Modul 12: Probabilistic Reasoning System

Introduction to Probabilistic Reasoning System

KK IF – Teknik Informatika – STEI ITB

Inteligensi Buatan
(Artificial Intelligence)

Probabilistic Reasoning System (PRS)

Supervised Learning with Uncertainty (Non Deterministic)

Probability theory provides a quantitative way of encoding likelihood

Probability is a model of degree of belief

Given state(s) e, what is the probabilty that x happens → P(x|e)

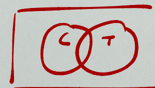
- Joint Probability Distribution
- Bayesian/ Belief Network

Joint Probability Distribution

- Random variables
  - Function: discrete domain → [0, 1]
  - Sums to 1 over the domain
    - Raining is a propositional random variable
    - Raining(true) = 0.2
      - P(Raining = true) = 0.2
    - Raining(false) = 0.8
      - P(Raining = false) = 0.8
- Joint distribution
  - Probability assignment to all combinations of values of random variables

Inference using Joint Probability Distribution

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576



$P(C \cup T) = P(C) + P(T) - P(C \cap T)$.

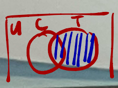
- For any proposition ϕ, sum the atomic events where it is true: $P(\phi) = \Sigma_{\omega:\omega \models \phi} P(\omega)$
- $P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$
- $P(cavity) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$
- P(cavity U toothache) = ?
- $P(\neg cavity \mid toothache)$ $= P(\neg cavity \cap toothache) / P(toothache)$

given

$= (0.016+0.064) / (0.108 + 0.012 + 0.016 + 0.064)$

$= 0.4$

↳ kalau udah gini, domainnya udah bukan U lagi, tapi T aja



domainnya yg biru aja.

Inference using Joint Probability Distribution

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- If you have n binary propositional variables □ requires 2^n numbers to build Joint Probability Distribution

□ Bayesian Network (We want to exploit independences in the domain)

Bayes' Rule:

$P(A \mid B) = P(A \cap B) / P(B)$

$\quad = P(B \mid A)\, P(A) / P(B)$

Conditioning:

$P(A) = P(A \mid B)\, P(B) + P(A \mid \neg B)\, P(\neg B)$

$\quad = P(A \cap B) + P(A \cap \neg B)$

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

$P(A \cap B) = P(A|B) \cdot P(B)$

$= P(B|A) \cdot P(A)$

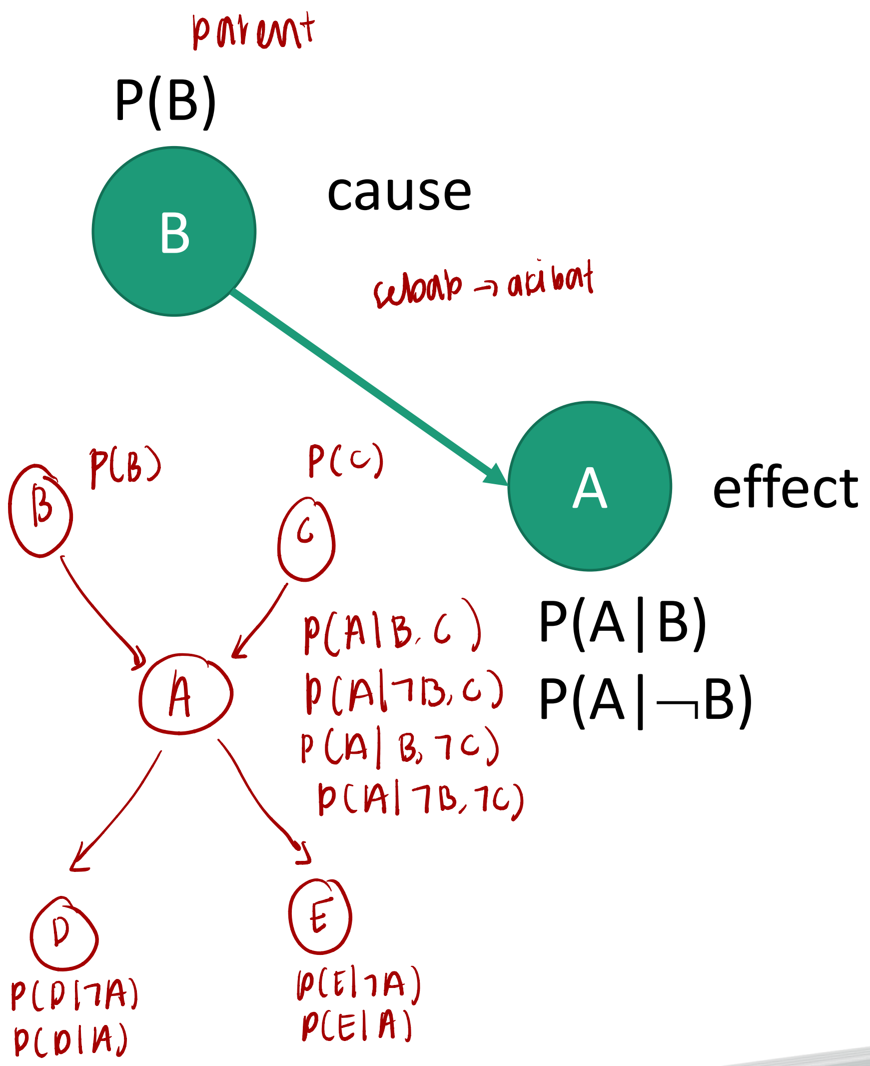
$P(\neg A|B) = 1 - P(A|B)$

$P(\neg A|\neg B) = 1 - P(A|\neg B)$

Structure of Bayesian Network

Nodes (variable)

Directed arc



Should be
Directed Acyclic Graph (DAG)

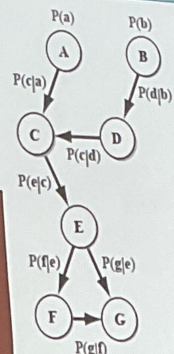
↓
karena nanti muter²
dan ga tau penyebabnya
(masa A → B, terus B → A ?)

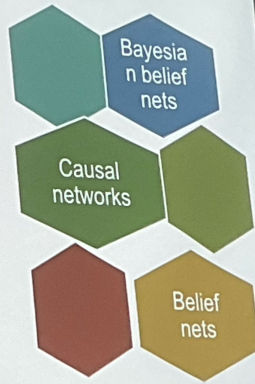
Structure of Bayesian Network

Representation of causal dependencies graphically (Hart et al.,2001)

BN have capability probabilistic reasoning like full joint probability distribution. It can answer any question about the domain.

