INSTITUT TEKNOLOGI BANDUNG

1920

Modul 4: Bayesian Network

01 What & Why

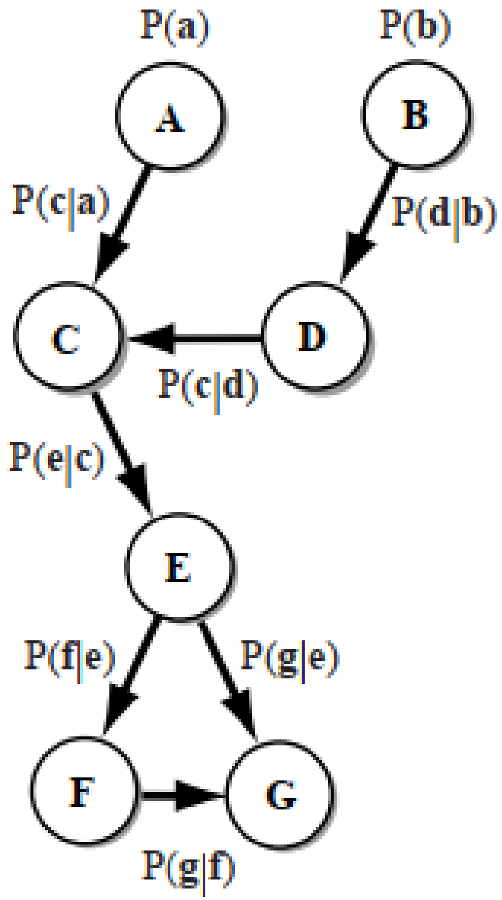
Masayu Leylia Khodra
(masayu@informatika.org)

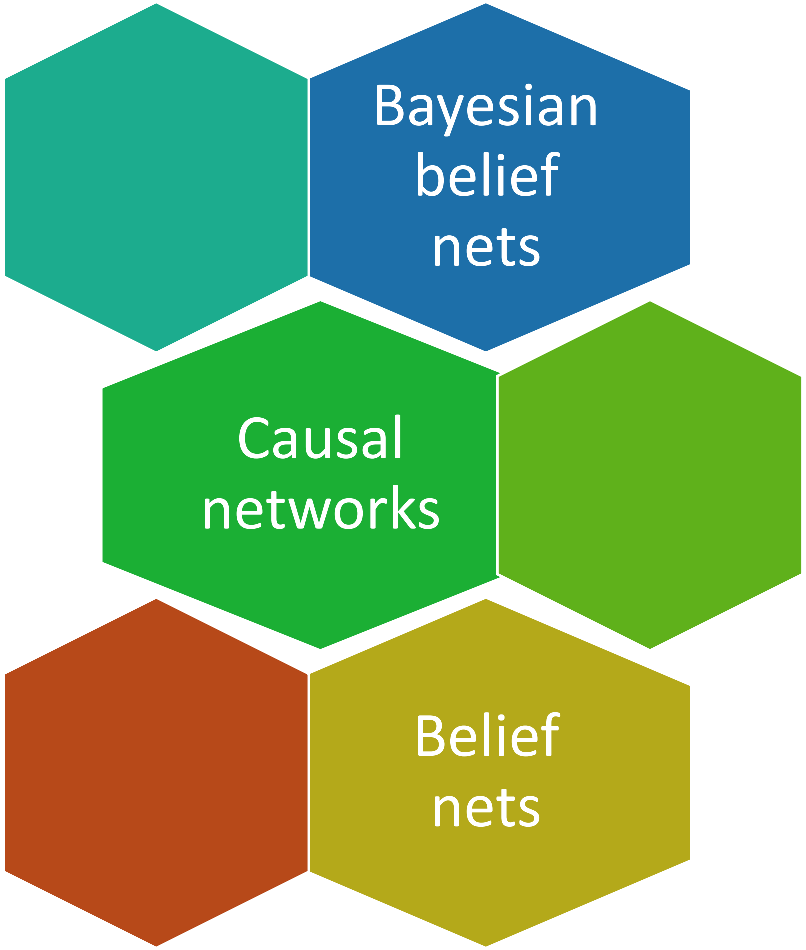
KK IF – Teknik Informatika – STEI ITB

Pengenalan Pola
(*Pattern Recognition*)

