FUDAN SUMMER SCHOOL INTRODUCTION TO AI

Probabilistic Reasoning and Decision Making

Day 4 – Conditional Independence, d-separation

Conditional independence revisited

full joint

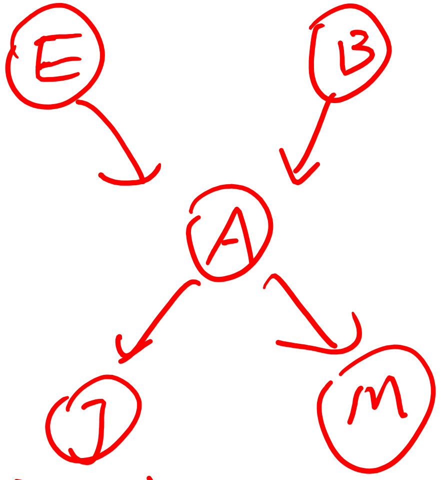
product

$$P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)\ldots P(X_n|X_1, X_2, \ldots, X_{n-1})$$

$$= \prod_{i=1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1})$$

always true

$P(J|A, E, B) = P(J|A)$



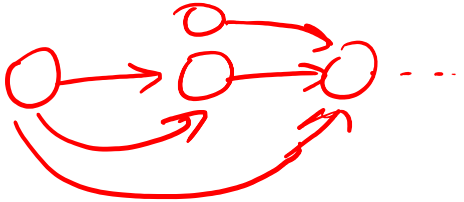
Suppose that in any given domain:

$P(E, B, A, J, M) = P(E)P(B)P(A|E, B)P(J|A)P(M|A)$

For Bayes Net

$$P(X_1 \ldots X_n) = \prod_{i=1}^{n} P(X_i | Pa(X_i))$$

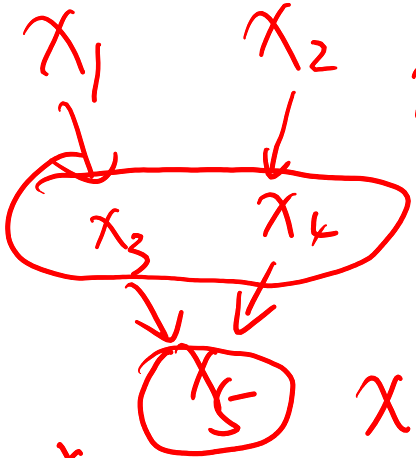
How to build a Bayes Net? : DAG



1. Choose random variables $\{X_1, X_2, \cdots X_n\}$
2. Choose an ordering — choose parents & their ordering
3. While there are variables left:
   1. *Add node X_i to the BN*
   2. *Set parents of X_i to the minimal subset such that X_i is conditionally independent from non-parent nodes preceding i, given its parents.*
   3. *Define $P(X_i|parent(X_i))$ CPT*

$Pa(X_i)$	$X_i = 1$
combinations	prob

all the nodes preceding X_i



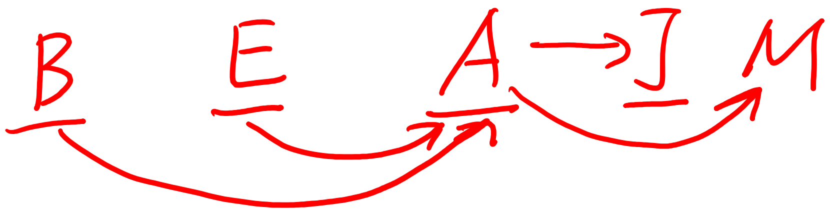
$Pa(X_i)$

X_i

$P(X_5 | X_1, X_2, X_3, X_4) = P(X_5 | X_3 X_4)$

How to build a Bayes Net?

1. Choose random variables
2. Choose an ordering
3. While there are variables left:
    1. *Add node X_i to the BN*
    2. *Set parents of X_i to the minimal subset such that X_i is conditionally independent from non-parent nodes preceding i, given its parents.*
    3. *Define $P(X_i|parent(X_i))$ CPT*



$P(J|A)$

$CPT \propto 2^2$

Which is the better ordering to choose?