-- How we exploit Independence?



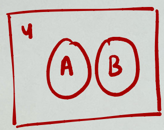


Independence

- A and B are independent iff
  - $P(A \cap B) = P(A) \cdot P(B)$
  - $P(A \mid B) = P(A)$
  - $P(B \mid A) = P(B)$

ga ada effect satu sama lain

- Independence is essential for efficient probabilistic reasoning
- A and B are conditionally independent given C iff
  - $P(A \mid B, C) = P(A \mid C)$
  - $P(B \mid A, C) = P(B \mid C)$
  - $P(A \cap B \mid C) = P(A \mid C) \cdot P(B \mid C)$



Example of Independence

- X is late (X)
- Traffic Jam (T)
- Y is late (Y)
- None of these propositions are independent of one other
- X and Y are conditionally independent given T

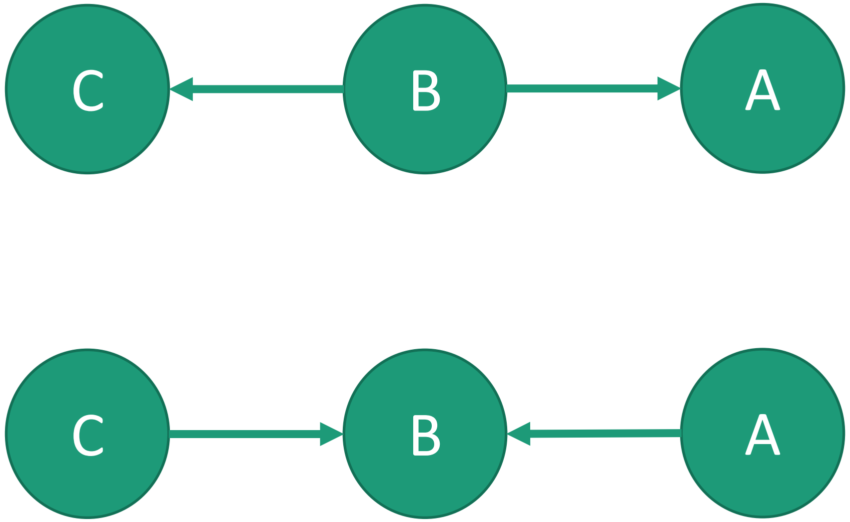
Types of Connections in Bayesian Network

Forward | Backward

Serial

Diverging

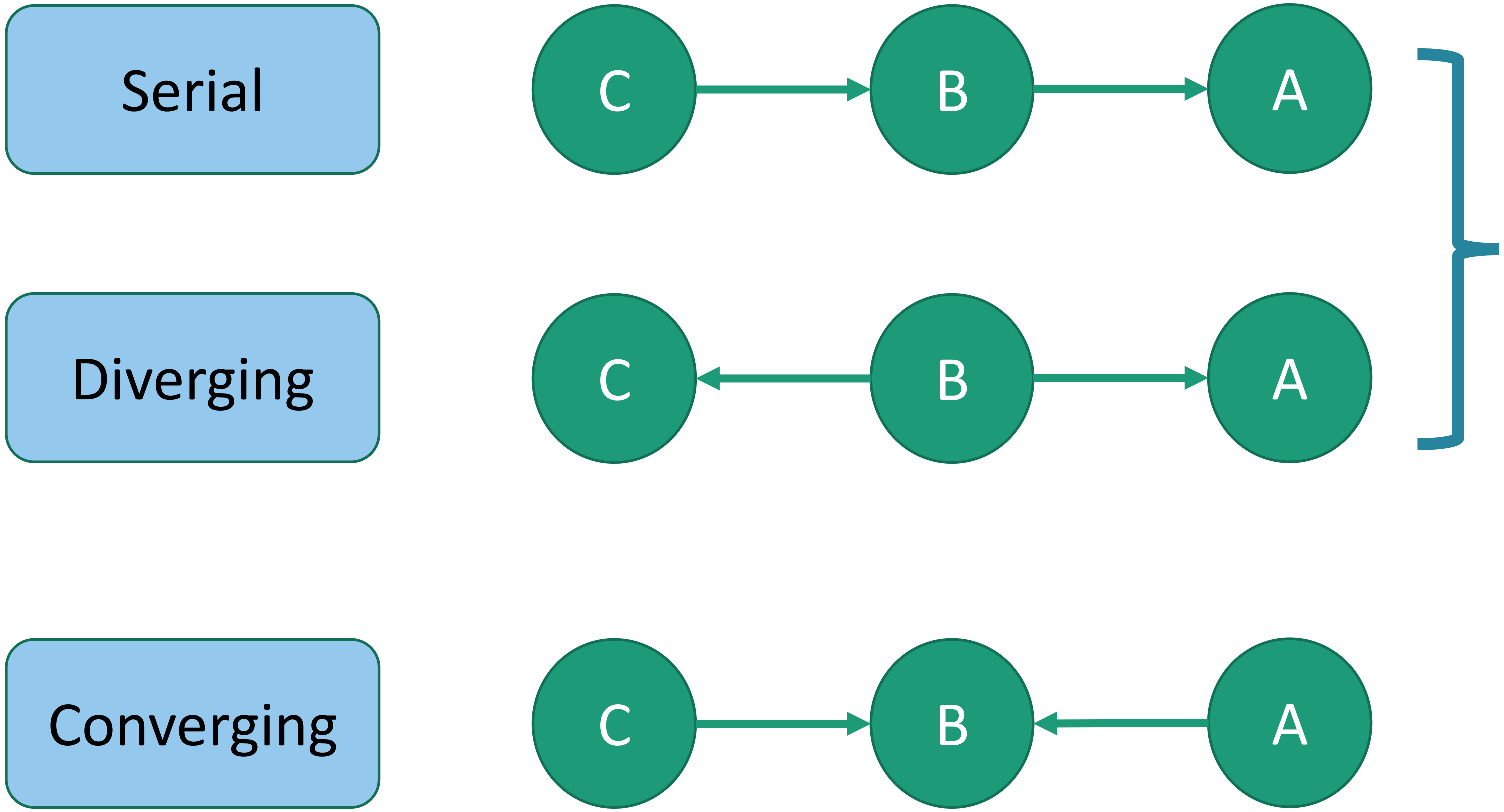
Converging



Why? To exploit Independence

↳ diawal tidak berhubungan, pas ada informasi dari B atau informasi dari variabel lain ke B, C dan A bisa aja berhubungan

