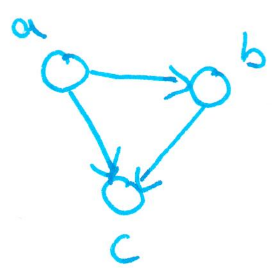
Lecture 22 - Graphical Models

This lecture is based on chapter 8 in "Pattern Recognition and Machine Learning" by Christopher M. Bishop.

Graphical models are helpful tools to visualize the structure of a prob. model and to gain insights about conditional independence.

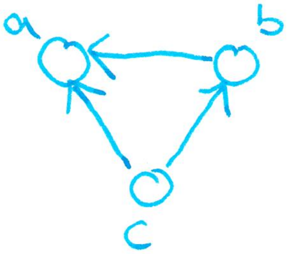
Bayesian Networks

- Directed acyclic graph (DAG)
- nodes: random variables
- edges : prob. relationships



joint distribution:
$p(a,b,c) = p(c|a,b) \cdot p(b|a) \cdot p(a)$

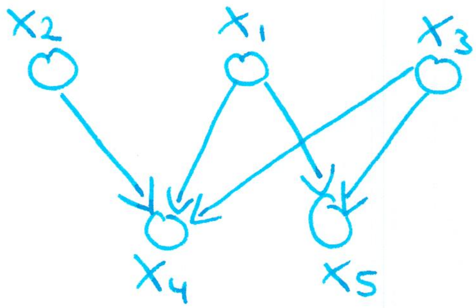
$\Rightarrow$ The graphical representation depends on the order we choose for the chain rule!



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

⇒ Both graphs are fully connected because there is a link between every pair of nodes.

⇒ It is the absence of links that conveys interesting information.



$$p(x_1,x_2,x_3,x_4,x_5) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4|x_1,x_2,x_3) \cdot p(x_5|x_1,x_3)$$

⇒ For a set of K variables we have

$$p(x) = \prod_{k=1}^{K} p(x_k|pa_k)$$

with pa_k being all parents of x_k

and $x = [x_1, x_2, \dots, x_K]$

Ancestral sampling:

⇒ We want to sample from $p(x_1, \dots, x_K)$

⇒ Assume ordering such that there is no link from a node to any lower numbered node

⇒ Each node has a higher number than any of its parents.

1. Draw $\hat{x}_1$ from $p(x_1)$
2. Go through all nodes in order and draw $\hat{x}_n \sim p(x_n | pa_n)$. with pa_n set to the corresponding previously drawn values.

How do we get $p(x_2, x_4)$?

⇒ marginalize $p(x_2, x_4) = \sum_{x_{\setminus\{2,4\}}} p(x_1, x_2, \dots, x_K)$

⇒ draw from joint pdf and ignore all $\{\hat{x}_{j \neq 2,4}\}$.

Conditional independence:

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

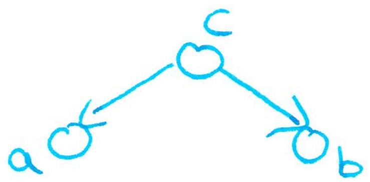
$\Rightarrow$ a is conditionally independent of b given c.

$\Rightarrow p(a,b|c) = p(a|c) \cdot p(b|c)$

$\Rightarrow$ We write $a \perp\!\!\!\perp b \mid c$

Some examples:

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



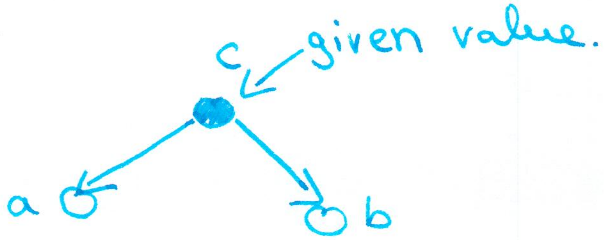
Are a and b independent?

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

$\Rightarrow$ Does not factorize to $p(a) \cdot p(b)$

$\Rightarrow$ not independent.

Now we condition on c:



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

$$= p(a|c)\cdot p(b|c) \quad \Rightarrow \quad a \perp\!\!\!\perp b \mid c$$

<u>Note:</u> The path from a to b via c is tail-to tail in c

$\Rightarrow$ a and b are dependent, unless c is given and blocks the path.

New graph: $a \longrightarrow c \longrightarrow b$

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

Are a and b independent?

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

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

$$= p(a)\cdot p(b|a) \neq p(a)\cdot p(b)$$

$\Rightarrow$ a and b are not independent.

