z to Y is called "tail-to-tail."

$$p(z, y|x) = p(z|x)p(y|x)$$



Two parents The path from X to Y is called "head-to-head."

$$p(x, y|z) \neq p(x|z)p(y|z)$$



Cascaded The path from X to Z is called "head-to-tail."

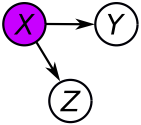
$$p(x, z|y) = p(x|y)p(z|y)$$

Please check Section 8.2 of Bishop's book.

D-separation: path blocking

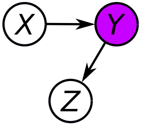
Two-kids

tail-to-tail



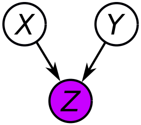
Cascaded

head-to-tail

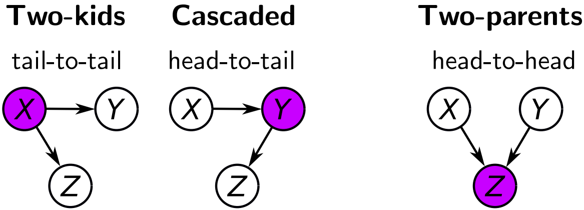


Two-parents

head-to-head



D-separation: path blocking

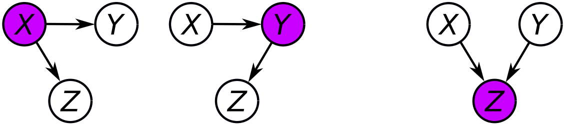


The purple node **"blocks" the path** in two-kids/tail-to-tail & cascaded/head-to-tail $\rightarrow$ conditional independence.

D-separation: path blocking

Two-kids **Cascaded** **Two-parents**

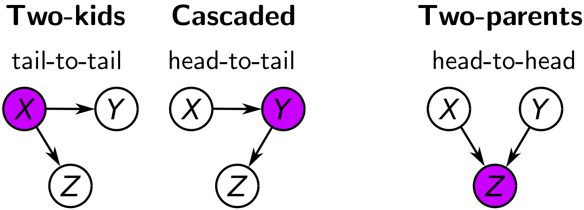
tail-to-tail head-to-tail head-to-head



The purple node **"blocks" the path** in two-kids/tail-to-tail & cascaded/head-to-tail $\rightarrow$ conditional independence.

The purple node **"unblocks" the path** in two-parents/head-to-head $\rightarrow$ conditional dependence.

D-separation: path blocking

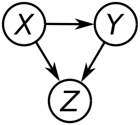


The purple node **"blocks" the path** in two-kids/tail-to-tail & cascaded/head-to-tail $\rightarrow$ conditional independence.

The purple node **"unblocks" the path** in two-parents/head-to-head $\rightarrow$ conditional dependence.
In the two-parents, Z or any descendant of Z will unblock the path.

D-separation: 6-dimensional Gaussian

[1D] Let X, Y and Z three random variables following:



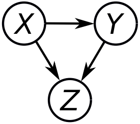
$$X = \varepsilon_x,$$
$$Y = \varepsilon_y + pX,$$
$$Z = \varepsilon_z + cX + kY,$$

with

$$\epsilon_*, \sim \mathcal{N}(\mu_*, \nu_*),$$
$$p, c, k \in \mathbb{R}.$$

D-separation: 6-dimensional Gaussian

[1D] Let X, Y and Z three random variables following:

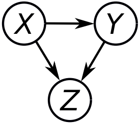


$$X = \varepsilon_x,$$
$$Y = \varepsilon_y + pX,$$
$$Z = \varepsilon_z + cX + kY,$$

with

$\epsilon_*, \sim \mathcal{N}(\mu_*, \nu_*)$,
$p, c, k \in \mathbb{R}$.

[2D] Let X, Y and Z three random vectors following:



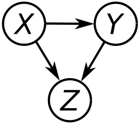
$$X = \varepsilon_x,$$
$$Y = \varepsilon_y + PX,$$
$$Z = \varepsilon_z + CX + KY,$$

with

$\epsilon_*, \sim \mathcal{N}(M_*, N_*)$,
$P, C, K \in \mathbb{R}^{2\times 2}$.

D-separation: 6-dimensional Gaussian

[1D] Let X, Y and Z three random variables following:

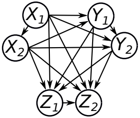


$$X = \varepsilon_x, \quad Y = \varepsilon_y + pX, \quad Z = \varepsilon_z + cX + kY,$$

with

$\epsilon_*, \sim \mathcal{N}(\mu_*, \nu_*)$,
$p, c, k \in \mathbb{R}$.

[2D] Let X, Y and Z three random vectors following:



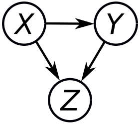
$$X = \varepsilon_x, \quad Y = \varepsilon_y + PX, \quad Z = \varepsilon_z + CX + KY,$$

with

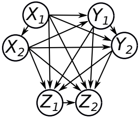
$\epsilon_*, \sim \mathcal{N}(M_*, N_*)$,
$P, C, K \in \mathbb{R}^{2\times 2}$.
$X = (X_1, X_2)^\top$
(same for Y and Z).