Independence in Connection



Knowing C will tell us about A, but if we know B, knowing C will tell us nothing about A (C and A conditionally independent or d-separated)

Knowing C will tell us nothing about A without knowing B, but if we see evidence about B, C and A becomes dependent

Example of Bayesian Network Structure

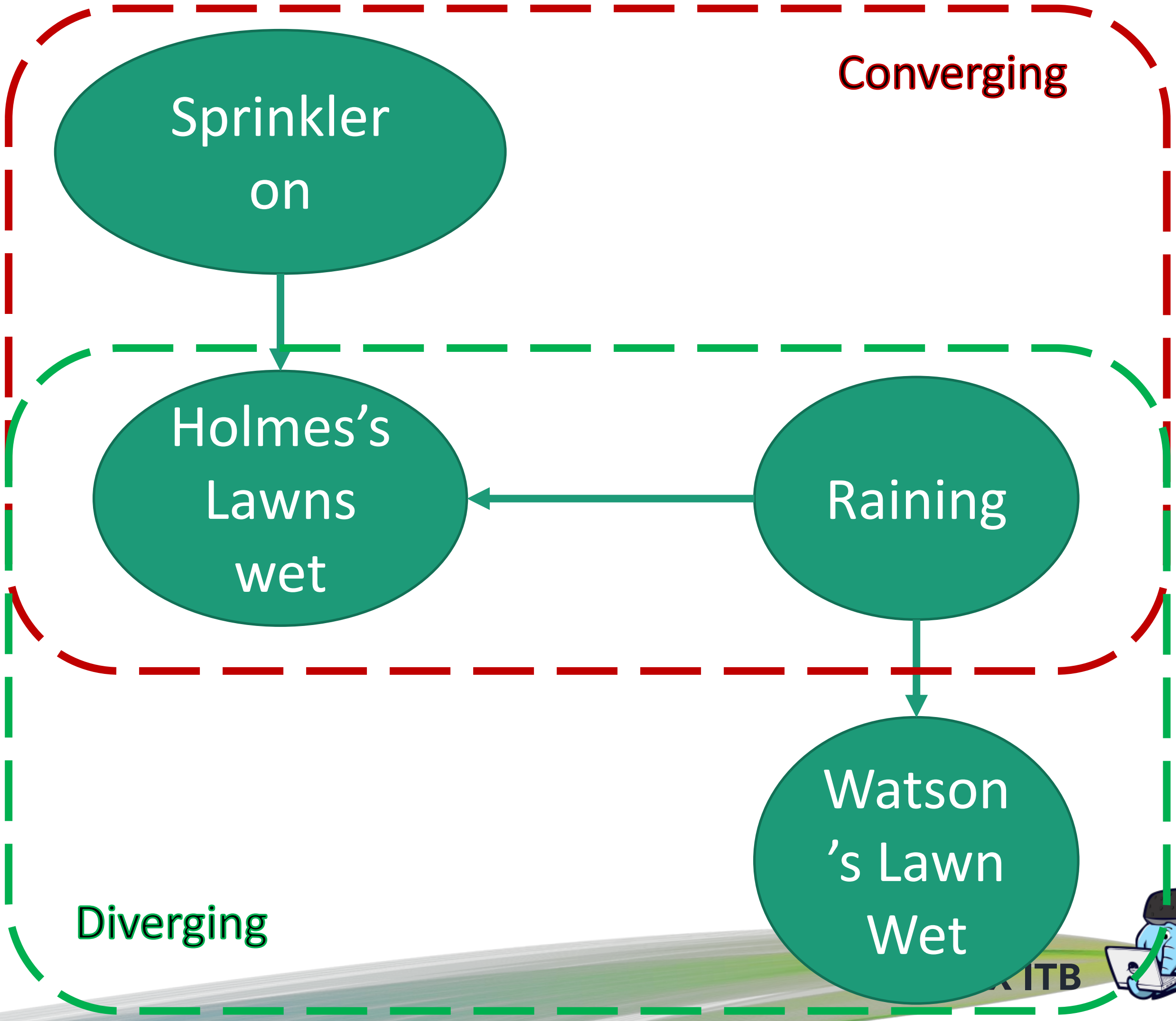
(from Kaelbling of MIT)

Holmes and Watson are neighbor

Holmes wakes up to find his lawn wet

Sprinkler on?

Raining?



Example of Bayesian Network with Likelihood

(from Kaelbling of MIT)

