Elective Module:

ML Applications

Chapter 12

PGMs:

Probabilistic Graphical Models

PGM

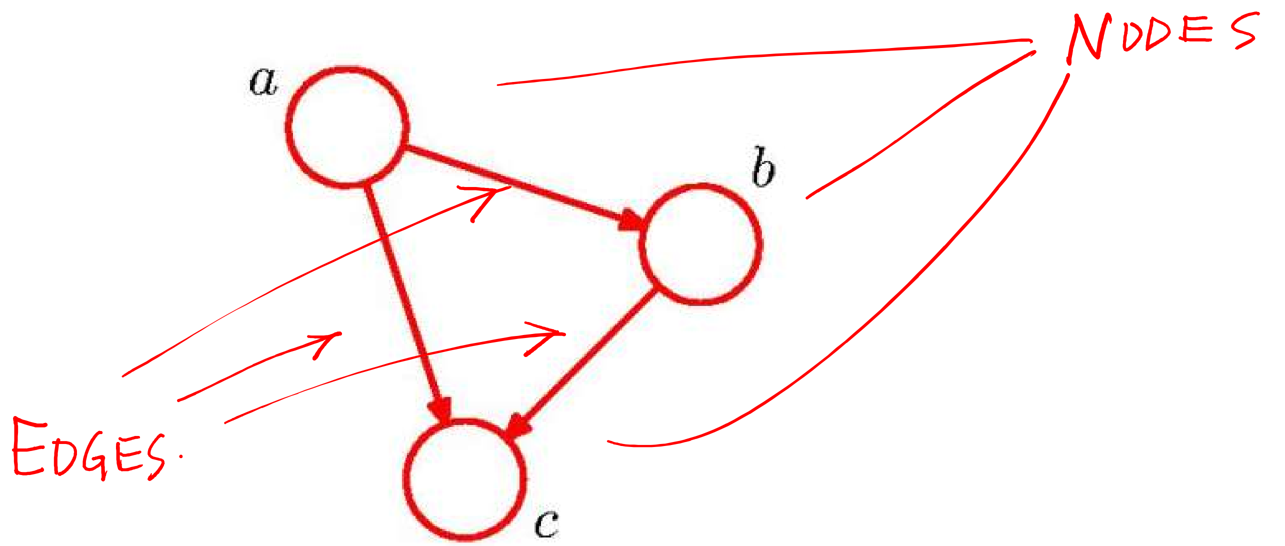
- Probabilistic Graphical Models
- These are VISUAL REPRESENTATION of probability distributions
- Fusion of probability and graph theories

PGM

- Probabilistic Graphical Models
- These are visual representations of probability distributions
- Fusion of probability and graph theories

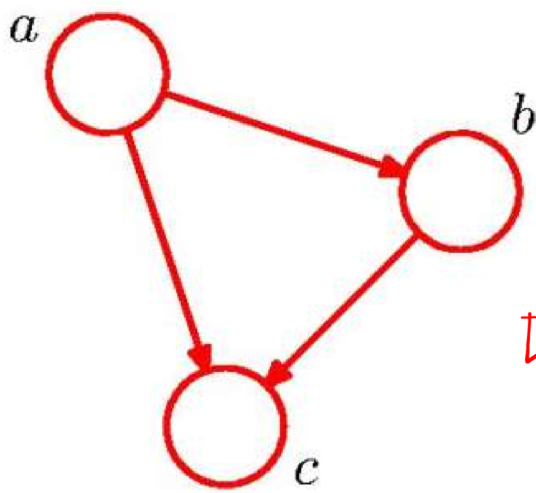
Graph

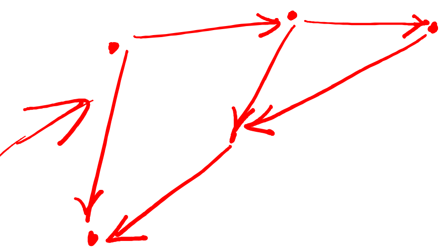
- Graph consists of ___EDGES and NODES.___