In this case, $M_* \in \mathbb{R}^2$ and $N_* \in \mathbb{R}^{2\times 2}$.
Each ϵ_* is a **multivariate Gaussian**.
$V = (X^\top, Y^\top, Z^\top)^\top$ is also a multivariate Gaussian.

D-separation: 6-dimensional Gaussian (II)

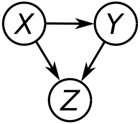


$$X = \varepsilon_x,$$
$$Y = \varepsilon_y + pX,$$
$$Z = \varepsilon_z + cX + kY,$$

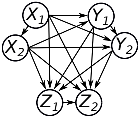


$$X = \varepsilon_x,$$
$$Y = \varepsilon_y + PX,$$
$$Z = \varepsilon_z + CX + KY,$$

D-separation: 6-dimensional Gaussian (II)



$$X = \varepsilon_x,$$
$$Y = \varepsilon_y + pX,$$
$$Z = \varepsilon_z + cX + kY,$$

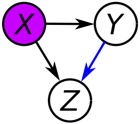


$$X = \varepsilon_x,$$
$$Y = \varepsilon_y + PX,$$
$$Z = \varepsilon_z + CX + KY,$$

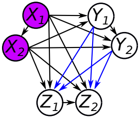
The covariance matrix Σ in the 2D-case writes (check Bishop's book):

$$\Sigma^{-1} = \begin{pmatrix} N_x^{-1} + P^\top N_y^{-1} P + C^\top N_z^{-1} C & N_y^{-1} P & N_z^{-1} C \\ P^\top N_y^{-1} & N_y^{-1} + K^\top N_z^{-1} K & N_z^{-1} K \\ C^\top N_z^{-1} & K^\top N_z^{-1} & N_z^{-1} \end{pmatrix}$$

D-separation: 6-dimensional Gaussian (II)



$$X = \varepsilon_x, \quad Y = \varepsilon_y + pX, \quad Z = \varepsilon_z + cX + kY,$$



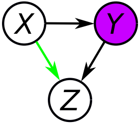
$$X = \varepsilon_x, \quad Y = \varepsilon_y + PX, \quad Z = \varepsilon_z + CX + KY,$$

The covariance matrix Σ in the 2D-case writes (check Bishop's book):

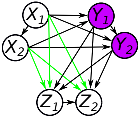
$$\Sigma^{-1} = \begin{pmatrix} N_x^{-1} + P^\top N_y^{-1} P + C^\top N_z^{-1} C & N_y^{-1} P & N_z^{-1} C \\ P^\top N_y^{-1} & N_y^{-1} + K^\top N_z^{-1} K & N_z^{-1} K \\ C^\top N_z^{-1} & K^\top N_z^{-1} & N_z^{-1} \end{pmatrix}$$

Two-kids (tail-to-tail): $p(y, z|x) = p(y|x)p(z|x)$

D-separation: 6-dimensional Gaussian (II)



$$X = \varepsilon_x, \quad Y = \varepsilon_y + pX, \quad Z = \varepsilon_z + cX + kY,$$



$$X = \varepsilon_x, \quad Y = \varepsilon_y + PX, \quad Z = \varepsilon_z + CX + KY,$$

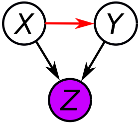
The covariance matrix Σ in the 2D-case writes (check Bishop's book):

$$\Sigma^{-1} = \begin{pmatrix} N_x^{-1} + P^\top N_y^{-1} P + C^\top N_z^{-1} C & N_y^{-1} P & N_z^{-1} C \\ P^\top N_y^{-1} & N_y^{-1} + K^\top N_z^{-1} K & N_z^{-1} K \\ C^\top N_z^{-1} & K^\top N_z^{-1} & N_z^{-1} \end{pmatrix}$$

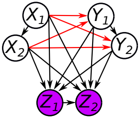
Two-kids (tail-to-tail): $p(y, z|x) = p(y|x)p(z|x)$

Cascaded (head-to-tail): $p(x, z|y) = p(x|y)p(z|y)$

D-separation: 6-dimensional Gaussian (II)



$$X = \varepsilon_x,\quad Y = \varepsilon_y + pX,\quad Z = \varepsilon_z + cX + kY,$$



$$X = \varepsilon_x,\quad Y = \varepsilon_y + PX,\quad Z = \varepsilon_z + CX + KY,$$

The covariance matrix Σ in the 2D-case writes (check Bishop's book):

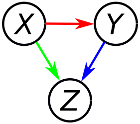
$$\Sigma^{-1} = \begin{pmatrix} N_x^{-1} + P^\top N_y^{-1} P + C^\top N_z^{-1} C & N_y^{-1} P & N_z^{-1} C \\ P^\top N_y^{-1} & N_y^{-1} + K^\top N_z^{-1} K & N_z^{-1} K \\ C^\top N_z^{-1} & K^\top N_z^{-1} & N_z^{-1} \end{pmatrix}$$

Two-kids (tail-to-tail): $p(y, z|x) = p(y|x)p(z|x)$

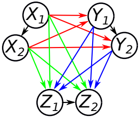
Cascaded (head-to-tail): $p(x, z|y) = p(x|y)p(z|y)$