Again we condition on c:

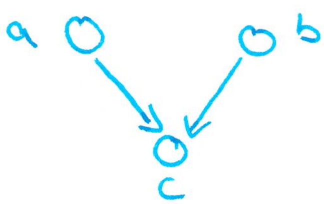
$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)\cdot p(c|a)\cdot p(b|c)}{p(c)}$$

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

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

This time c is head-to-tail with respect to the path from a to b.

Now the tricky one:



Note that c is head-to-head

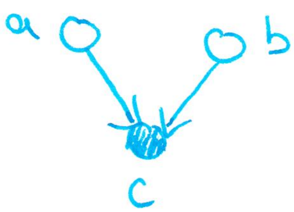
$$p(a,b,c) = p(a)\cdot p(b)\cdot p(c|a,b)$$

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

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

$$= p(a)\cdot p(b) \quad \Rightarrow \quad a \perp\!\!\!\perp b \mid \emptyset$$

Now for the conditioning on c:



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

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

$\Rightarrow$ Does not factorize into $p(a|c)\cdot p(b|c)$!

$\Rightarrow$ Observing c renders a and b dependent.
This is called explaining away.

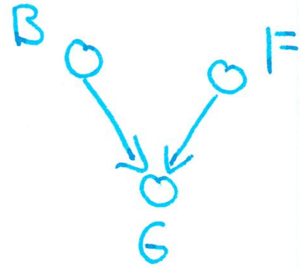
Example: Car fuel system

B: Battery state (full: 1, flat: 0)

F: Fuel tank state (full: 1, empty: 0)

G: Fuel gauge (indicating full: 1, empty: 0)

We have:



Let's make up some numbers:

$p(B=1) = 0.9 \qquad p(F=1) = 0.9$

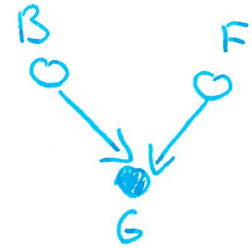
$p(G=1 \mid B=1, F=1) = 0.8$

$p(G=1 \mid B=1, F=0) = 0.2$

$p(G=1 \mid B=0, F=1) = 0.2$

$p(G=1 \mid B=0, F=0) = 0.1$

we observe $G=0$ $\Rightarrow$



If we know nothing about the Battery status, what is the prob of our fuel being empty?

$p(F=0) = 0.1$

$$p(F=0 \mid G=0) = \frac{p(G=0 \mid F=0) \cdot p(F=0)}{p(G=0)} \approx 0.257$$

$\Rightarrow p(F=0 \mid G=0) > p(F=0)$ which makes sense, because the fuel gauge gave us some information.

What if we discover that the battery is flat?

$$p(F=0 \mid G=0, B=0) = \frac{p(G=0 \mid B=0, F=0) \cdot p(F=0)}{\sum_F p(G=0 \mid B=0, F) \cdot p(F)}$$

$$\approx 0.111$$

Our confidence in the empty tank has decreased, because the empty battery is also an explanation for the fuel gauge status

$\Rightarrow$ The empty battery explains away the observation that the fuel gauge is empty.

Note that this would also work similarly if instead of observing G directly we would observe a descendant of G, e.g. our front passenger's report that the fuel gauge reads empty.

The more general term for these types of independence analysis is <u>d-separation</u>

Let A, B, C be arbitrary non-intersecting sets of nodes.

We want to know if $A \perp\!\!\!\perp B \mid C$

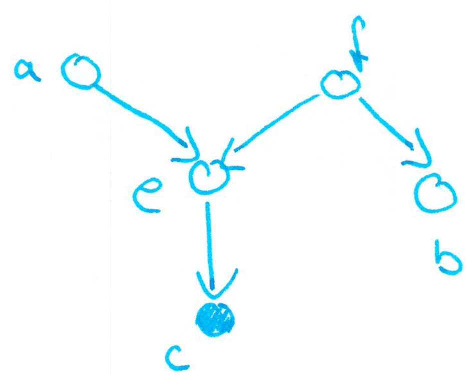
⇒ Consider all possible paths from any node in A to any node in B.

⇒ Any such path is blocked if it includes a node such that:

- The node is in C and the arrows meet head-to-tail or tail-to-tail.
- The node is not in C (and neither are any of it's descendants), and the arrows meet head-to-head.