Graph

- Graph consists of nodes and edges

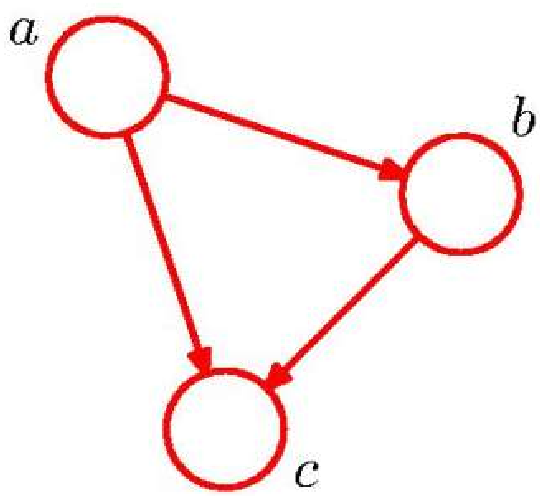




DIRECTED EDGES.

PGM

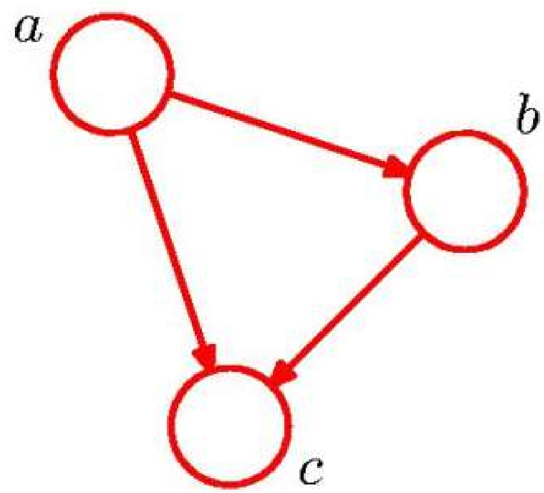
- Each node represents a Random variable.



Probabilistic Graphical Model.

PGM

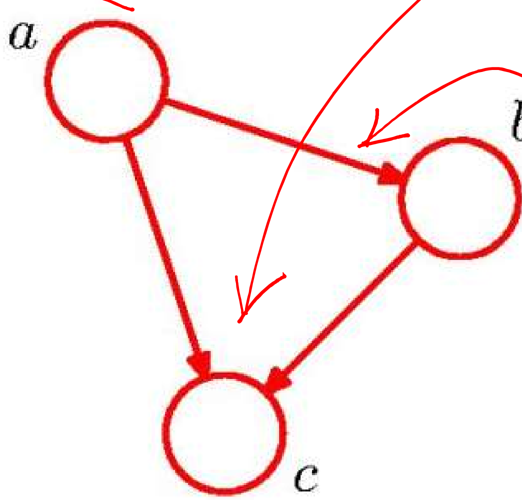
- Each node represents a random variable.



PGM

- Edge represents a Probabilistic Relationship.

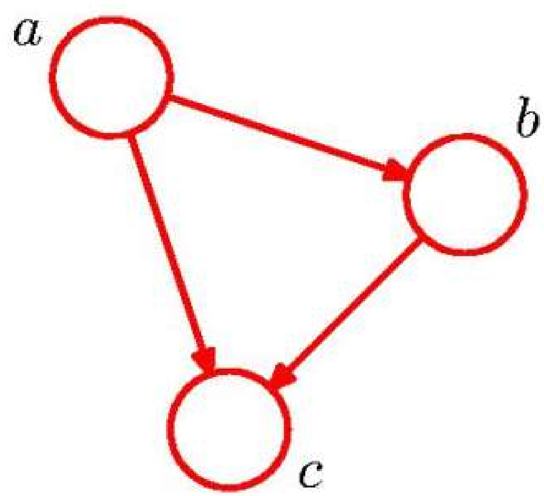
RVs



Probabilistic Relationship.