Holmes and Watson are neighbor

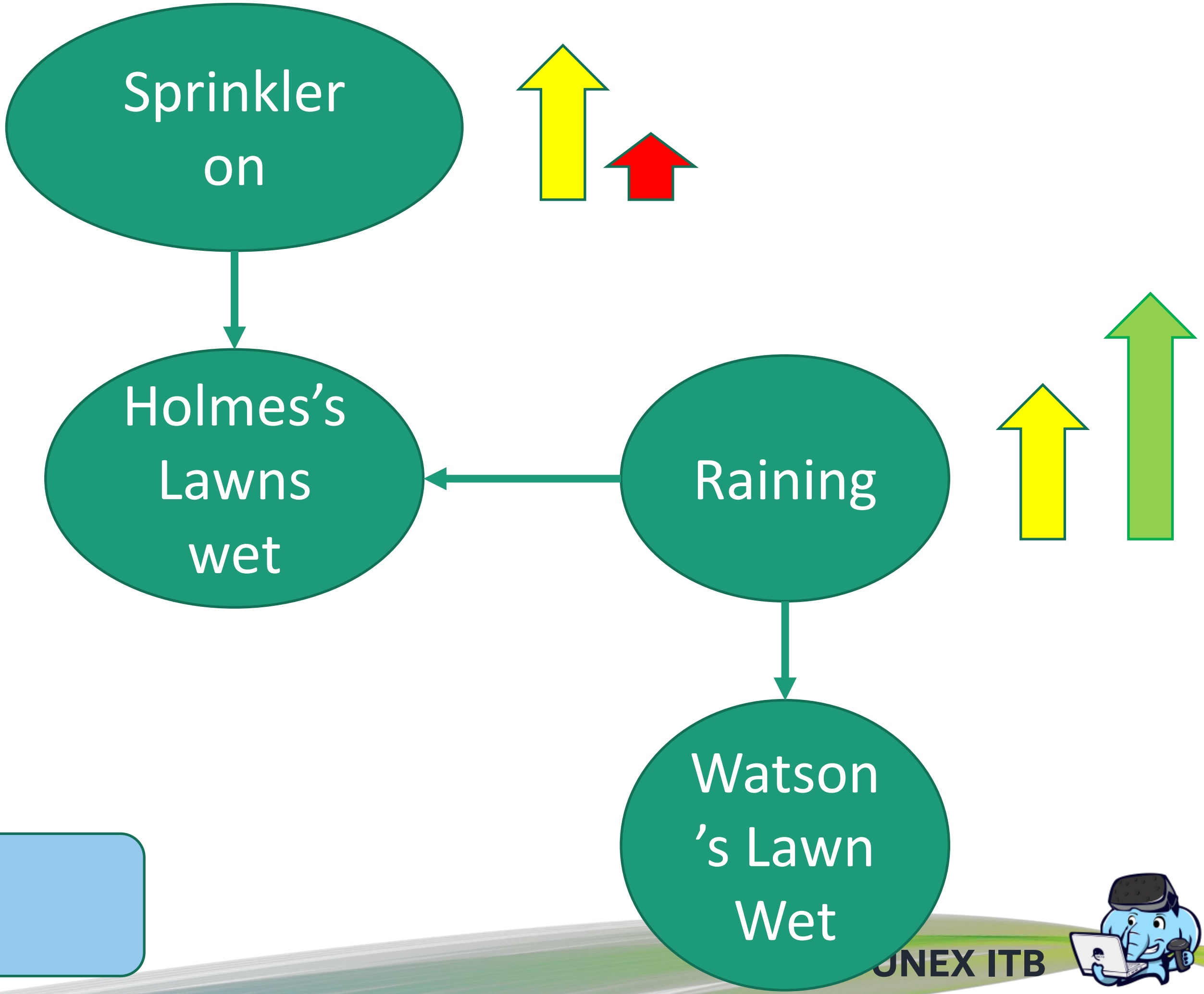
Holmes wakes up to find his lawn wet

Sprinkler on?

Raining?

Watson’s lawn and he sees it is wet too

Likelihood → Probability



Modul 12: Probabilistic Reasoning System

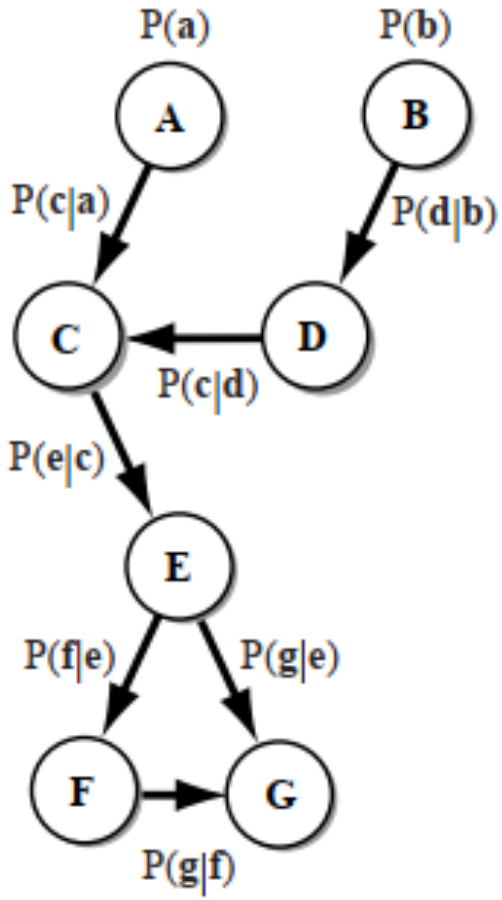
Bayesian Network: What & Why

KK IF – Teknik Informatika – STEI ITB

Inteligensi Buatan
(*Artificial Intelligence*)

Bayesian Networks: What

Representation of causal dependencies graphically (Hart et al.,2001)



Bayesian belief nets

Causal networks

Belief nets

Why Bayesian Networks ?

BN have capability probabilistic reasoning like full joint probability distribution. It can answer any question about the domain.

Full joint probability distribution can become intractably large as the number of variables grows.

BN: Independence and conditional independence relationships can greatly reduce the number of probabilities

lightness	width	category	Prob
1.5	14.6	salmon	0.3
…	…	…	…
8.3	15	Sea bass	0.25
…	…	…	…

In full joint probability distribution, each combination of variable values has information how probable it is.

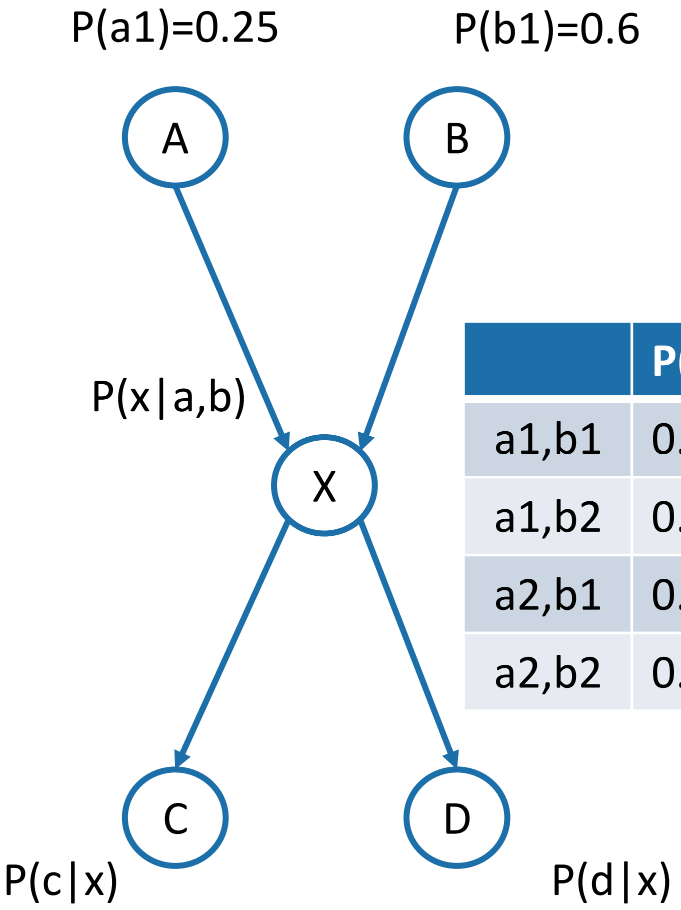
BN Components

Structure

- Node (variables)
- Directed arcs (acyclic graph)