A. Root events first (i.e. causes), e.g. B, E, A, J, M

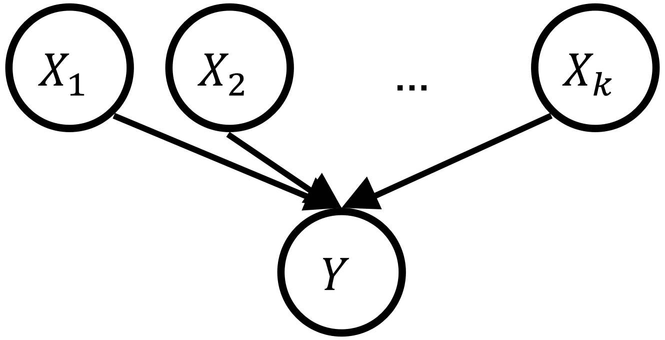
B. Child events first (i.e. consequences), e.g. M, J, A, E, B

C. It doesn't matter

$CPT \propto 2^{(\text{max \# of parents any node has})}$

$P(J|A,E,B) \propto 2^3$

Representing CPTs



How do we represent the CPT?

Option 1: CPT

X_1	X_2	---	X_k	$P(Y=1)$
0	0	---	0	$P(Y=1 \mid X_1=0, X_2=0, \dots X_k=0)$
0	0	---	1	
⋮	⋮		⋮	⋮
1	1	---	1	$P(Y=1 \mid X_1=1, X_2=1, \dots X_k=1)$

2^k entries

advantage: very easy to read

disadvantage: size exponential

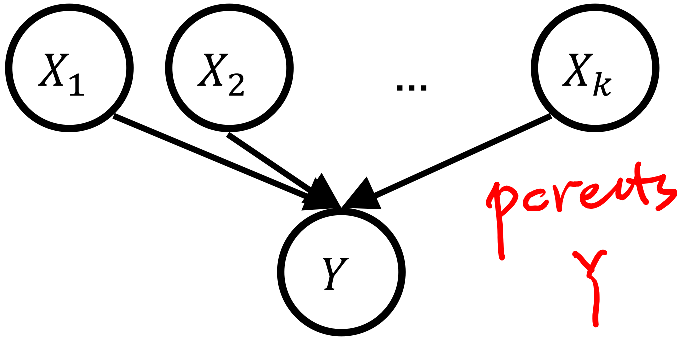
$k=20 \rightsquigarrow 2^{20} = 1M$

X_1	X_2	X_3	$P(Y=1)$

8 nums

decimal 5 $\Leftarrow$ 1 0 1 index

Representing CPTs



parents activate Y

advantage: save space (just do calculation)

disadvantage: deterministic

How do we represent the CPT?

option 2: OR logic

make $P(Y=1)$ deterministic of its parents

$$P(Y=1 \mid X_1 \dots X_k) = OR(X_1 \dots X_k)$$

$$OR(X_1 \dots X_k) = \begin{cases} 1 & \text{if any } x_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

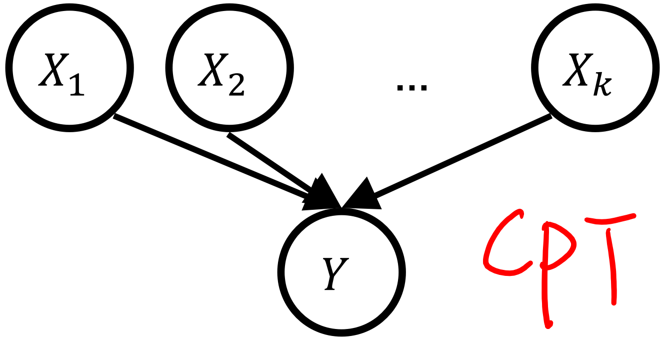
$$P(Y=0 \mid X \dots X_k) = 1 - OR(X_1 \dots X_k)$$