PGM

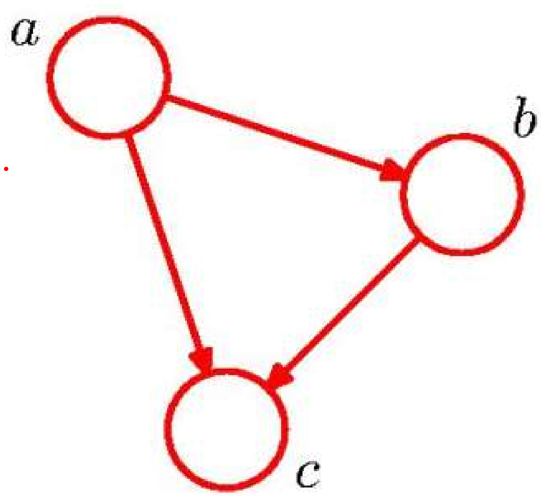
- Edge represents a probabilistic relationship.



Bayesian Networks (BNs)

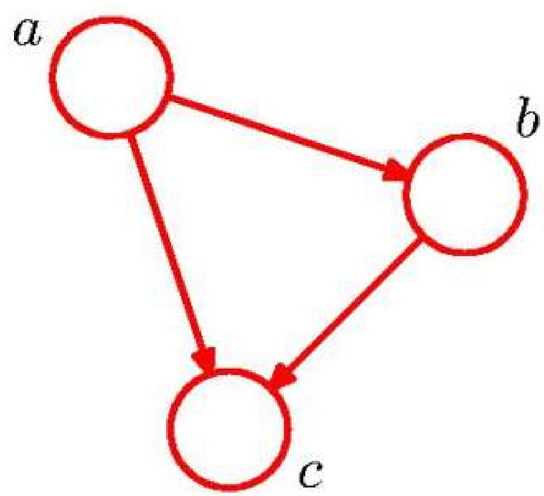
Types of PGMs.

- These are Directed Graphical models.
- Arrows show Directionality.



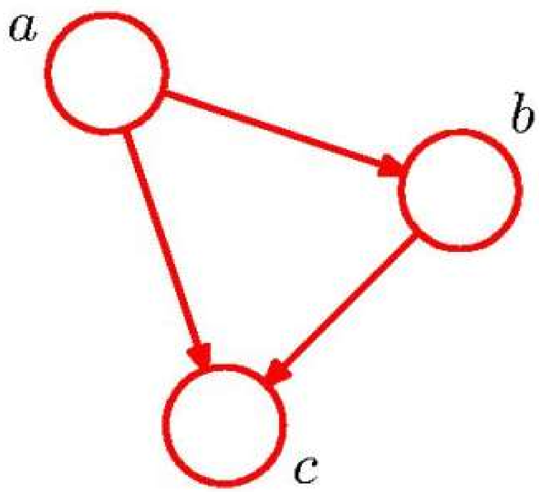
Bayesian Networks (BNs)

- These are directed graphical models.
- Arrows show Directionality.



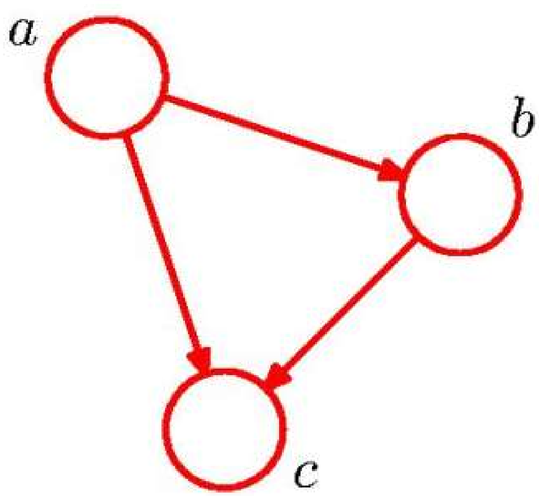
Bayesian Networks (BNs)

- Arrows capture Causal Relationships. between random variables.