Two-parents (head-to-head): $p(x, y|z) \neq p(x|z)p(y|z)$

D-separation: 6-dimensional Gaussian (II)



$$X = \varepsilon_x, \quad Y = \varepsilon_y + pX, \quad Z = \varepsilon_z + cX + kY,$$



$$X = \varepsilon_x, \quad Y = \varepsilon_y + PX, \quad Z = \varepsilon_z + CX + KY,$$

The covariance matrix Σ in the 2D-case writes (check Bishop's book):

$$\Sigma^{-1} = \begin{pmatrix} N_x^{-1} + P^\top N_y^{-1} P + C^\top N_z^{-1} C & N_y^{-1} P & N_z^{-1} C \\ P^\top N_y^{-1} & N_y^{-1} + K^\top N_z^{-1} K & N_z^{-1} K \\ C^\top N_z^{-1} & K^\top N_z^{-1} & N_z^{-1} \end{pmatrix}$$

Two-kids (tail-to-tail): $p(y, z|x) = p(y|x)p(z|x)$

Cascaded (head-to-tail): $p(x, z|y) = p(x|y)p(z|y)$

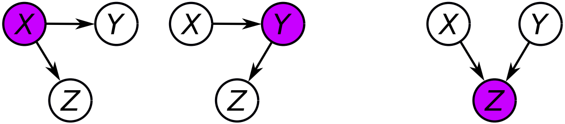
Two-parents (head-to-head): $p(x, y|z) \neq p(x|z)p(y|z)$

This gives us hope for more complicated models!

D-separation: path blocking (revisited)

Remember ?

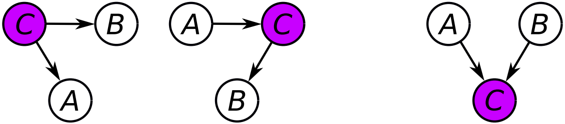
Two-kids tail-to-tail | **Cascaded** head-to-tail | **Two-parents** head-to-head



D-separation: path blocking (revisited)

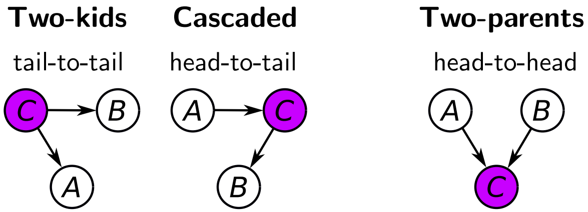
Remember ? Let me change the variable names...

Two-kids — tail-to-tail | **Cascaded** — head-to-tail | **Two-parents** — head-to-head



D-separation: path blocking (revisited)

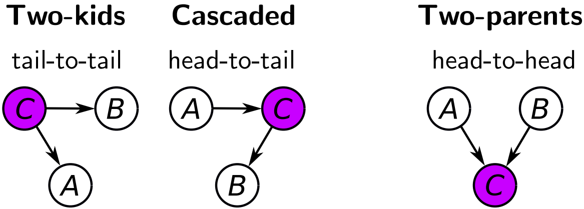
Remember ? Let me change the variable names...



Tail-to-tail & head-to-tail $\rightarrow A \perp\!\!\!\perp B \mid C$.

D-separation: path blocking (revisited)

Remember ? Let me change the variable names...

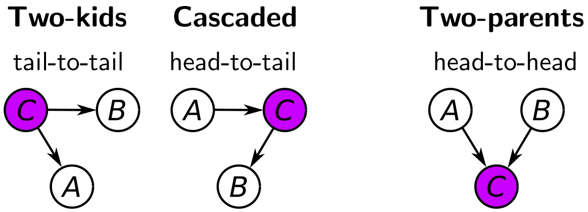


Tail-to-tail & head-to-tail $\rightarrow A \perp\!\!\!\perp B \mid C$.

Head-to-head $\rightarrow A \not\perp\!\!\!\perp B \mid C$ or any descendant of C.

D-separation: path blocking (revisited)

Remember ? Let me change the variable names...



Tail-to-tail & head-to-tail $\rightarrow A \perp\!\!\!\perp B \mid C$.

Head-to-head $\rightarrow A \not\perp\!\!\!\perp B \mid C$ or any descendant of C.

$\Rightarrow$ Nodes within tail-to-tail or head-to-tail can be in C and nodes within head-to-head or any of their descendents must not be in C.

D-separation: path blocking (definition)

Definition: blocked path

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A path from A to B is said to be blocked by C if it includes a node that either:

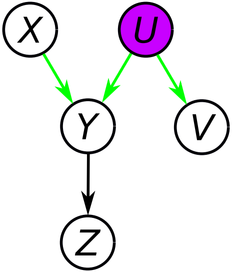
- the path meets tail-to-tail or head-to-tail at the node and the node is in C;
- the path meets head-to-head at the node and neither the node nor any of its descendants are in C.

D-separation: path blocking (definition)

Definition: blocked path

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A path from A to B is said to be blocked by C if it includes a node that either:

- the path meets tail-to-tail or head-to-tail at the node and the node is in C;
- the path meets head-to-head at the node and neither the node nor any of its descendants are in C.



