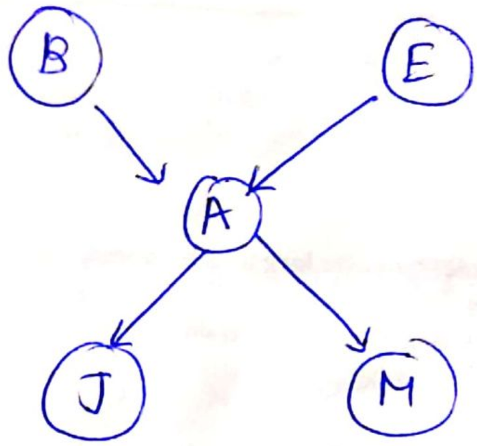
Review

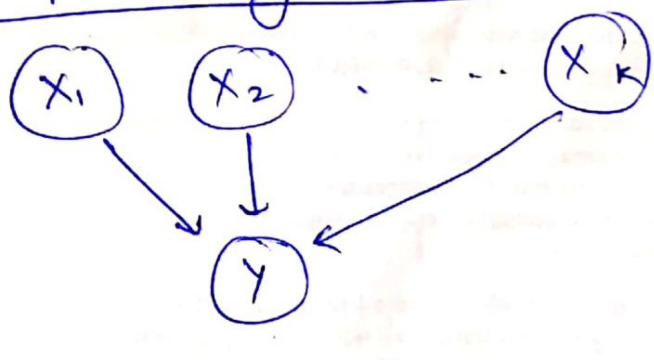


Burglary?
Earthquake?
Alarm?
John Calls?
Mary calls?

* Belief Network $BN = DAG + CPTs$
* Conditional Independence on all other ancestors given pare

$$P(x_i \mid x_1, \ldots x_{i-1}) = P(x_i \mid pa(x_i))$$

Representing CPTs



For simplicity, assume binary variables i.e. $\{0,1\}$ are the values

$x_i \in \{0,1\}$, $Y \in \{0,1\}$

How to represent $P(Y=1 \mid x_1, x_2, \ldots$

① Lookup table $O(2^k)$ rows store arbitrary CPT

x_1	x_2	. . .	x_k	$P(Y=1 \mid x_1, \ldots x_k)$
0	0			0.1
1	0			0.9
⋮	⋮			⋮
⋮	⋮			0.6

2^k combos

too many combinations if k is large.

(2) Deterministic CPT

"AND" function as a CPT $P(Y=1 \mid X_1, \ldots X_k) = \prod_{i=1}^{k} X_i$

"OR" function as a CPT $P(Y=0 \mid X_1, \ldots X_k) = \prod_{i=1}^{k} (1 - X_i)$

(3) Noisy-OR CPT

Use k numbers $p_i \in [0,1]$ to parametrize all 2^k entries in CPT.

$$P(Y=0 \mid X_1, X_2, \ldots X_k) = \prod_{i=1}^{k} (1-p_i)^{X_i}$$

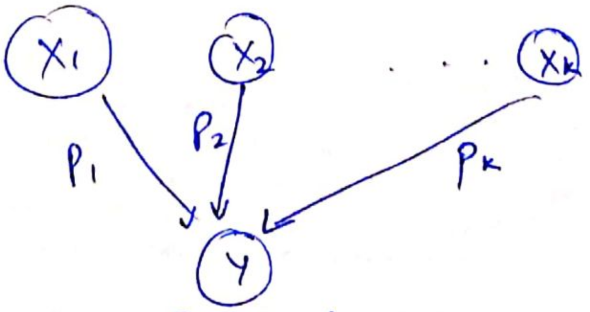
$$P(Y=1 \mid X_1, X_2, \ldots X_k) = 1 - \prod_{i=1}^{k} (1-p_i)^{X_i}$$

Why "Noisy-OR"? Look at prob. that $Y=1$ when exactly one parent is "on" (equal to 1):