Bayesian Networks (BNs)

- Arrows capture Causal relationships between random variables.



Joint distribution

- Consider random variables

$$x_1, x_2, \ldots, x_k$$

$$p(x_1, x_2, \ldots, x_k) = ?$$

Joint distribution

- Consider random variables

$$x_1, x_2, \ldots, x_K$$

Joint distribution

- The joint PDF can be simplified as

$$p(x_1, x_2, \dots, x_K)$$

$$= p(x_1) \times p(x_2|x_1) \times p(x_3|x_1, x_2) \times \cdots \times p(x_K|x_1, x_2, \dots, x_{K-1}).$$

Holds for all RVs.

Chain Rule.

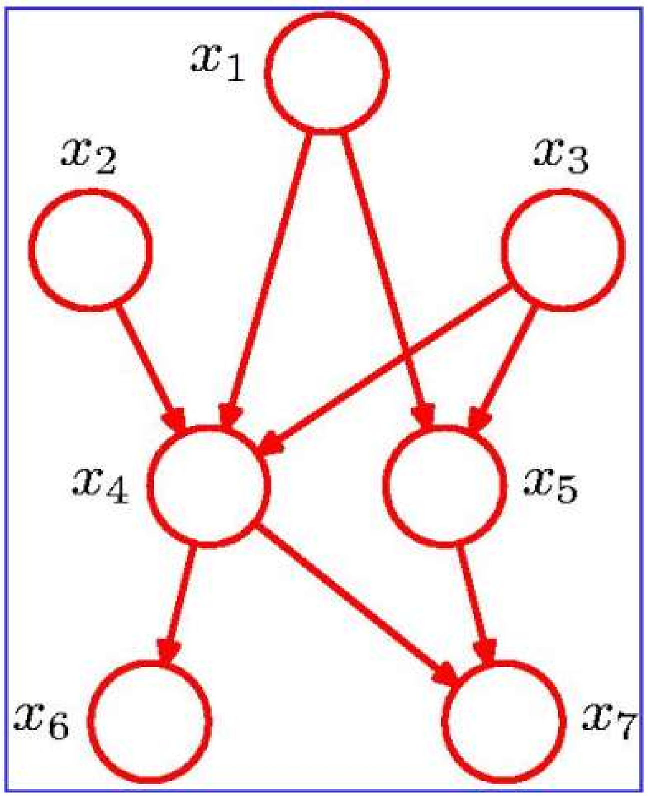
Joint distribution

- The joint PDF can be simplified as

$$p(x_1, x_2, \ldots, x_K)$$
$$= p(x_1) \times p(x_2|x_1) \times p(x_3|x_1, x_2) \times$$
$$\ldots p(x_K|x_1, x_2, \ldots, x_{K-1})$$

Example BN

Valid For any $x_1, x_2, \ldots, x_6, x_7$.