Numerical parameters

- Prior probability
- Probability conditional tables



	P(x1|a,b)
a1,b1	0.3
a1,b2	0.7
a2,b1	0.6
a2,b2	0.8

	P(c1|x)
x1	0.3
x2	0.6

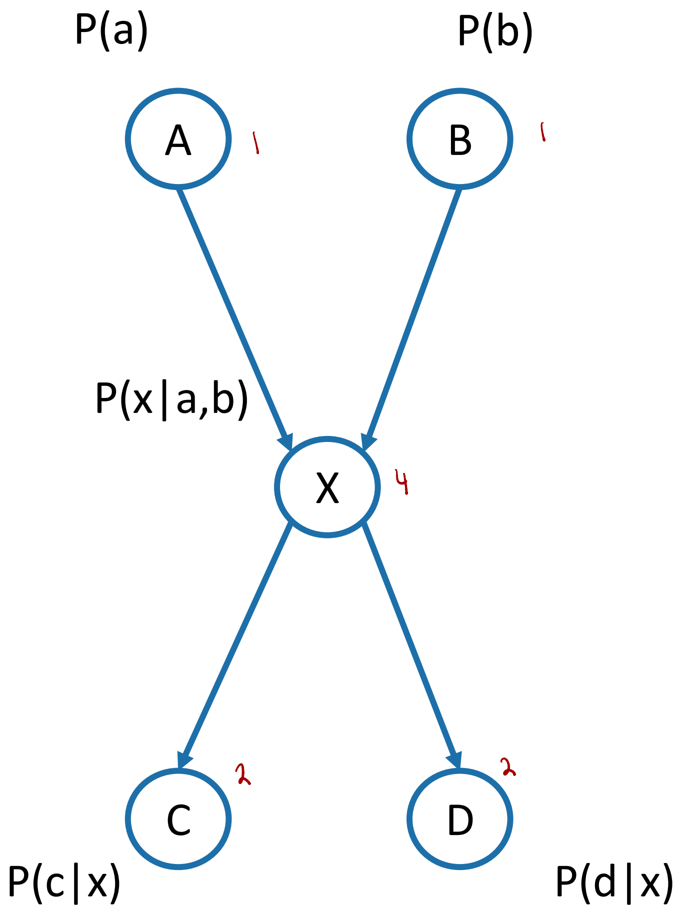
	P(d1|x)
x1	0.3
x2	0.6

Reduction Number of Probabilities

In a domain with N binary propositional variables (2 possibilities value), one needs 2^N numbers to specify the joint probability distribution. N=5: need 32 probabilities

Independence and conditional independence relationships among variables can greatly reduce the number of probabilities that need to be specified in order to define the full joint distribution (Russel & Norvig, 2013)

For 5 binary variables with casual networks: need 2+2+8+4+4=20 probabilities (or 10 with complements).

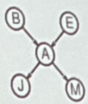


Bayesian Network as Joint Probability Distribution (chain rule)

- The full joint distribution is defined as the product of the local conditional distributions:

$$\mathbf{P}(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

$\boxed{P(A \mid B) = 1 - P(A \mid \neg B)}$



J ketika dikasih A, jadi ga perlu lagi berhubungan dengan B. dst. ⇒ Jadi J dan B conditionally independent.

e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$= \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a)\, \mathbf{P}(a \mid \neg b, \neg e)\, \mathbf{P}(\neg b)\, \mathbf{P}(\neg e)$$

$P(\neg J, M, \neg A, B, E)$

$= P(\neg J \mid \neg A) \cdot P(M \mid \neg A) \cdot P(\neg A \mid B, E) \cdot P(B) \cdot P(E)$

$P(J, M, A, B) = P(J \mid A) \cdot P(M \mid A) \cdot P(A \mid B, E) \cdot P(B) \cdot P(E)$

$+ P(J \mid A) \cdot P(M \mid A) \cdot P(A \mid \underbrace{B, \neg E}) \cdot P(B) \cdot \underbrace{P(\neg E)}$

harus sama

$P(M, A, B, E) = P(J \mid A) \cdot P(M \mid A) \cdot P(A \mid B, E) \cdot P(B) \cdot P(E)$

$+ P(\neg J \mid A) \cdot P(M \mid A) \cdot P(A \mid B, E) \cdot P(B) \cdot P(E)$

$= \underbrace{(P(J \mid A) + P(\neg J \mid A))}_{1} \cdot P(M \mid A) \cdot P(A \mid B, E) \cdot P(B) \cdot P(E)$

$P(A) =$ hanya perlu B dan E saja soalnya yang J dan M nanti jadi 1 juga

$= P(A \mid B, E) \cdot P(B) \cdot P(E) +$

$P(A \mid \neg B, E) \cdot P(\neg B) \cdot P(E) +$

$P(A \mid B, \neg E) \cdot P(B) \cdot P(\neg E) +$

$P(A \mid \neg B, \neg E) \cdot P(\neg B) \cdot P(\neg E)$

Bayesian Network as Joint Probability

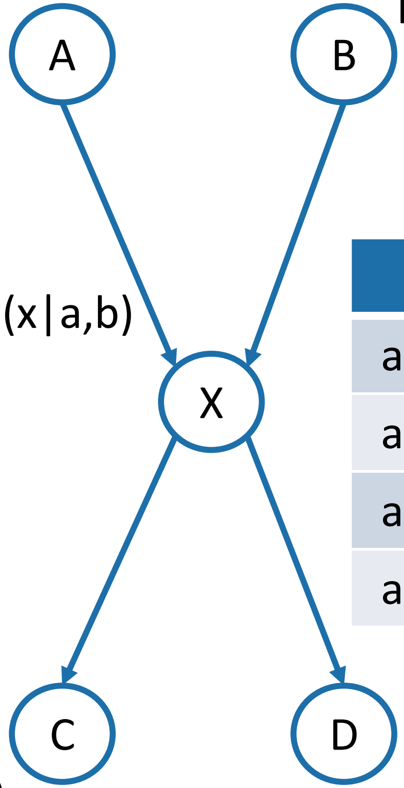
- We can determine the value of any entry in the joint probability.
- P(a2,b1,x2,c2,d1)
  =P(a2)P(b1)P(x2|a2,b1)P(c2|x2)P(d1|x2)
  =0.75*0.6*0.4*0.4*0.6
  =0.0432