⇒ If all paths are blocked then $A \perp\!\!\!\perp B \mid C$

Example:



Consider the path from a to b: $A = \{a\}$

$B = \{b\}$

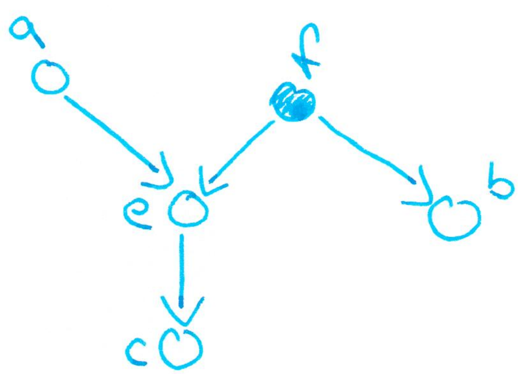
$C = \{c\}$

node f: $f \notin C$, but descendant $c \in C$, and
arrows are tail-to-tail
$\Rightarrow$ not blocked

node e: $e \notin C$ and head-to-head, but c is
a descendant and $c \in C$
$\Rightarrow$ not blocked

$\Rightarrow A \not\!\perp\!\!\!\perp B \mid C$

Second example:



A: $\{a\}$
B: $\{b\}$
C: $\{f\}$

node f: $f \in C$ and tail-to-tail
$\Rightarrow$ blocked

$\Rightarrow a \perp\!\!\!\perp b \mid f$

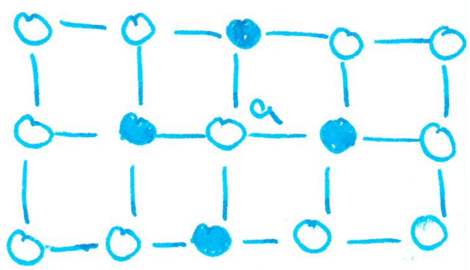
node e: (just for fun) $e \notin C$ and neither are any descendants
arrows are head-to-head
$\Rightarrow$ blocked

Undirected Graphs ⇒ Markov Random Field

d-separation simplifies to graph separation for undirected graphical models

⇒ Imagine a graph with all nodes in C removed. If there exists no path between any node in A to any node in B then $A \perp\!\!\!\perp B \mid C$.

You have already used this:



a is independent of all other nodes given C.

Factorization for undirected graphs:

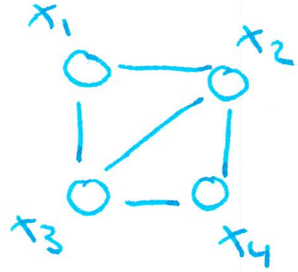
→ Two nodes x_i and x_j that are not connected by a link must be indipendent given all other nodes of the graph

$$p(x_i, x_j \mid X_{\setminus\{i,j\}}) = p(x_i \mid X_{\setminus\{i,j\}}) \cdot p(x_j \mid X_{\setminus\{i,j\}})$$

Clique: Fully connected subset of nodes

maximal clique: cannot make it any larger

Example:



a clique: $\{x_1, x_2\}$

a maximal clique: $\{x_1, x_2, x_3\}$

Let: C: a clique

x_C: variables/nodes of C

$\varphi(C)$: potential function

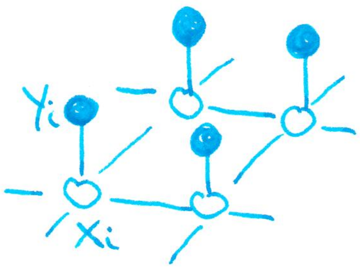
$\varphi(x_c) \geq 0$

$$p(x) = \frac{1}{Z} \prod_C \varphi_C(x_C)$$

$$Z = \sum_x \prod_C \varphi_C(x_C)$$

└→ pretty painful to compute.

Example: Image denoising



Y_i: observed noisy pixel value

X_i: underlying true pixel value

$x_i, y_i \in \{-1, 1\}$

$$E(x,y) = h\sum_i x_i - \beta\sum_{\{i,j\}} x_i x_j - \eta\sum_i x_i y_i$$

$$p(x,y) = \frac{1}{Z}\exp\{-E(x,y)\}$$

$\Rightarrow$ In the homework we just averaged over samples.

$\Rightarrow$ often you want $\underset{x}{\operatorname{argmax}}\, p(x|y)$

$\Rightarrow$ Iterated Conditional Modes (ICM)

- set all y_i
- start with $x = y$
- iteratively pick a x_i and flip its state
- compute old and new energy
- choose state with lower energy
- repeat

$\Rightarrow$ greedy, local optimum

$\Rightarrow$ For certain $\varphi(x_c)$ graph cuts gives global optimum.