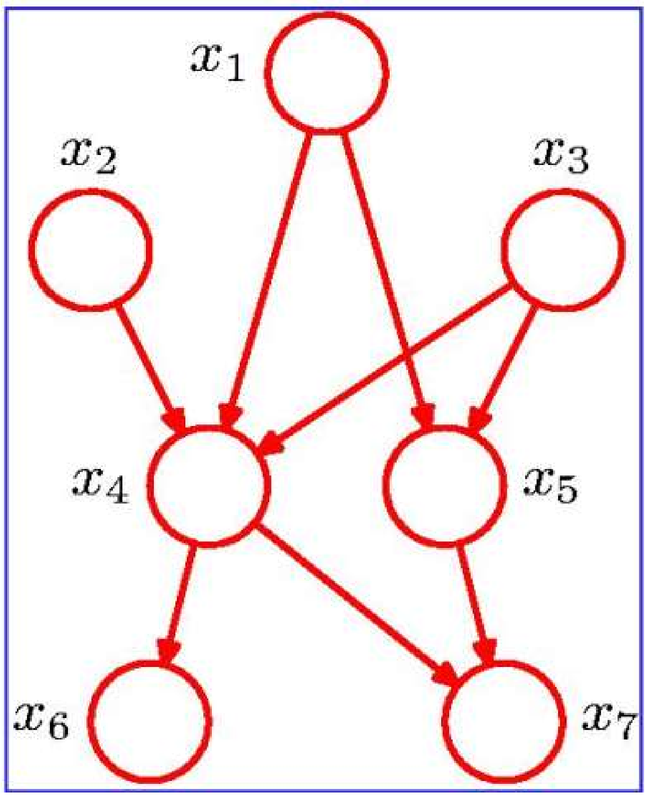


- The joint PDF for this can be simplified as follows

$$p(x_1, x_2, x_3, x_4, x_5, x_6, x_7) =$$

$$= p(x_1) \times p(x_2 | x_1) \times p(x_3 | x_1, x_2)$$
$$\times p(x_4 | x_1, x_2, x_3) \times p(x_5 | x_1, x_2, x_3, x_4)$$
$$\times p(x_6 | x_1, x_2, x_3, x_4, x_5) .$$
$$\times p(x_7 | x_1, x_2, x_3, x_4, x_5, x_6)$$

Valid for any $x_1, x_2, \dots, x_7$



Example BN

- The joint PDF for this can be simplified as follows

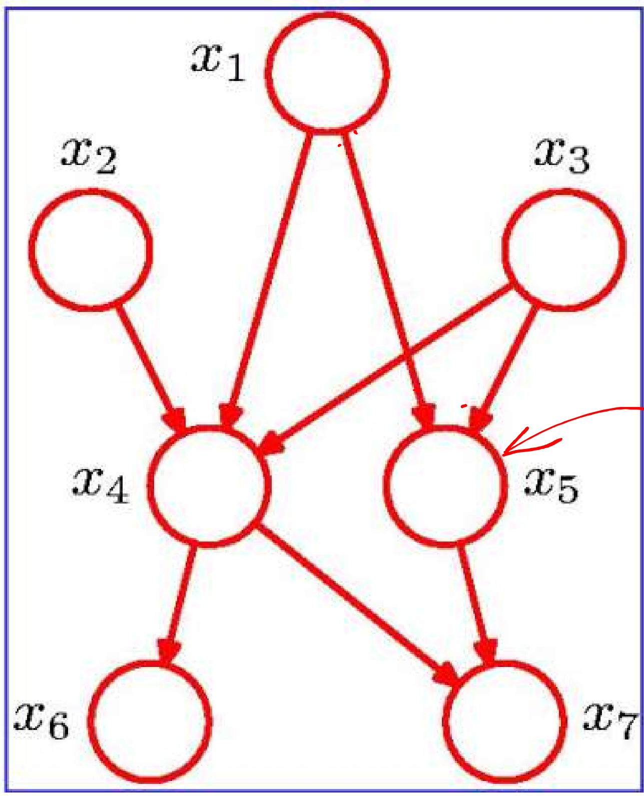
$$p(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$
$$= p(x_1) \times p(x_2|x_1) \times p(x_3|x_1, x_2) \times p(x_4|x_1, x_2, x_3) \times p(x_5|x_1, x_2, x_3, x_4) \times p(x_6|x_1, x_2, x_3, x_4, x_5) \times p(x_7|x_1, x_2, x_3, x_4, x_5, x_6)$$