P(a1)=0.25

p(a2) = 1-0.25
= 0.75

P(b1)=0.6

p(b2) = 1-0.6
= 0.4



P(x|a,b)

	P(x1|a,b)
a1,b1	0.3
a1,b2	0.7
a2,b1	0.6
a2,b2	0.8

P(c|x)

	P(c1|x)
x1	0.3
x2	0.6

p(c2|x)

x1	1-0.3 = 0.7
x2	1-0.6 = 0.4

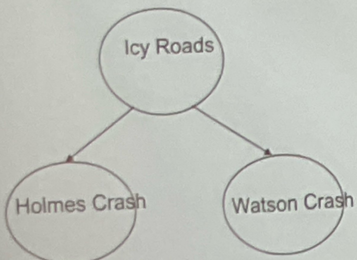
P(d|x)

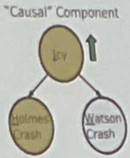
	P(d1|x)
x1	0.3
x2	0.6

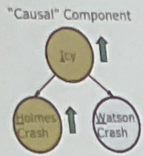
Example of Bayesian Network

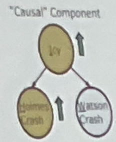
Icy Roads

Inspector Smith is waiting for Holmes and Watson, who are driving (separately) to meet him. It is winter. His secretary tells him that Watson has had an accident. He says, "It must be that the roads are icy. I bet that Holmes will have an accident too. I should go to lunch." But, his secretary says, "No, the roads are not icy, look at the window." So, he says, "I guess I better wait for Holmes."

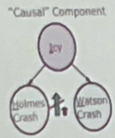






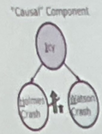


H and W are dependent,



H and W are dependent,

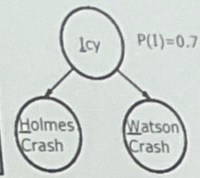
Holmes mungkin kecelakaan tapi informasi watson kecelakaan udah ga pengaruh



H and W are dependent, but conditionally independent given I

Icy Roads with Numbers

	P(H| I)
I	0.8
¬I	0.1

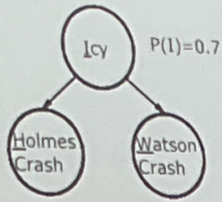


	P(W| I)
I	0.8
¬I	0.1

Probability that Watson Crashes:

$$P(W) = P(W|I)\,P(I) + P(W|\neg I)\,P(\neg I)$$
$$= 0.8 \cdot 0.7 + 0.1 \cdot 0.3$$
$$= 0.56 + 0.03$$
$$= 0.59$$

Icy Roads with Numbers



	P(H| I)
I	0.8
¬I	0.1

	P(W| I)
I	0.8
¬I	0.1

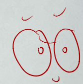
Probability of Icy given Watson (Bayes' Rule):

$$P(I \mid W) = P(W \mid I)\, P(I) \,/\, P(W)$$

→ universenya dah terbatas bukan U lagi. → watson sudah kecelakaan

$$= 0.8 \cdot 0.7 / 0.59$$

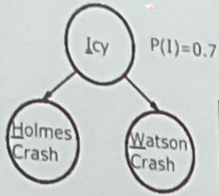
$$= 0.95$$



We started with P(I) = 0.7; knowing that Watson crashed raised the probability to 0.95

Icy Roads with Numbers

	P(H| I)
I	0.8
¬I	0.1



	P(W| I)
I	0.8
¬I	0.1

Probability of Holmes given Watson :

$P(H|W) = P(H|W,I)P(I|W) + P(H|W,\neg I)\ P(\neg I|\ W)$

$= P(H|I)P(I|W) + P(H|\neg I)\ P(\neg I|\ W)$

$= 0.8 \cdot 0.95 + 0.1 \cdot 0.05$

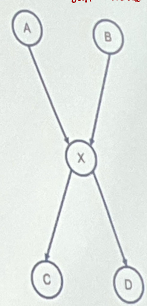
$= 0.765$

We started with P(H) = 0.59; knowing that Watson crashed raised the probability to 0.765

- X represents the fish : x1=salmon and x2=sea bass.
- X is influenced by A and B.
- A represents time of year: a1 = winter, a2 = spring, a3 = summer and a4 = autumn. Probability distribution on A in uniform.
- B represents geographical area where the fish was caught: b1 = north Atlantic and b2 = south Atlantic. The probabilities that any fish came from those areas are 0.6 and 0.4.
- C represents lightness with c1 = light, c2 = medium and c3 = dark
- D represents thickness with d1 = wide and d2 = thin.

kalo ditanya berapa angka
↳ tulis aja semua

if binang kan 20 (10 + complement)
but if case kyk gini just write all.



The probability that the fish was caught in the summer in the north Atlantic and is a sea bass that is dark and thin.

The probability that the fish was caught in the summer (a3) in the north Atlantic (b1) and is a sea bass (x2) that is dark (c3) and thin (d2).

$p(a3, b1, x2, c3, d2)$

Inference: Example

$$P(a3,b1,x2,c3,d2) = P(a3)P(b1)P(x2|a3,b1)P(c3|x2)P(d2|x2) = 0.012$$

P(a)

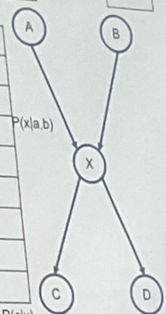
P(a1)	P(a2)	P(a3)	P(a4)
0.25	0.25	0.25	0.25

P(b)

P(b1)
0.6

P(x|a,b)

	P(x1|a,b)
a1,b1	0.3
a1,b2	0.7
a2,b1	0.6
a2,b2	0.8
a3,b1	0.4
a3,b2	0.1
a4,b1	0.2
a4,b2	0.3



P(c|x)

	P(c1|x)	P(c2|x)	P(c3|x)
x1	0.6	0.2	0.2
x2	0.2	0.3	0.5

P(d|x)

	P(d1|x)
x1	0.3
x2	0.6

Classification

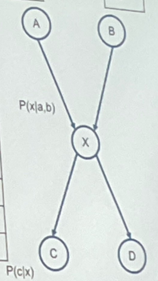
P(a1)	P(a2)	P(a3)	P(a4)
0.25	0.25	0.25	0.25

P(b1)
0.6

Classify the fish that is light (c1) and caught in the south Atlantic (b2), but we do not know what time of year the fish was caught nor its thickness.

