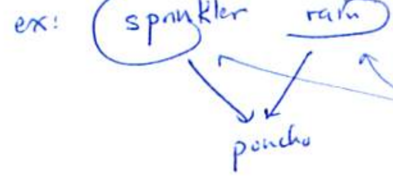
Did announcements, people say they will know their teams. Then had to post, so missed first part.

Kaggle req'd format to be consistent w/ Kaggle

- give clarification question:

ex: 

- marginal independence (does not imply conditional independence)

$X_i \perp X_j \mid \phi \not\Rightarrow$ that $X_i \perp X_j$ given you see a poncho!

↳ def: $P_\theta(X_i, X_j) = P_\theta(X_i) P_\theta(X_j)$

independent under a particular distribution, such as the marginal.

ex: $X_1 : \begin{cases} 1 & \text{heads} \\ 0 & \text{tails} \end{cases}$ $\quad X_2 : \begin{cases} 1 & \text{heads} \\ 0 & \text{tails} \end{cases}$

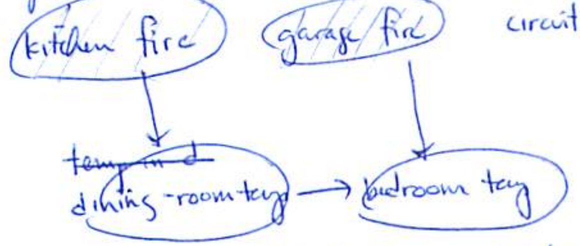
X_1 & X_2 are clearly indep.

But it isn't true that X_1 & X_2 are indep of $Y = X_1 + X_2$.

ex: if you know $Y = 2$ & $X_1 = 1$, then you know X_2.

So, basic idea: you know conditional (in)dependences and marginal (in)dependence.

Another Example: fires in the house.



circuitboard fire } marginally indep.

(b/c dining room is close to bedroom)

But, if DRT > BT, would guess P(kitchen fire on) > P(garage fire happening).

Exercise: given X_1, X_2, X_3, X_4, X_5 and practitioner tells you that:

$X_4 \perp X_5 \mid X_3$ $\quad X_4 \perp X_1 \mid X_3$ $\quad X_5 \perp X_1 \mid X_3$

$X_4 \perp X_2 \perp X_3$ $\quad X_5 \perp X_2 \mid X_3$

source of fire, say. ← [$X_1 \perp X_2$

circles = involves X_5

box = involves X_4

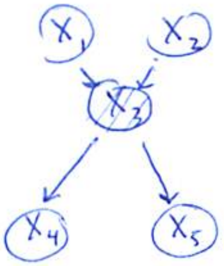
What kind of distributions would be allowed?

Well, $P(X_1, X_2, \dots, X_5) = P(X_5 \mid X_1, \dots, X_4) \cdot P(X_1, \dots, X_4)$

$= \left[P(X_5 \mid X_1, \dots, X_4) \cdot P(X_4 \mid X_1, X_2, X_3) \cdot P(X_3 \mid X_1, X_2) \cdot P(X_2 \mid X_1) \, P(X_1) \right] \cdot P(X_1, \dots, X_4)$

$= \left[\underbrace{P(X_5 \mid X_3)}_{\text{by circled stuff}} \underbrace{P(X_4 \mid X_3)}_{\text{by boxed stuff}} \cdot P(X_3 \mid X_1, X_2) \underbrace{P(X_2)}_{\text{by indep.}} \cdot P(X_1) \right] \cdot P(X_1, \dots, X_4)$

but this looks pretty cumbersome; let's capture this in a better representation. edges ~~Arrows~~ = {causality / generation}



] → unobserved but you want to infer

] → ~~unobserved~~ unobserved "dynamic".

] → observed.

~~variable~~ variables = nodes

Formally, a Bayes network is defined by a DAG G.

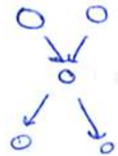
And, $P_\theta(X_1, \ldots, X_N) = \prod_{i=1}^{N} P_\theta(X_i \mid \text{Parents}(X_i))$ means P_θ factorizes over G.

So, they model distributions that can be factorized in this way.

q: can there be links from grandparents & grandchildren?

a: yes - undirected loops are OK.

semantics.

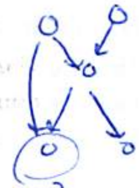


chromosomes of grandparent

chromosomes of grandchild

[even tho' grandparents "caused" the grandchildren, don't need to know the grandparents]

vs.



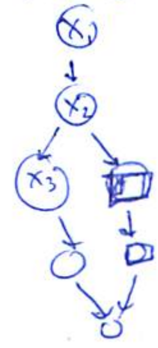
what ~~they~~ grandparents know

what grandchildren know

↳ so, you learned both from grandparent & parent, "directly".

(i.e., also need to know what grandparent to know what grandchild knows)

"Local markov property": each var is cond. indep of its non-descendants given its parents.



given parent X_2, X_3 is cond indep of the □'s.

note: given a DAG, we can ~~need~~ rename according to topological sort, so that if $i<j$, then X_i is not descended from X_j.

So then we can write $P_\theta(X_1, \ldots, X_N) = \prod_{i=1}^{N} P_\theta(X_i \mid X_1, \ldots, X_{i-1})$

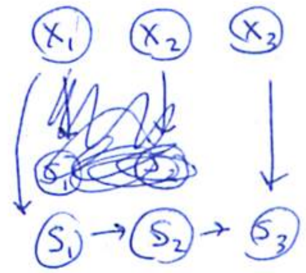
Some exagles of graphical models.

Ex: you have 3 variables, all indep coin flips.

Q: <no response>

X_1 X_2 X_3 (we're doing unobserved as colored).

Ex: you are going to model $S_i = \sum_{j=1}^{i} X_i$. Two diff models:



$P(S_1 = 1 \mid X_1 = x_1) = 1$

$P(S_2 = s \mid X_2 = x_2, S_1 = s_1) = \mathbb{1}_{s = s_1 + s_2}$

another valid model, is a diff parametrization.

$S_1 \to S_2 \to S_3$

when S_2 is 'generated' by a process that adds a random variable,

$$P(S_2 \mid S_1) = \begin{cases} S_1 + 0 & \text{w/ prob} = 1/2 \\ S_1 + 1 & \text{w/ prob} = 1/2 \end{cases}$$

$$P_\theta(s_3 \mid s_2) = \sum_{x_3 = \{0,1\}} P_\theta(s_3, x_3 \mid s_2)$$

you have this info.