BN Property

- Following property holds for BN,

$$p(x_k \mid x_1, x_2, \ldots x_{k-1}) = p(x_k \mid P_k).$$

- $\mathcal{P}_k$ denotes parents of $\boldsymbol{x_k}$

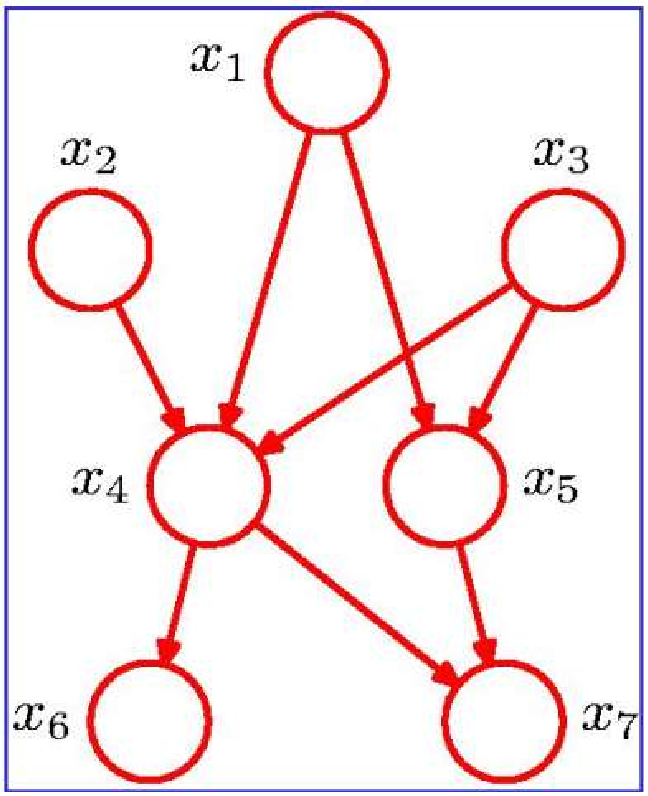


$P_5 = \{x_1, x_3\}$

$P_7 = \{x_4, x_5\}$

BN Property

- Following property holds for BN,
  $p(x_k|x_1, x_2, \ldots, x_{k-1}) = p(x_k|\mathcal{P}_k)$