Maximum a posterior probability:
P(x1|c1,b2) vs P(x2|c1,b2)

	P(x1|a,b)
a1,b1	0.3
a1,b2	0.7
a2,b1	0.6
a2,b2	0.8
a3,b1	0.4
a3,b2	0.1
a4,b1	0.2
a4,b2	0.3



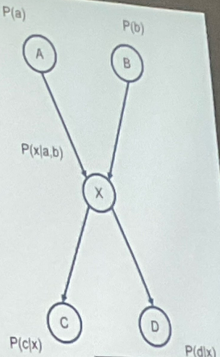
	P(c1|x)	P(c2|x)	P(c3|x)
x1	0.6	0.2	0.2
x2	0.2	0.3	0.5

	P(d1|x)
x1	0.3
x2	0.6

Classification (2)

Q: query
e: evidence of all variables
$P(Q|\mathbf{e})=P(q,e)/P(e)=\alpha P(Q,e)$ $\frac{1}{p(e)}$ dianggap α

$P(x1|c1,b2)=P(x1,c1,b2)/P(c1,b2)$
$=\alpha \sum P(x1,a,b2,c1,d)$
$=\alpha\sum P(a).P(b2).P(x1|a,b2).\ P(c1|x1).\ P(d|x1)$
$=\alpha P(b2).P(c1|x1)\ \sum P(a).\ P(x1|a,b2).\ P(d|x1)$
$=\alpha P(b2).P(c1|x1)\ [\sum P(a).\ P(x1|a,b2)][\sum P(d|x1)]=\alpha\ 0.114$



	P(d1|x)
x1	0.3
x2	0.6

$P(x1|c1,b2)=P(x1,c1,b2)/P(c1,b2)$
$=\alpha P(b2).P(c1|x1).[\sum P(a).P(x1|a,b2)]. [\sum P(d|x1)]= \alpha\ 0.114$

$P(x2|c1,b2)=P(x2,c1,b2)/P(c1,b2)$
$=\alpha P(b2).P(c1|x2)\ [\sum P(a).P(x2|a,b2)][\sum P(d|x2)]= \alpha\ 0.042$

Normalize:

$P(x1|c1,b2)=0.73$ $: \frac{0.114\alpha}{0.114\alpha+0.042\alpha}$

$P(x2|c1,b2)=0.27$ $: \frac{0.0.42\alpha}{0.114\alpha+0.042\alpha}$

Decision: x1=salmon

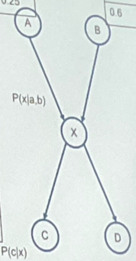
P(a)

P(a1)	P(a2)	P(a3)	P(a4)
0.25	0.25	0.25	0.25

P(b)

P(b1)
0.6

	P(x1|a,b)
a1,b1	0.3
a1,b2	0.7
a2,b1	0.6
a2,b2	0.8
a3,b1	0.4
a3,b2	0.1
a4,b1	0.2
a4,b2	0.3