Bayesian Networks: What

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





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

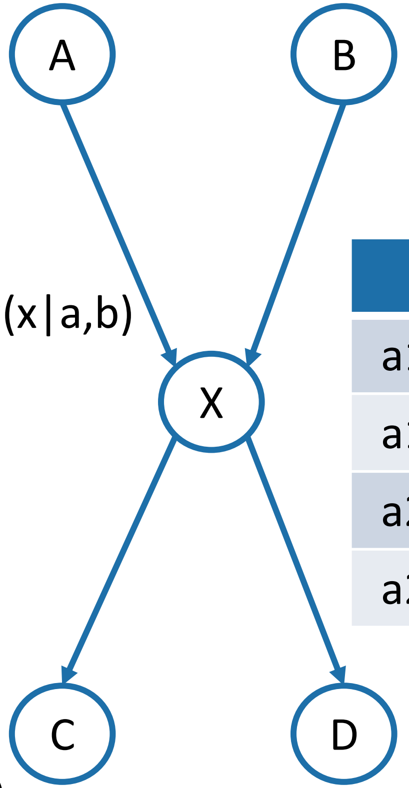
- Node (variables)
- Directed arcs (acyclic graph)

Numerical parameters

- Prior probability
- Probability conditional tables

P(a1)=0.25

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

P(c|x)

	P(c1|x)
x1	0.3
x2	0.6

P(d|x)