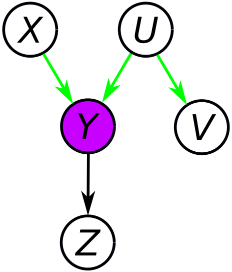
- Is the path from X to V blocked by U?

D-separation: path blocking (definition)

Definition: blocked path

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A path from A to B is said to be blocked by C if it includes a node that either:

- the path meets tail-to-tail or head-to-tail at the node and the node is in C;
- the path meets head-to-head at the node and neither the node nor any of its descendants are in C.



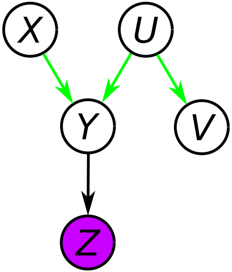
- Is the path from X to V blocked by U? Yes
- Is the path from X to V blocked by Y?

D-separation: path blocking (definition)

Definition: blocked path

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A path from A to B is said to be blocked by C if it includes a node that either:

- the path meets tail-to-tail or head-to-tail at the node and the node is in C;
- the path meets head-to-head at the node and neither the node nor any of its descendants are in C.



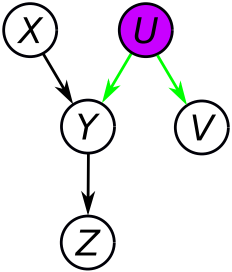
- Is the path from X to V blocked by U? Yes
- Is the path from X to V blocked by Y? No
- Is the path from X to V blocked by Y?

D-separation: path blocking (definition)

Definition: blocked path

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A path from A to B is said to be blocked by C if it includes a node that either:

- the path meets tail-to-tail or head-to-tail at the node and the node is in C;
- the path meets head-to-head at the node and neither the node nor any of its descendants are in C.



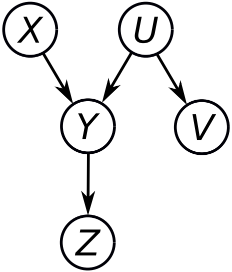
- Is the path from X to V blocked by U? Yes
- Is the path from X to V blocked by Y? No
- Is the path from X to V blocked by Y? No, because of Z.
- Is the path from Y to V blocked by U?

D-separation: path blocking (definition)

Definition: blocked path

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A path from A to B is said to be blocked by C if it includes a node that either:

- the path meets tail-to-tail or head-to-tail at the node and the node is in C;
- the path meets head-to-head at the node and neither the node nor any of its descendants are in C.



- Is the path from X to V blocked by U? Yes
- Is the path from X to V blocked by Y? No
- Is the path from X to V blocked by Y? No, because of Z.
- Is the path from Y to V blocked by U? Yes

D-separation: definition

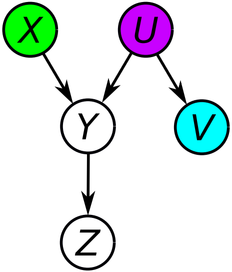
Definition: D-separation

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A and B are D-separated by C, if all paths from any node from A to B are blocked by C.

D-separation: definition

Definition: D-separation

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A and B are D-separated by C, if all paths from any node from A to B are blocked by C.

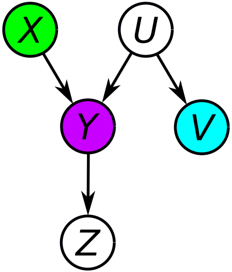


- Is $\{X\}$ D-separated from $\{V\}$ by $\{U\}$?

D-separation: definition

Definition: D-separation

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A and B are D-separated by C, if all paths from any node from A to B are blocked by C.

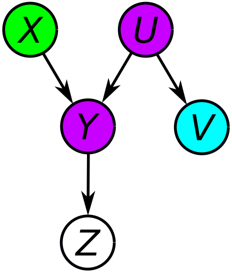


- Is $\{X\}$ D-separated from $\{V\}$ by $\{U\}$? Yes
- Is $\{X\}$ D-separated from $\{V\}$ by $\{Y\}$?

D-separation: definition

Definition: D-separation

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A and B are D-separated by C, if all paths from any node from A to B are blocked by C.

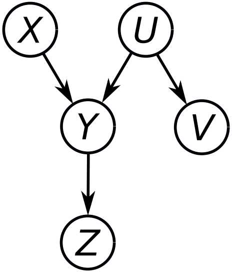


- Is $\{X\}$ D-separated from $\{V\}$ by $\{U\}$? Yes
- Is $\{X\}$ D-separated from $\{V\}$ by $\{Y\}$? No
- Is $\{X\}$ D-separated from $\{V\}$ by $\{Y, U\}$?

D-separation: definition

Definition: D-separation

Let A, B and C be three non-intersecting sets of nodes of a directed acyclic graph. A and B are D-separated by C, if all paths from any node from A to B are blocked by C.

