Conditional Independence

An important concept in modeling multidimensional probability distributions is conditional independence

Consider 3 R.V.'s a, b, c
where the conditional distribution of a given b and c is such that it does not depend on the value of b, so that

$$p(a \mid b, c) = p(a \mid c)$$

then we can say $a \perp\!\!\!\perp b \mid c)$ is conditionally independent of b given c.
In such a case the joint factorizes as

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

which shows that a and b ~~and~~ statistically independent given c.

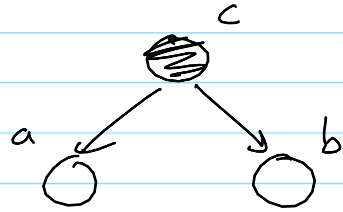
Conditional independence is important for model simplicity (as seen re: param counts) and for computation

Graphical models allow conditional independence properties to be read directly from the graph.

General framework called

d-separation

1 of 3 examples



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

if none observed note that a & b are not independent, i.e.

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

if c is <u>observed</u> then

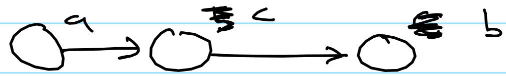
$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = p(a|c)\,p(b|c)$$

and we see that $a \perp\!\!\!\perp b \mid c$

<u>Graphical interpretation</u>

Node c is <u>tail-to-tail</u> and when c is observed it "blocks" the path from a to b rendering them conditionally independent given c

2 of 3 examples



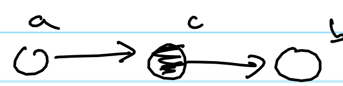
$$p(a,b,c) = p(a)\,p(c|a)\,p(b|c)$$

To test if a & b are a priori independent we can marginalize over c

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

so $a \not\perp\!\!\!\perp b \mid \emptyset$, i.e. a and b are not independent a priori

observing c



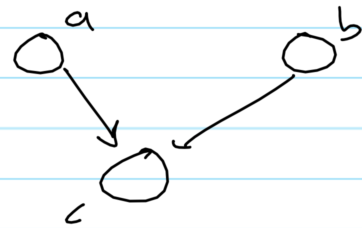
renders $a \perp\!\!\!\perp b \mid c$, to see this consider

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

$$= p(a|c)\,p(b|c) \quad \blacksquare$$

in this graph c is <u>head-to-tail</u> w.r.t the path from a to b. When we observe c along such a path, a and b are rendered conditionally independent.

3 of 3 examples

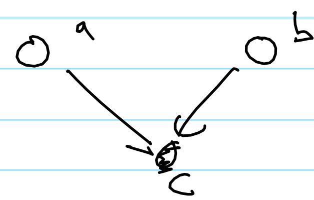


$$p(a,b,c) = p(a)\,p(b)\,p(c \mid a,b)$$

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

So, unlike examples 1 and 2 $a \perp\!\!\!\perp b \mid \emptyset$

Now observing c induces dependency between a and b



i.e. $a \not\!\perp\!\!\!\perp b \mid c$

Note $p(a,b \mid c) = p(a,b,c) / p(c)$

$$= p(a)\,P(b)\,p(c \mid a,b) \,/\, p(c)$$

$$\neq p(a \mid c)\,p(b \mid c)$$

Graphically c is head-to-head. When c is unobserved it blocks the path and vice versa.

Important note if $\exists$ a directed path from x to y then y is a descendant of x. A head-to-head will become unblocked if the node or any of its descendants is observed.

PRINTABLE NOTEPAD 1

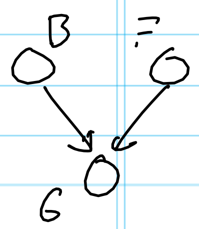
Explaining away

This last example is worth going into in more detail: (an example)

Vars

$B=\{0,1\}$ battery charged or not
$F=\{0,1\}$ fuel full or empty
$G=\{0,1\}$ gauge is either showing full or empty

A priori $p(B=1)=.9$, $p(F=1)=.9$

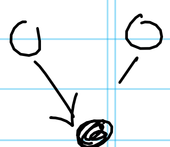


$P(G=1 \mid B,F)$

B \ F	0	1
0	.1	.2
1	.2	.8

$P(G=0 \mid B,F)$

B \ F	0	1
0	.9	.8
1	.8	.2



Before any observation $p(F=0)=.1$
Then, suppose we observe that $G=0$. What's the post'r prob that the fuel tank is empty?

$$p(F=0 \mid G=0) = \frac{P(G=0 \mid F=0)\,p(F=0)}{p(G=0)}$$

$$p(G=0 \mid F=0) = \sum_{B\in\{0,1\}} p(G=0 \mid B, F=0)\,p(B) = .81$$

$$p(G=0) = \sum_{B\in\{0,1\}} \sum_{F\in\{0,1\}} p(G=0 \mid B,F)\,p(B)\,p(F) = .315$$

yielding $p(F=0 \mid G=0) = \dfrac{.81(.1)}{.315} \doteq .257$

this is intuitive since it is greater than $P(F=0)$, i.e. conditioning on information does what you would expect.

Now, though, if we also observe $B=0$, that the battery is flat we see that

$$p(F=0 \mid G=0, B=0) = \frac{p(G=0 \mid B=0, F=0)\, p(F=0)}{\sum_{F \in \{0,1\}} p(G=0 \mid B=0, F)\, p(F)} \simeq .111$$

which is less than .257. Observing $B=0$ "explains away" the $G=0$ observation. That $P(F=0 \mid G=0, B=0)$ is affected by the value / observation of B is an illustration of the fact that $F \not\perp\!\!\!\perp B \mid G$

D-Separation

Consider a general directed graph in which A, B, and C are arbitrary nonintersecting sets of nodes. We want to know if $A \perp\!\!\!\perp B \mid C$.

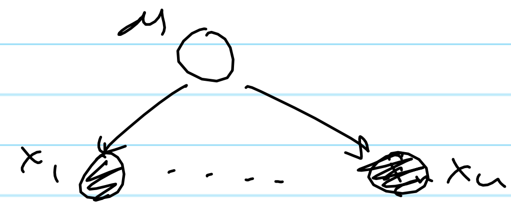
To do this we must consider all possible paths from <u>any</u> node in A to <u>any</u> node in B. Any such path is blocked if

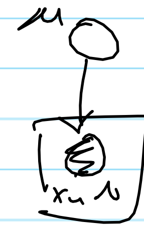
(a) the arrows on the path meet either head-to-tail or tail-to-tail at the node and the node is in set C or

(b) the arrows meet head-to-head at the node and neither the node nor any of its descendants is in the set C.

If <u>all</u> paths are blocked, then A is said to be d-separated from B by C and the joint distribution will satisfy $A \perp\!\!\!\perp B \mid C$.

<u>Concept</u> iid data & graphical models





$$x_i \perp\!\!\!\perp x_j \mid \mu \quad \forall i,j \text{ s.t. } i \neq j$$

all paths blocked tail-to-tail on μ.

Integrating over μ introduces dependency

μ is a <u>latent</u> variable