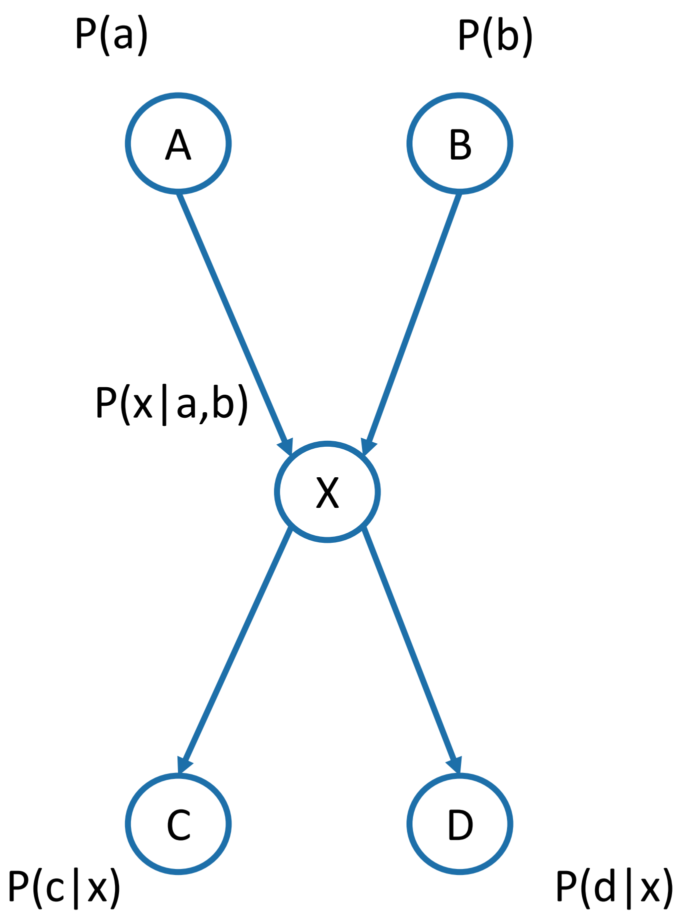
	P(d1|x)
x1	0.3
x2	0.6

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

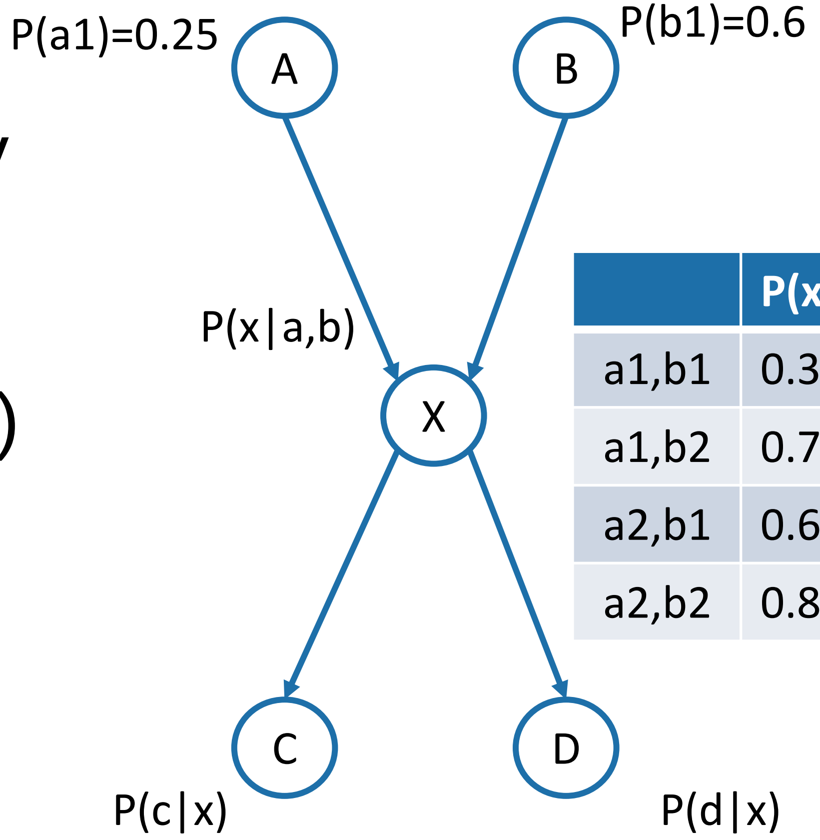
Independence and conditional independence relationships among variables can greatly reduce the number of probabilities that need to be specified in order to define the full joint distribution (Russel & Norvig, 2013)

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



Bayesian Network as Joint Probability

- 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(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

Summary

Bayesian Network

BN vs Joint Probability

Classification using BN

Modul 4: Bayesian Network

02 Classification using BN

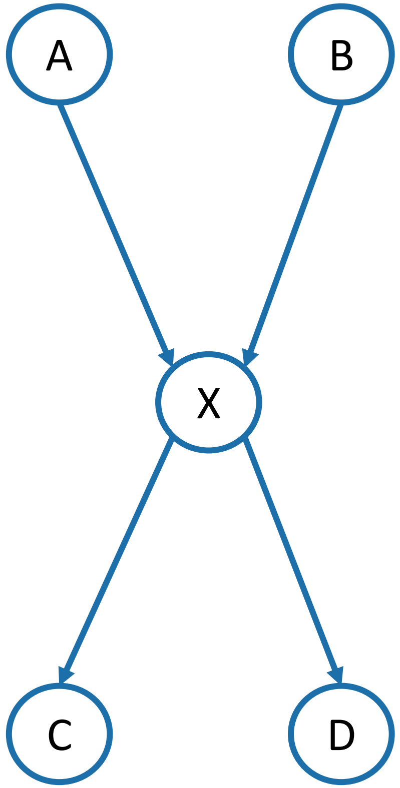
Masayu Leylia Khodra
(masayu@informatika.org)

KK IF – Teknik Informatika – STEI ITB

Pengenalan Pola
(*Pattern Recognition*)

Belief Network from Human Expert

- 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.