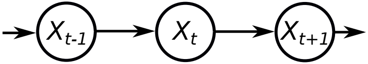


- Is $\{X\}$ D-separated from $\{V\}$ by $\{U\}$? Yes
- Is $\{X\}$ D-separated from $\{V\}$ by $\{Y\}$? No
- Is $\{X\}$ D-separated from $\{V\}$ by $\{Y, U\}$? Yes

Markovian dependencies

Markov models: introduction

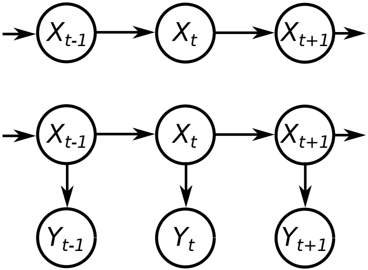
Principle: each variable depends **only** on its closer neighbours. Examples:



Markov chain.

Markov models: introduction

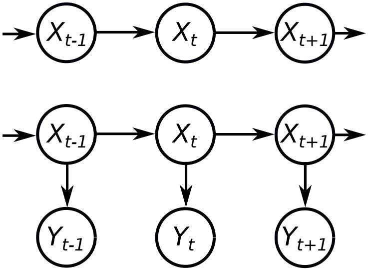
Principle: each variable depends **only** on its closer neighbours. Examples:



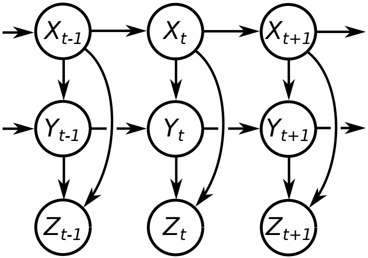
Markov chain (top).
Hidden Markov chain (bottom).

Markov models: introduction

Principle: each variable depends **only** on its closer neighbours. Examples:



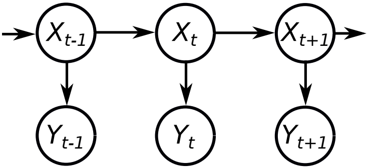
Markov chain (top).
Hidden Markov chain (bottom).



Double hidden Markov chain.

D-separation in Markov models

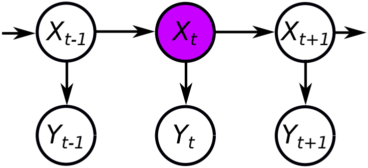
With the following model:



Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

D-separation in Markov models

With the following model:

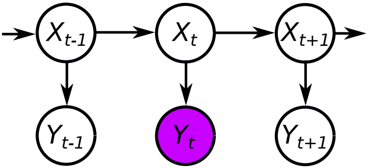


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? [1 minute]

D-separation in Markov models

With the following model:

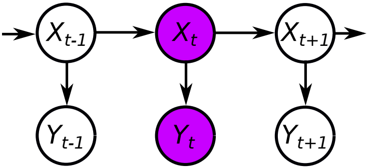


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? Yes
- $\{Y_t\}$? [1 minute]

D-separation in Markov models

With the following model:

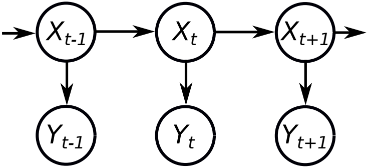


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? Yes
- $\{Y_t\}$? No
- $\{X_t, Y_t\}$? [1 minute]

D-separation in Markov models

With the following model:

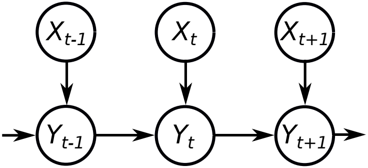


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? Yes
- $\{Y_t\}$? No
- $\{X_t, Y_t\}$? Yes

D-separation in Markov models (II)

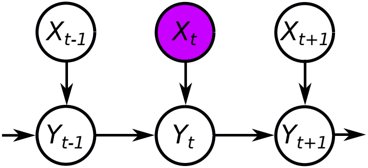
With the following model:



Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

D-separation in Markov models (II)

With the following model:

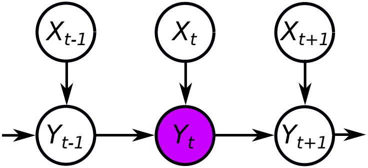


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? [1 minute]

D-separation in Markov models (II)

With the following model:

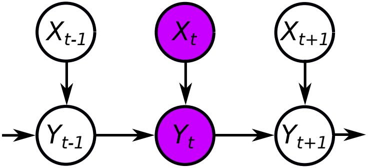


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? No
- $\{Y_t\}$? [1 minute]

D-separation in Markov models (II)

With the following model:

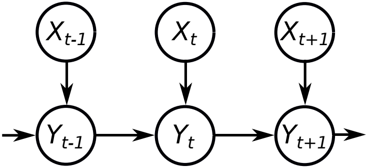


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? No
- $\{Y_t\}$? Yes
- $\{X_t, Y_t\}$? [1 minute]

D-separation in Markov models (II)

With the following model:

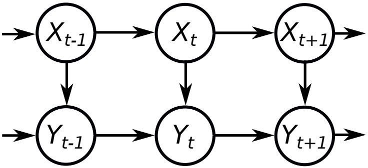


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? No
- $\{Y_t\}$? Yes
- $\{X_t, Y_t\}$? Yes

D-separation in Markov models (III)