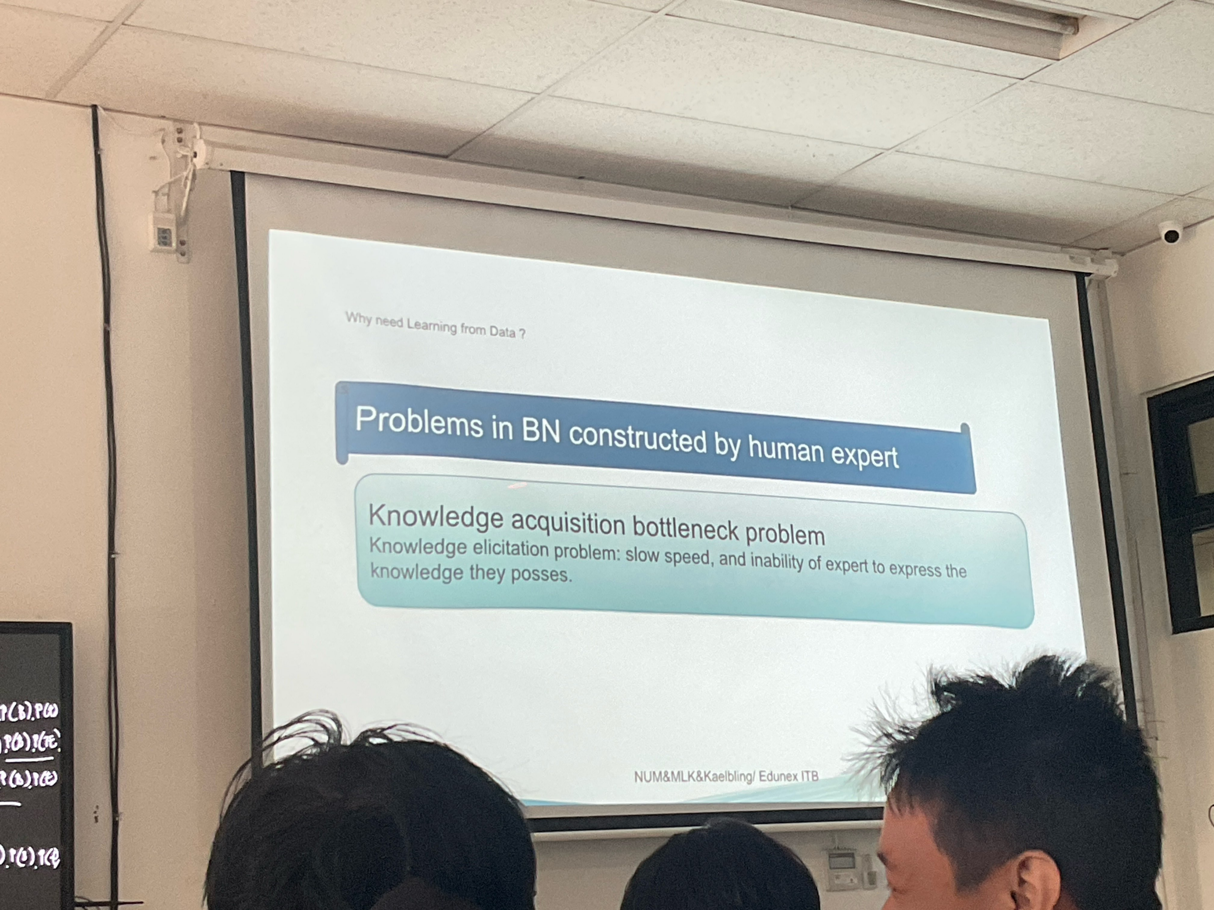


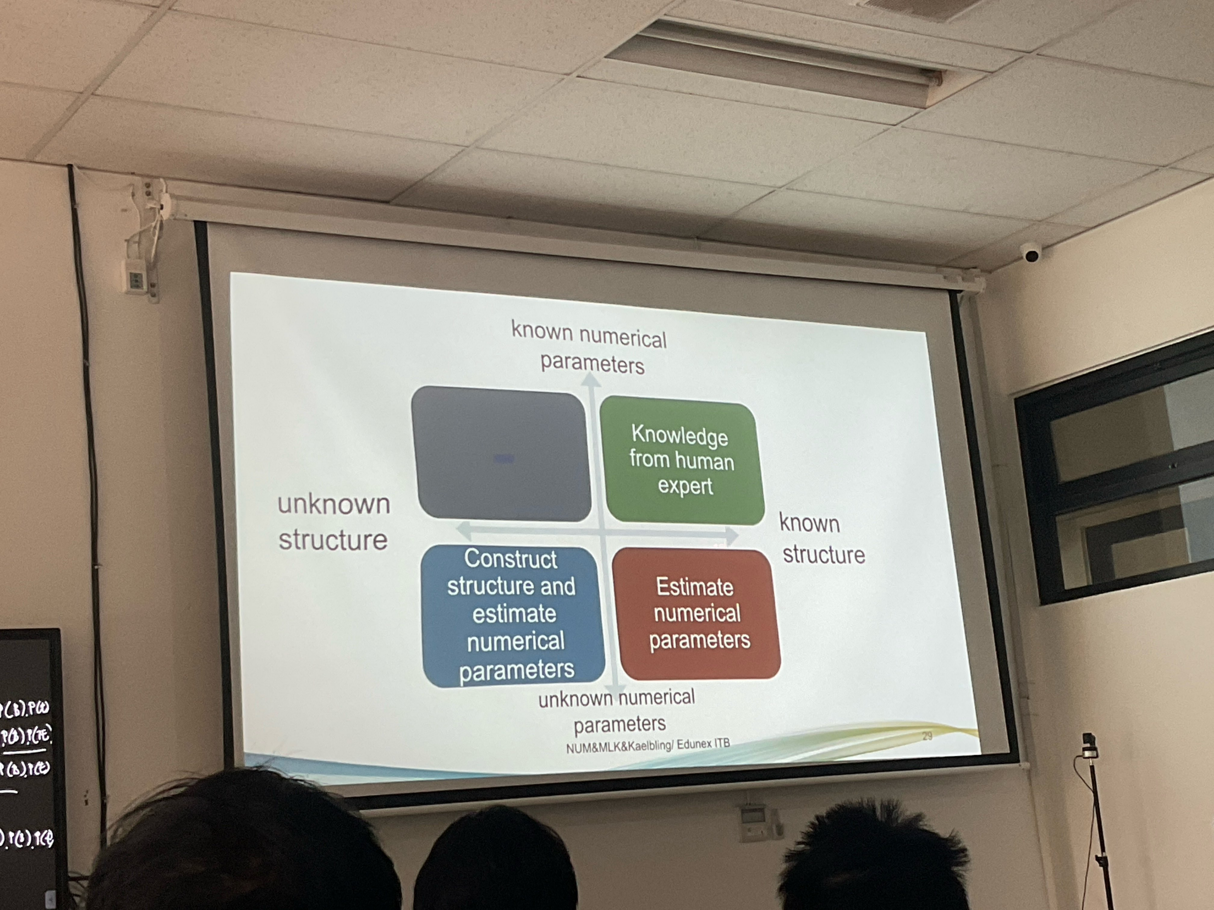
P(c|x)

	P(c1|x)	P(c2|x)	P(c3|x)
x1	0.6	0.2	0.2
x2	0.2	0.3	0.5

P(d|x)

	P(d1|x)
x1	0.3
x2	0.6


Why need Learning from Data ?
Problems in BN constructed by human expert
Knowledge acquisition bottleneck problem
Knowledge elicitation problem: slow speed, and inability of expert to express the knowledge they posses.
NUM&MLK&Kaelbling/ Edunex ITB


known numerical parameters
Knowledge from human expert
unknown structure
known structure
Construct structure and estimate numerical parameters
Estimate numerical parameters
unknown numerical parameters
NUM&MLK&Kaelbling/ Edunex ITB
29

Parameter Estimation

Given structure with m nodes

Given a data set

$D = \{<v_1^1,..,v_m^1>,..,<v_1^k,..,v_m^k>\}$

Count $\#(V_i{=}T)$, $\#(V_i{=}F)$, $\#(V_i{=}T, V_j{=}T)$, $\#(V_i{=}T, V_j{=}F)$

Variable V_i with no parent

$$P(V_i) \approx \frac{\#(V_i = T)}{k}$$

Variable V_i with parent V_j

$$P(V_i|V_j) \approx \frac{\#(V_i = T, V_j = T)}{\#(V_j = T)}$$

$$P(V_i|\neg V_j) \approx \frac{\#(V_i = T, V_j = F)}{\#(V_j = F)}$$

Smoothing: Avoid Probability = 0

Variable V_i with no parent

$$P(V_i) \approx \frac{\#(V_i = T) + 1}{k + 2}$$

Variable V_i with parent V_j

$$P(V_i|V_j) \approx \frac{\#(V_i = T, V_j = T) + 1}{\#(V_j = T) + 2}$$

$$P(V_i|\neg V_j) \approx \frac{\#(V_i = T, V_j = F) + 1}{\#(V_j = F) + 2}$$

Summary

Bayesian Network

BN vs Joint Probability

Classification using BN

THANK YOU