Inference in Bayesian Network

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)

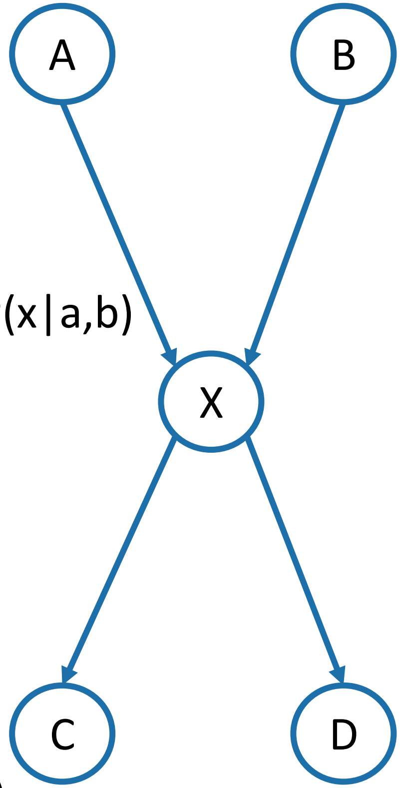
P(a1)	P(a2)	P(a3)	P(a4)
0.25	0.25	0.25	0.25

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

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.

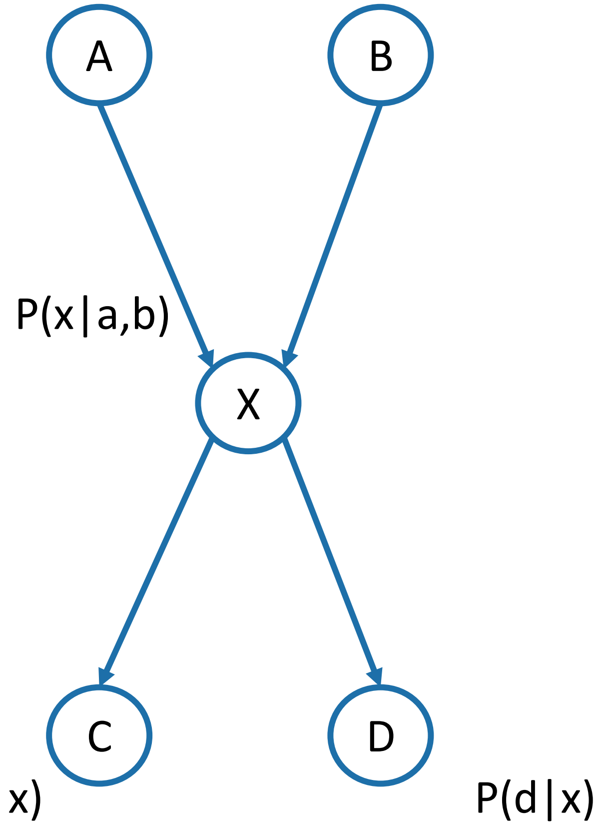
Maximum a posterior probability:
P(x1|c1,b2) vs P(x2|c1,b2)

P(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(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