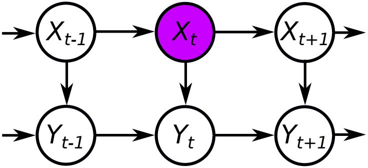
With the following model:



Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

D-separation in Markov models (III)

With the following model:

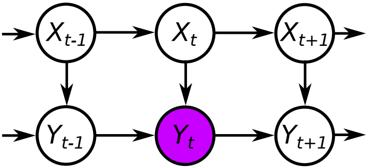


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? [1 minute]

D-separation in Markov models (III)

With the following model:

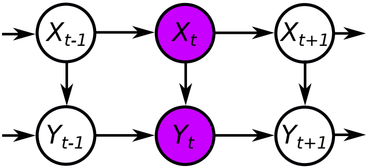


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? No
- $\{Y_t\}$? [1 minute]

D-separation in Markov models (III)

With the following model:

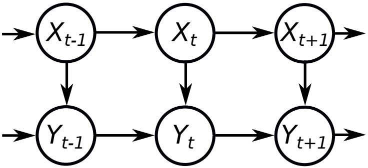


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? No
- $\{Y_t\}$? No
- $\{X_t, Y_t\}$? [1 minute]

D-separation in Markov models (III)

With the following model:

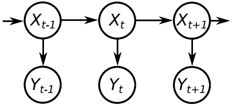
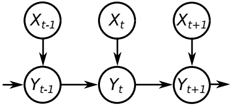
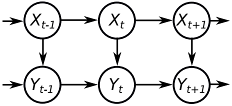


Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by ...

- $\{X_t\}$? No
- $\{Y_t\}$? No
- $\{X_t, Y_t\}$? Yes

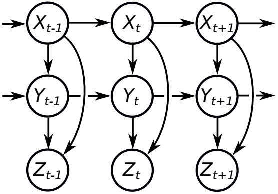
D-separation in Markov models: summary

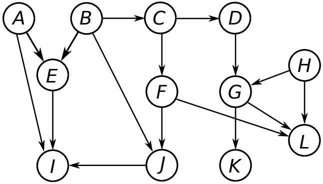
Is $\{X_{t-1}\}$ D-separated from $\{Y_{t+1}\}$ by (left column) in (top row)?

			
$\{X_t\}$	Yes	No	No
$\{Y_t\}$	No	Yes	No
$\{X_t, Y_t\}$	Yes	Yes	Yes

More complex models

Let's play a little bit with these two models:

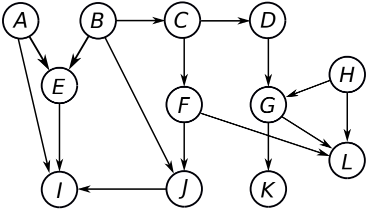




Find one example of D-separation and one example of non D-separation (in whatever case). [You've got 5 minutes]

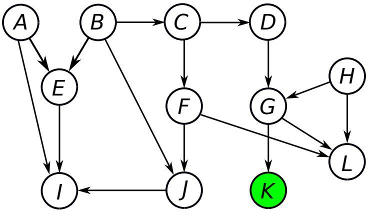
Markov blanket (or boundary)

Model example:



Markov blanket (or boundary)

Model example:



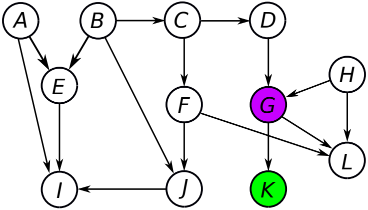
For a given node K, what is the minimal set of variables $\mathcal{B}_K$ so that:

$$p(K|\text{all except } K) = p(K|\mathcal{B}_K)?$$

You've got 3 minutes!

Markov blanket (or boundary)

Model example:

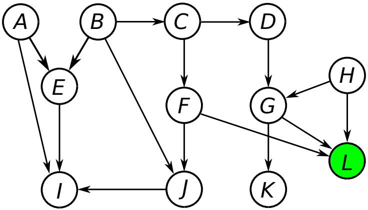


For a given node K, what is the minimal set of variables $\mathcal{B}_K$ so that:

$$p(K|\text{all except } K) = p(K|\mathcal{B}_K)? \qquad \mathcal{B}_K = \{G\}$$

Markov blanket (or boundary)

Model example:



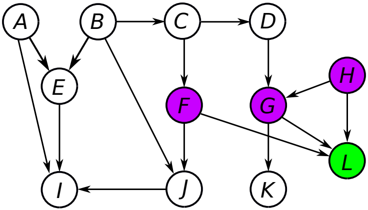
For a given node K, what is the minimal set of variables $\mathcal{B}_K$ so that:

$$p(K|\text{all except } K) = p(K|\mathcal{B}_K)? \qquad \mathcal{B}_K = \{G\}$$

For L? You've got 3 minutes!

Markov blanket (or boundary)

Model example:



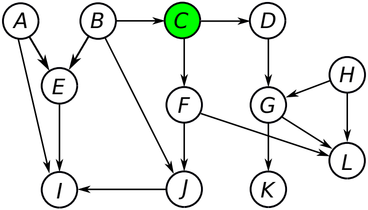
For a given node K, what is the minimal set of variables $\mathcal{B}_K$ so that:

$$p(K|\text{all except } K) = p(K|\mathcal{B}_K)? \qquad \mathcal{B}_K = \{G\}$$

For L? $\quad \mathcal{B}_L = \{F, G, H\}$ because F, G, H are **parents** of L.