Representing CPTs

How do we represent the CPT?

option 3: middle ground: noisy-or

use k-parameters to model the relationship $p_i \in [0,1]$, $1 \le i \le k$



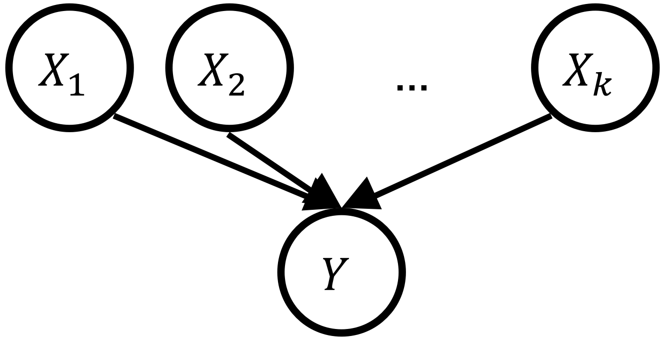
CPT: 2^k entries

noisy-or: k params (reduce the size a lot)

$$P(Y=0 \mid \underbrace{x_1 \cdots x_k}_{Pa(Y)}) = \prod_{i=1}^{k} (1-p_i)^{x_i}$$ assuming x_i is $\{0,1\}$

$$P(Y=1 \mid x_1 \cdots x_k) = 1 - \prod_{i=1}^{k} (1-p_i)^{x_i}$$

Noisy-OR



What is the value of $P(Y = 1|X_1 = 0, X_2 = 0, \ldots, X_k = 0)$?

A. 0
B. p_i
C. 1
D. None of these
E. Not enough information given

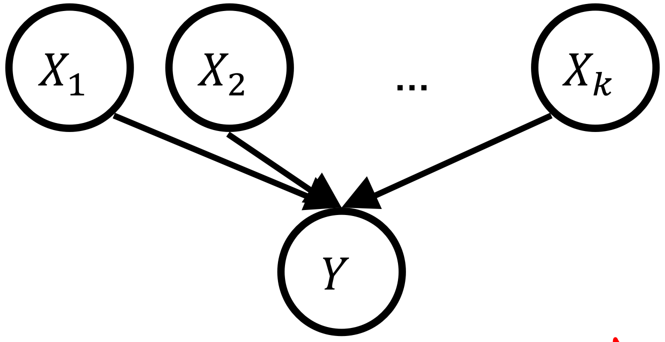
$$P(Y=1 \mid X_1 \cdots X_k) = 1 - \prod_{i=1}^{k} (1-p_i)^{X_i}$$

$$P(Y=1 \mid X_1=0,\ X_2=0, \ldots X_k=0)$$

$$= 1 - \underbrace{\prod_{i=1}^{k} (1-p_i)^0}_{1} = 1 - 1 = 0$$

if all parents are inactive, the child is guaranteed to be inactive

Noisy-OR



What is the value of $P(Y = 1|X_1 = 0, X_2 = 0, \ldots, X_i = 1, \ldots, X_k = 0)$?

A. 0

B. p_i

C. 1

D. None of these

E. Not enough information given

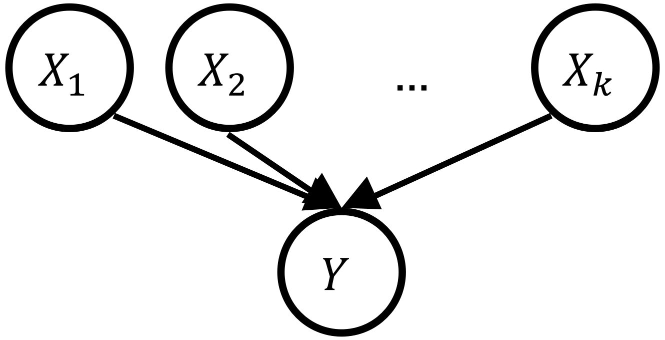
p_i: the probability that X_i alone will activate Y.