$P(Q|\mathbf{e})=P(q,e)/P(e)=\alpha P(Q,e)$

$P(x1|c1,b2)=P(x1,c1,b2)/P(c1,b2)$

$=\alpha \sum P(x1,\mathbf{a},b2,c1,\mathbf{d})$

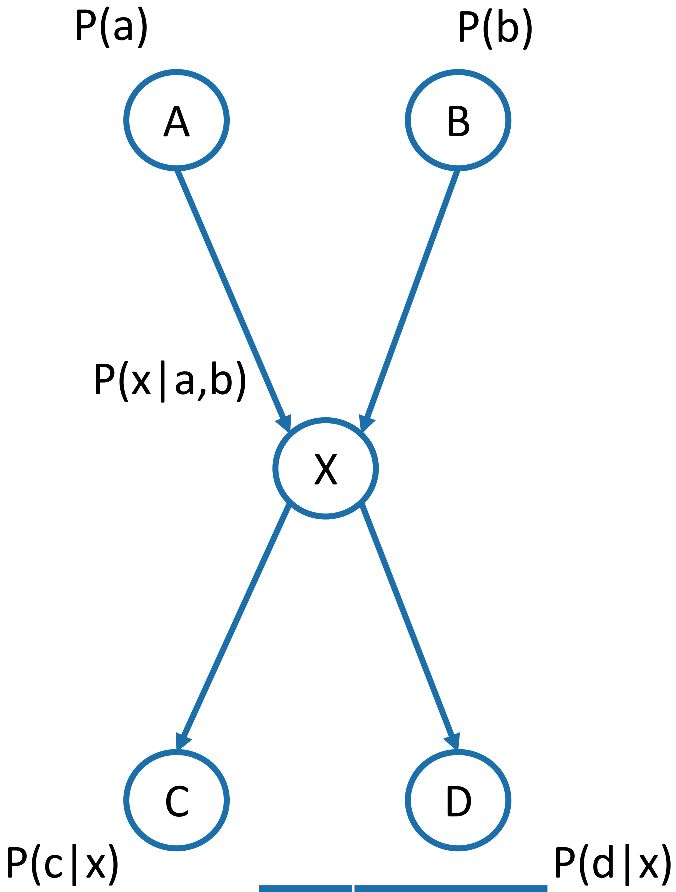
$=\alpha\sum P(\mathbf{a}).P(b2).P(x1|\mathbf{a},b2).\ P(c1|x1).\ P(\mathbf{d}|x1)$

$=\alpha P(b2).P(c1|x1)\ \sum P(\mathbf{a}).\ P(x1|\mathbf{a},b2).\ P(\mathbf{d}|x1)$

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

$P(a1).\ P(x1|a1,b2)+$
$P(a2).\ P(x1|a2,b2)+$
$P(a3).\ P(x1|a3,b2)+$
$P(a4).\ P(x1|a4,b2)$

$P(d1|x1)+P(d2|x1)=1.0$



	P(d1|x)
x1	0.3
x2	0.6

Classification (3)

P(x1|c1,b2)=P(x1,c1,b2)/P(c1,b2)
=αP(b2).P(c1|x1).[∑P(a).P(x1|a,b2)].
[∑P(d|x1)]= α 0.114

P(x2|c1,b2)=P(x2,c1,b2)/P(c1,b2)
=αP(b2).P(c1|x2) [∑P(a).
P(x2|a,b2)][∑P(d|x2)]= α 0.042

Normalize:
P(x1|c1,b2)=0.73
P(x2|c1,b2)=0.27
Decision: x1=salmon

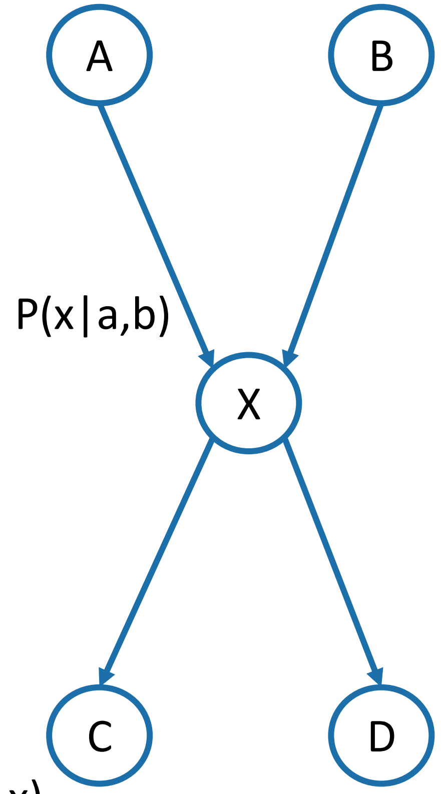
P(a)

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

Summary

BN from human expert

Inference complete evidence

Inference incomplete evidence

INSTITUT TEKNOLOGI BANDUNG

1920