$$P(Y=1 \mid X_1=0, X_2=0, \ldots X_{j-1}=0, X_j=1, X_{j+1}=0, \ldots X_k=0)$$

$$= 1 - (1-p_1)^0 \cdots (1-p_j)^1 \cdots (1-p_k)^0$$

$$= 1 - (1-p_j) = p_j$$

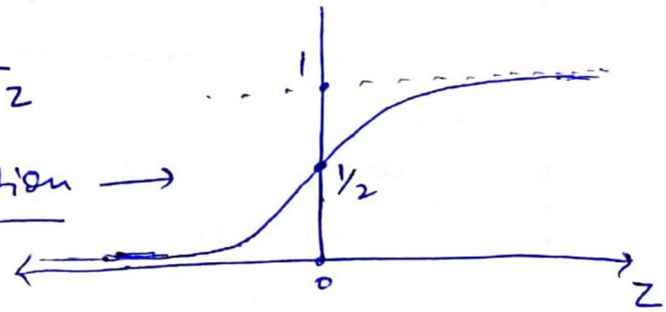


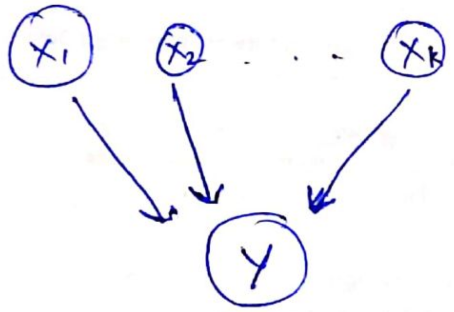
Intuitively, $p_i \in [0,1]$ is trigger prob. that $\underbrace{X_i = 1}_{\text{by itself}}$ causes $Y=1$. Setting all $p_i = 1$ for $i=1,2,\ldots k$ recovers logical OR.

Example : parents are diseases and child is symptoms.
The more diseases you have, the more probability of symptoms.

④ Sigmoid CPT
Use k real numbers θ_i to parametrize 2^k rows of CPT. Let $\sigma(z) = \dfrac{1}{1+e^{-z}}$

Sigmoid function →





$$P(Y=1 \mid X_1, X_2, \dots X_k) = \sigma\left(\sum_{i=1}^{k} \theta_i X_i\right)$$

Also known as logistic regression in Statistics or activation function in Neural Nets.

- If θ_i strongly negative, then $X_i = 1$ tends to inhibit / suppress $Y=1$
- If θ_i strongly positive, then $X_i=1$ tends to activate / excite $Y=1$
- Unlike a Noisy-OR CPT, sigmoid CPT can mix inhibition / excitation.

Conditional Independence

- A node X_i is conditionally independent of its non-parent ancestors given its parent.

$$P(X_i \mid X_1, X_2, \dots X_{i-1}) = P(X_i \mid pa(X_i))$$

- More generally:
  Let X, Y, and E refer to disjoint sets of nodes in BN
  When is X conditionally independent of Y given E?
  When is $P(X \mid E, Y) = P(X \mid E)$?
  When is $P(X, Y \mid E) = P(X \mid E)\, P(Y \mid E)$?

d-separation

(direction - dependent (in DAGs)
Relates conditional independence to graph theoretic properties.

Thm: $P(X, Y \mid E) = P(X \mid E)\, P(Y \mid E)$ if and only if every (undirected) path from a node in X to a node in Y is "d-separated" by E.

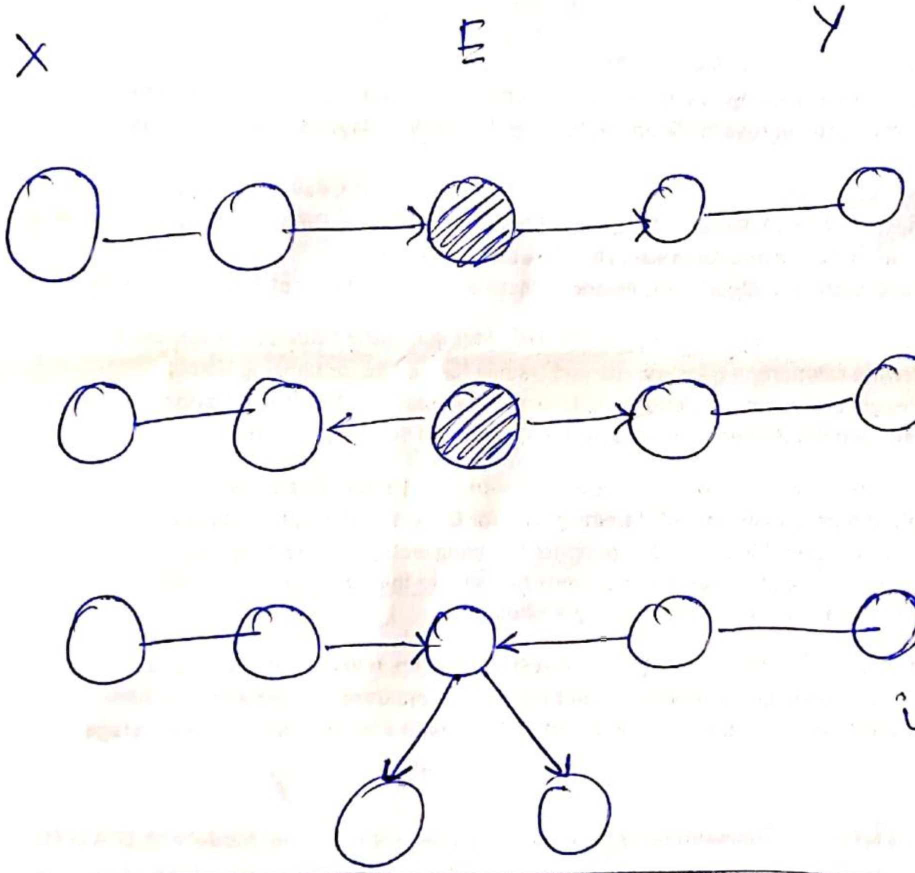
Def: a path π is d-separated if there exists a node $Z \in \pi$ for which one of three conditions hold:

1) $Z \in E$ with. $\longrightarrow Z \longrightarrow$ edges flow through the node Z

$\longleftarrow Z \longleftarrow$

2) $Z \in E$ with $\leftarrow Z \rightarrow$ edges divergee from Z

3) $Z \notin E$ ~~with~~ with $\rightarrow Z \leftarrow$ edges convergee on Z
descendants(z) $\notin E$ $\rightarrow (z) \leftarrow$

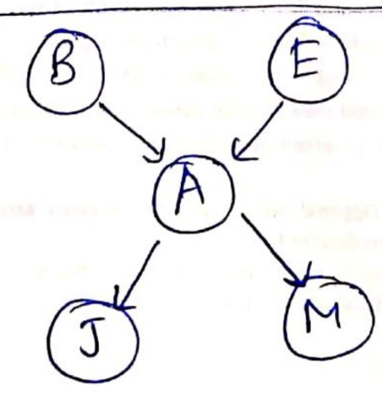


$Z \in E$ is an intervening event in a causal chain

$Z \in E$ is a common explanation or cause of X and Y.

$Z \notin E$, desc(z) $\notin E$
↓
is a common effect (unobserved).

Ex.



True or false?

1) $P(B|A,M) \stackrel{?}{=} P(B|A)$ True

path $B \rightarrow A \rightarrow M$

Checks out $\rightarrow A \rightarrow$ (1) ✓

2) $P(J,M|A) \stackrel{?}{=} P(J|A)\,P(M|A)$ **TRUE**.

$J \leftarrow A \rightarrow M$

satisfies (ii)

3) $P(B) \stackrel{?}{=} P(B|E)$ **TRUE**

$B \rightarrow A \leftarrow E$

Satisfies (iii)

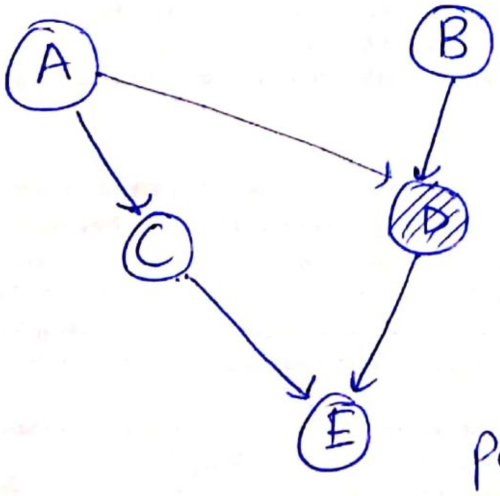
4) $P(B|M) \stackrel{?}{=} P(B|M,E)$ **FALSE**

$B \rightarrow A \leftarrow E$

A violates (i),(ii), (iii) → desc(A) is E

explaining away.

Ex.



true or false **FALSE**.

$P(B|D,E) \stackrel{?}{=} P(B|D)$

path $B \rightarrow D \rightarrow E$

Condition (i) d-separation

path $B \rightarrow D \leftarrow A \rightarrow C \rightarrow E$

violates cond(iii).

all 3 conds. violated.

* Proof that d-sep $\iff$ conditional independence.
  is not trivial.
* Algorithms exist for efficient tests of d-separation.

~~* P~~ $\boxed{\text{Inference}}$

* Problem.
  E – set of evidence nodes
  Q – set of query nodes.
  How to compute posterior distribution $P(Q|E)$?

* Types of inference
  - diagnostic reasoning from effects to causes.
    $P(B=1 | M=1)$
  - causal reasoning from causes to effects
    $P(M=1 | B=1)$
  - explaining away (about multiple causes)
    $P(B=1 | A=1, E=1)$
  - mixed reasoning $P(B=1, M=1 | J=1, A=1)$

When can inference be done efficiently?
i.e. (polynomial time in size of DAG and CPTs)