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

$$P(Y=1 | X_1=0 \cdots X_{i-1}=0 \quad X_i=1, \; X_{i+1}=0 \cdots X_k=0)$$

$$= 1-(1-p_i)^1 = p_i$$

Noisy-OR



Let $P(Y=1|X_1=0,\ldots,X_i=1,X_j=1,\ldots,X_k=0)=\frac{1}{3}$

and $p_i=\frac{1}{5}$

What is p_j?

A. $^1/_6$
B. $^2/_5$
C. $^1/_{15}$
D. $^2/_3$
E. I have no idea

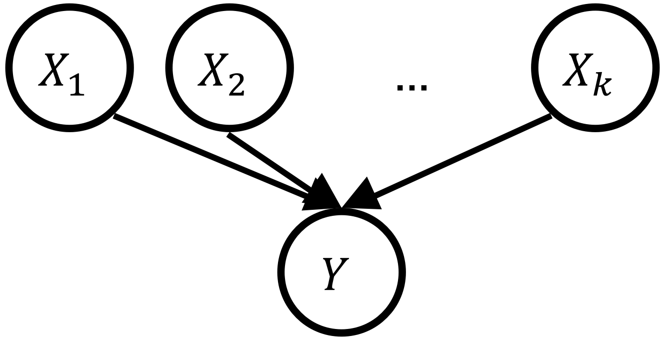
if only 3 parents noisy-or

CPT : $\boxed{p_1, p_2, p_3}$

CPT

X_1	X_2	X_3	$P(Y=1)$
0	0	0	0
0	0	1	p_3
0	1	0	p_2
0	1	1	$1-((1-p_2)(1-p_3))$
1	0	0	p_1
1	0	1	$1-(1-p_1)(1-p_3)$
1	1	0	$1-(1-p_1)(1-p_2)$
1	1	1	$1-(1-p_1)(1-p_2)(1-p_3)$

Noisy-OR



Let $P\left(Y=1|X_1=0,\ldots,X_i=1,X_j=1,\ldots,X_k=0\right)=\frac{1}{3}$

and $p_i=\frac{1}{5}$

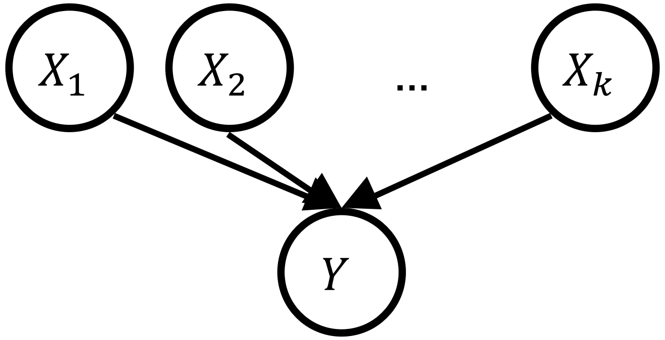
What is p_j?

A. $^1/_6$
B. $^2/_5$
C. $^1/_{15}$
D. $^2/_3$
E. I have no idea

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

$$1-\left(1-\frac{1}{5}\right)\left(1-p_j\right)=\frac{1}{3}$$

Noisy-OR



Is it true that P(Xi=1) = pi

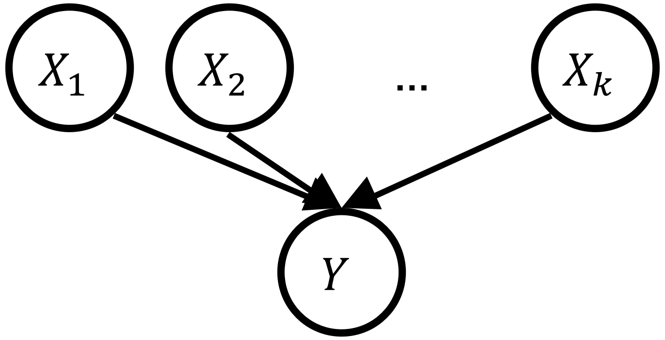
$p_i = P(x_i=1,$ other $x = 0)$

A. Yes

B. NO

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

Noisy-OR: Intuition



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

Let $P(Y=1|X_1=1, X_2=1, \dots, X_{i-1}=1, X_i=0, \dots, X_k=0) = a$ $= 1 - \prod_{j=1}^{i-1}(1-p_j)$