Markov blanket (or boundary)

Model example:



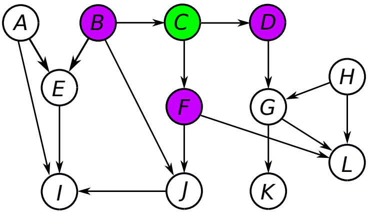
For a given node K, what is the minimal set of variables $\mathcal{B}_K$ so that:

$$p(K|\text{all except } K) = p(K|\mathcal{B}_K)? \qquad \mathcal{B}_K = \{G\}$$

For L? $\quad \mathcal{B}_L = \{F, G, H\}$ because F, G, H are **parents** of L.
For C? $\quad$ You've got 3 minutes!

Markov blanket (or boundary)

Model example:



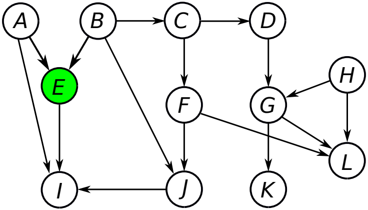
For a given node K, what is the minimal set of variables $\mathcal{B}_K$ so that:

$$p(K|\text{all except } K) = p(K|\mathcal{B}_K)? \qquad \mathcal{B}_K = \{G\}$$

For L? $\mathcal{B}_L = \{F, G, H\}$ because F, G, H are **parents** of L.
For C? $\mathcal{B}_C = \{B, D, F\}$ because B (F, D) is parent (**children**) of C.

Markov blanket (or boundary)

Model example:



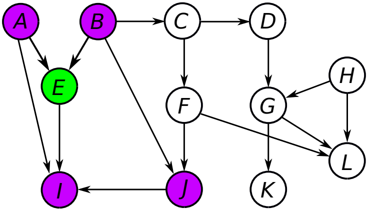
For a given node K, what is the minimal set of variables $\mathcal{B}_K$ so that:

$$p(K|\text{all except } K) = p(K|\mathcal{B}_K)? \qquad \mathcal{B}_K = \{G\}$$

For L? $\mathcal{B}_L = \{F, G, H\}$ because F, G, H are **parents** of L.
For C? $\mathcal{B}_C = \{B, D, F\}$ because B (F, D) is parent (**children**) of C.
For E? You've got 3 minutes!

Markov blanket (or boundary)

Model example:



For a given node K, what is the minimal set of variables $\mathcal{B}_K$ so that:

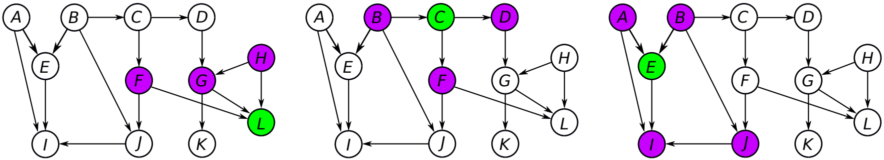
$$p(K|\text{all except } K) = p(K|\mathcal{B}_K)? \qquad \mathcal{B}_K = \{G\}$$

For L? $\mathcal{B}_L = \{F, G, H\}$ because F, G, H are **parents** of L.
For C? $\mathcal{B}_C = \{B, D, F\}$ because B (F, D) is parent (**children**) of C.
For E? $\mathcal{B}_E = \{A, B, I, J\}$ because A, B (I) are parents (children) of E and J is **co-parent** of E.

Markov blanket: definition

Definition of Markov blanket

The Markov blanket is the minimal set that D-separates a set of nodes from the rest of the graph.

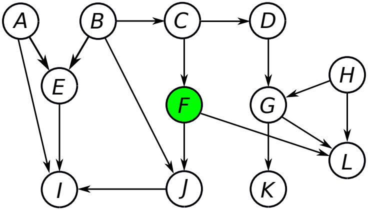


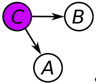
Construction of the Markov blanket

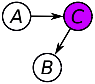
Given a directed acyclic graph, and a node X on that graph, the Markov blanket of X, $\mathcal{B}_x$ is the set of all parents, children and co-parents of X.

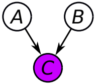
Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.



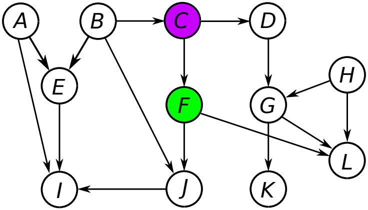
Blocked paths: 

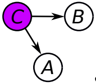
and 

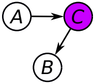
Unblocked paths: 

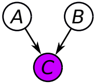
Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.



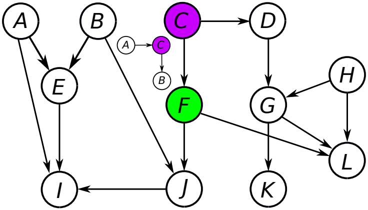
Blocked paths: 

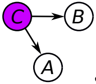
and 

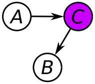
Unblocked paths: 

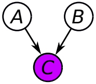
Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.