and $P(Y=1|X_1=1, X_2=1, \dots, X_{i-1}=1, X_i=1, X_{i+1}=0, \dots, X_k=0) = b$ $= 1 - \prod_{j=1}^{i}(1-p_j)$

What is the relationship between a and b?

A. $a \le b$

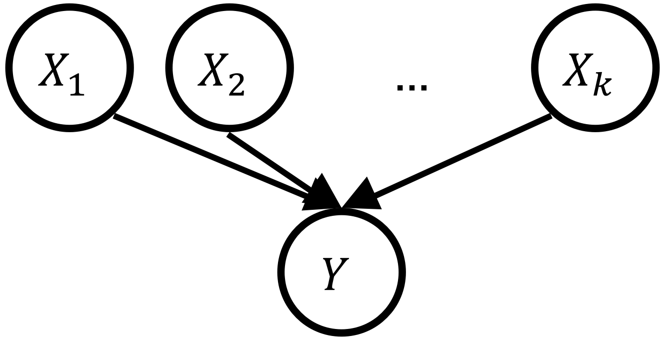
B. $a = b$

C. $a \ge b$

D. There is not enough information to determine the relationship

$$= 1 - \underbrace{(1-p_j)}_{\le 1}\prod_{j=1}^{i-1}(1-p_j)$$

Noisy-OR: Intuition



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

max of probability is when all $x_i = 1$

- When all parents are off, child is off with certainty (just like OR).
- Each parent that turns on increases (or at least does not decrease) the probability that the child will turn on.
- When all the parents are on, it still might not be certain that the child is on, but it is probably very likely.

Bayesian Networks: Network structure and conditional independence

- By our construction of a Bayes' net, we know that a node is conditionally independent of its non-parent ancestors given its parents:

$$P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)\ldots P(X_n|X_1, X_2, \ldots, X_{n-1})$$
$$= \prod_{i=1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1})$$
$$= \prod_{i=1}^{n} P\big(X_i|parents(X_i)\big)$$

But what about more general assertions of independence?

Bayesian Networks: Network structure and conditional independence

X, Y, E have no common nodes

$X: \{A, B\}$

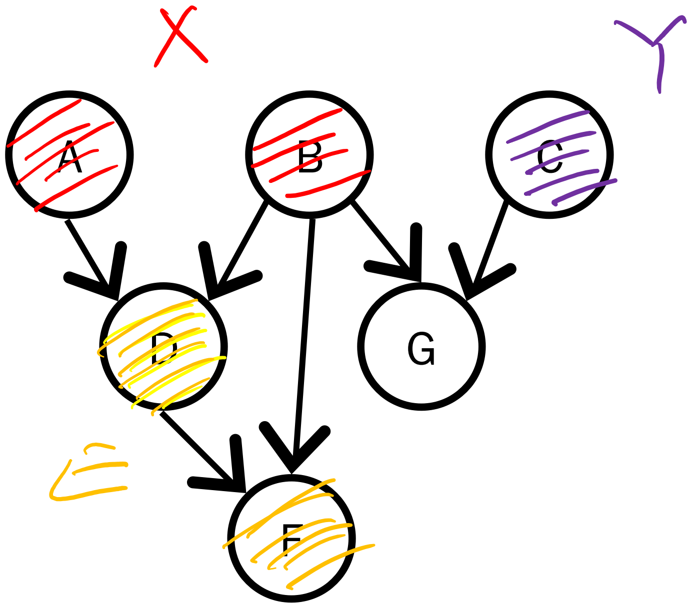
$Y: \{C\}$

$E: \{D, F\}$

- Let X, Y and E refer to disjoint sets of nodes. When is X conditionally independent of Y given E (evidence)?