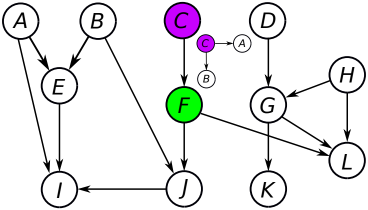
Blocked paths: 

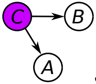
and 

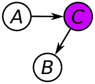
Unblocked paths: 

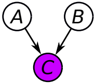
Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.



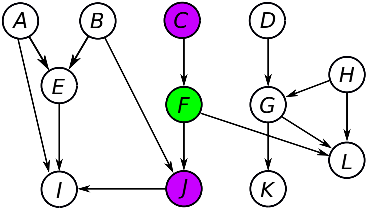
Blocked paths: 

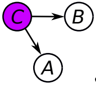
and 

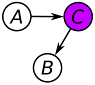
Unblocked paths: 

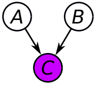
Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.



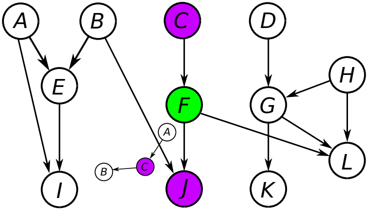
Blocked paths: 

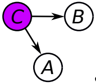
and 

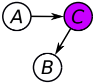
Unblocked paths: 

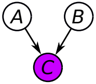
Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.



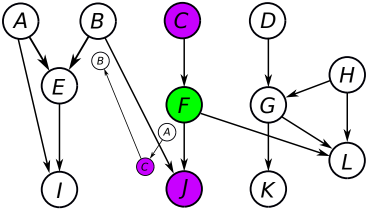
Blocked paths: 

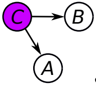
and 

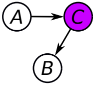
Unblocked paths: 

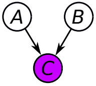
Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.



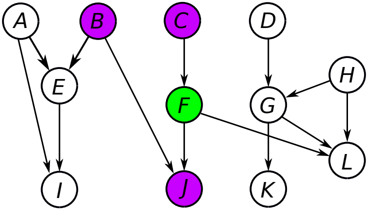
Blocked paths: 

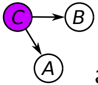
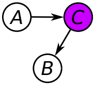
and 

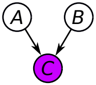
Unblocked paths: 

Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.

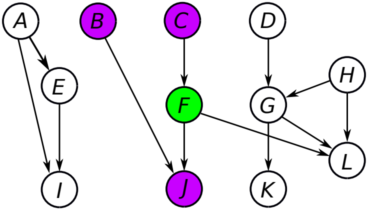


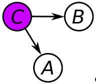
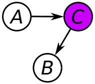
Blocked paths:  and 

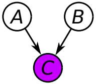
Unblocked paths: 

Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.

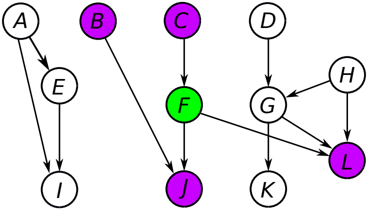


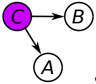
Blocked paths:  and 

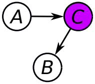
Unblocked paths: 

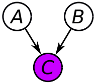
Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.



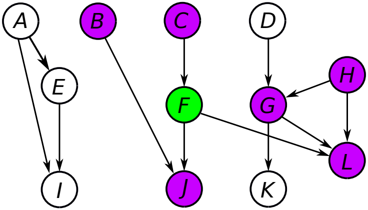
Blocked paths: 

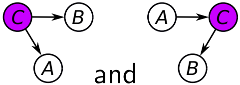
and 

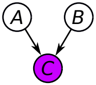
Unblocked paths: 

Markov blanket: extra-example

Aim: Isolate F from as many nodes as possible with blocked paths.

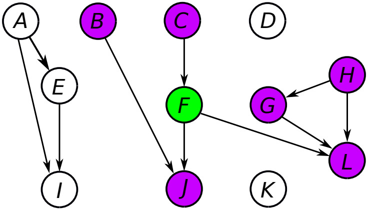


Blocked paths: 

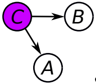
Unblocked paths: 

Markov blanket: extra-example

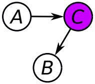
Aim: Isolate F from as many nodes as possible with blocked paths.



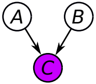
Blocked paths:



and

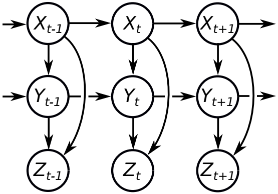


Unblocked paths:



Markov blanket: practice

Homerwork: Find the Markov blanket of X_t, of Y_t and of Z_t in:



Next two sessions Fei will explain the GMM in detail and the expectation-maximization (EM) algorithm. This provides the basis for the rest of the semester and for practical session #1.

See you in three weeks!!!