Write three equalities using $P(X|Y,E)$, $P(Y|X,E)$, $P(X,Y|E)$, etc, that indicate this conditional independence.

$$\begin{cases} P(X|Y,E) = P(X|E) \\ P(Y|X,E) = P(Y|E) \\ P(X,Y|E) = P(X|E)P(Y|E) \end{cases}$$



Bayesian Networks: Network structure and conditional independence

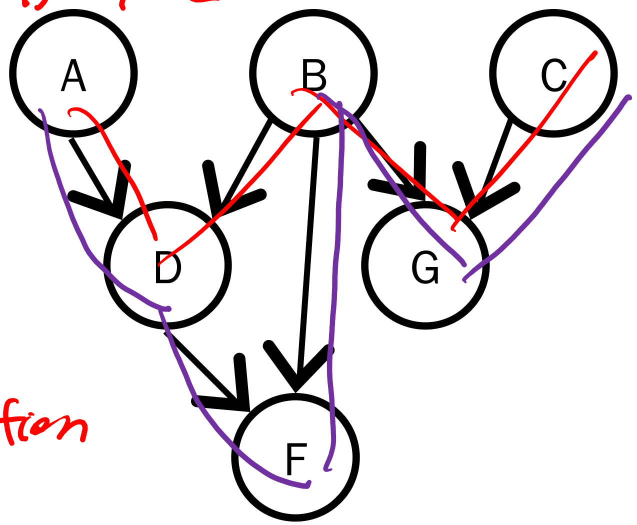
Every undirected path:
all possible undirected path
from any node in X to nodes in Y

- Let X, Y and E refer to disjoint sets of nodes. When is X conditionally independent of Y given E (evidence)?

X: {A, B}, Y: {C}

B-G-C

all possible path
A-D-B-G-C,
A-D-F-B-G-C,



Theorem: $P(X, Y|E) = P(X|E)P(Y|E)$ if and only if every UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

undirected path:
ignore edge direction

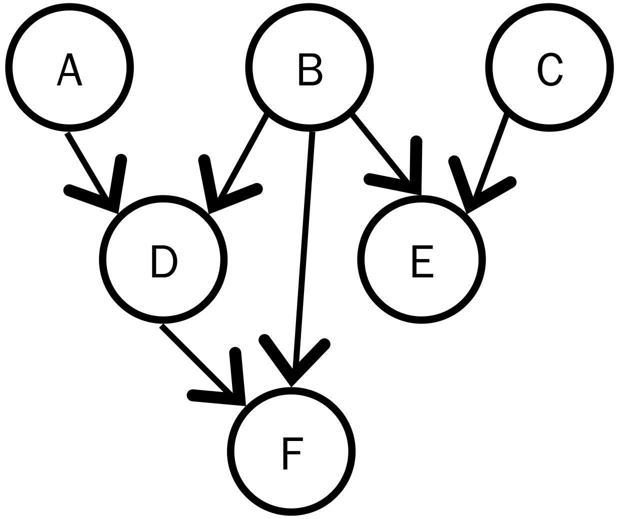
(we will not prove this)

undirected path from A to C

A-D-B-G-C

A-D-F-B-G-C

d-separation ("dependence" separation)



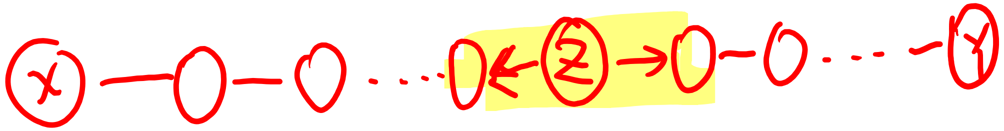
Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which one of the following three conditions hold:

arrows can be reversed

Condition 1:

Z blocks X & Y, $Z \in E$

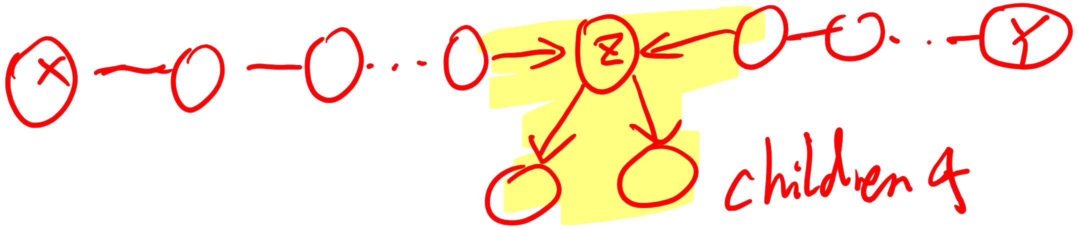
Condition 2:



$Z \in E$,

Z is the common cause

Condition 3:



children of Z

Z is the common consequence, $Z \notin E$, children of $Z \notin E$

Conditional Independence

$P(B|A,M)=P(B|A)$

A. True

B. False

X: {B} , Y: {M}, E: {A}

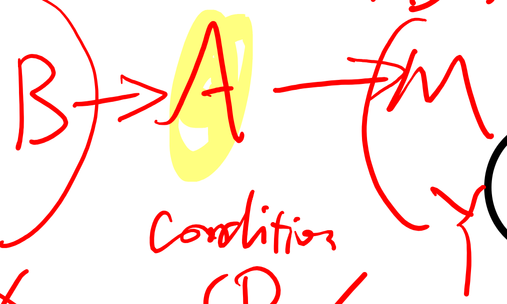
Theorem: $P(X,Y|E) = P(X|E)P(Y|E)$ if and only if **every** UNDIRECTED path from a node in X to a node in Y is "d-separated" ("blocked") by E.

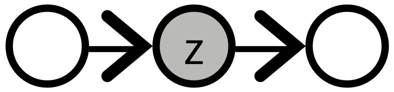
$B \to A \to M$

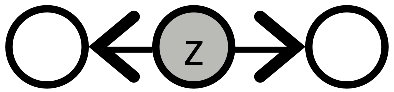
X Y

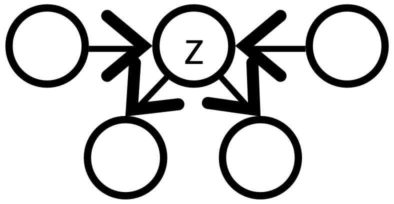
condition ① ✓

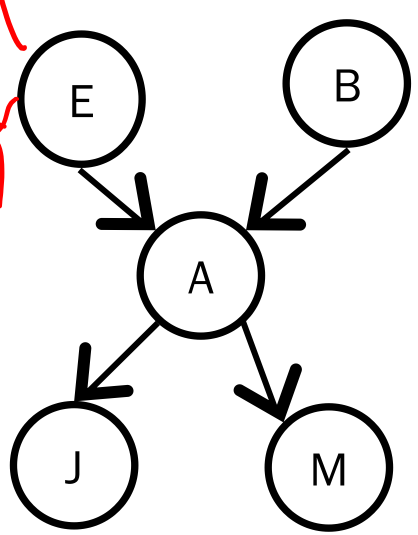
Definition: A path π (sequence of nodes connected by edges) is d-separated if there exists some node $z \in \pi$ for which **one** of the following three conditions hold:



1.  $z \in E$ (observed event in causal chain)

2.  $z \in E$ (observed common explanation)

3.  $z \notin E, descendants(z) \notin E$ (no observed common effect)



steps: ① list X, Y, E

② ignore all directions on edges, list all paths from X to Y

③ highlight evidence nodes, put back the direction of edges, check if all paths are d-sep