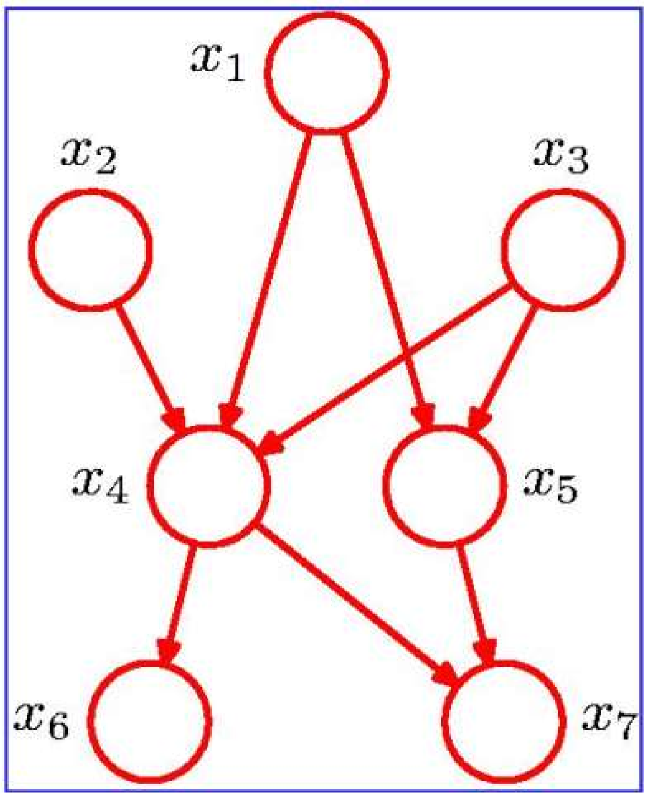


- $\mathcal{P}_k$ denotes parents of $\boldsymbol{x_k}$

Joint PDF

$$p(x_k|x_1, x_2, \ldots, x_{k-1}) = p(x_k|\mathcal{P}_k)$$

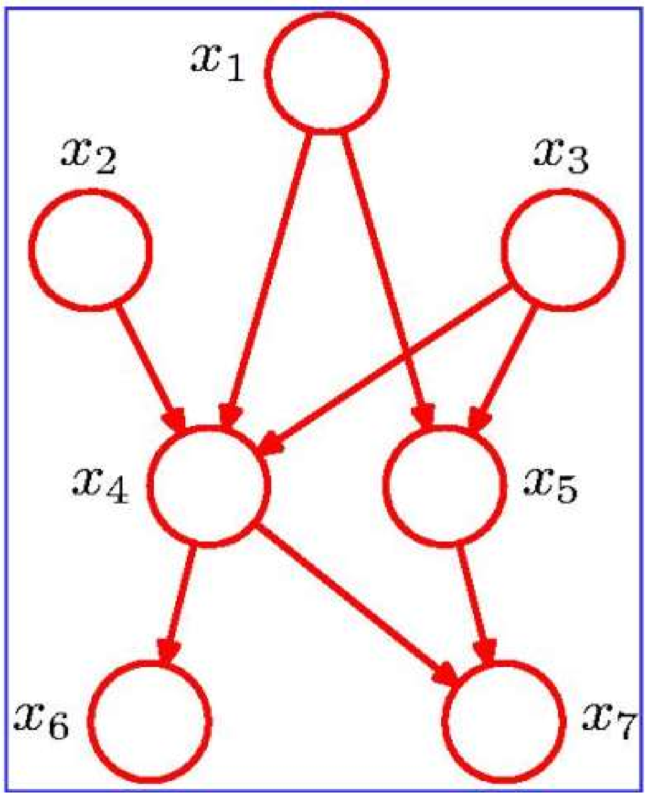
- Given parents P_k, $\boldsymbol{x_k}$ is Conditionally independent. of others!



Joint PDF

$$p(x_k|x_1, x_2, \ldots, x_{k-1}) = p(x_k|\mathcal{P}_k)$$

- Given parents, $\boldsymbol{x_k}$ is conditionally independent of others!



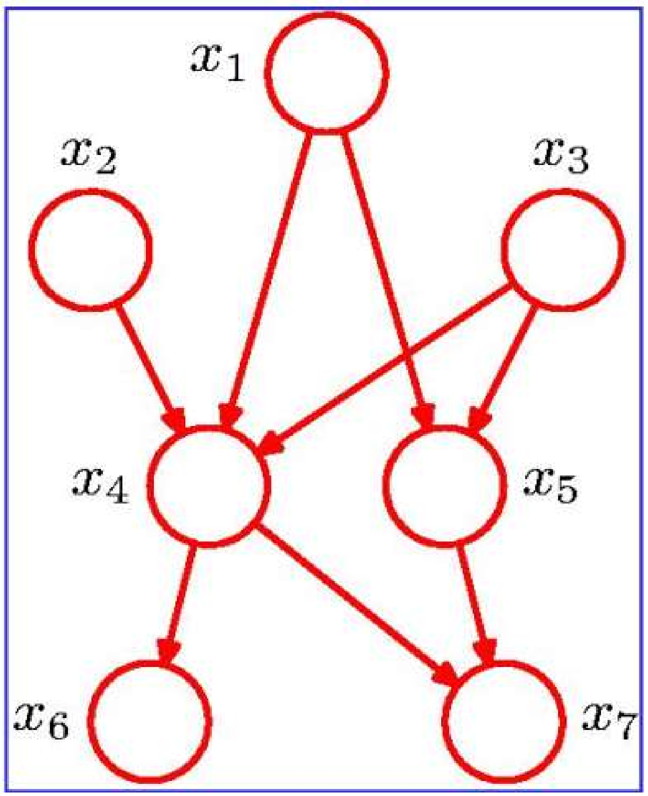
Joint PDF

- Therefore, for this graph

P_5

$$p(x_2|x_1) = p(x_2).$$

$$p(x_5|x_1, x_2, x_3, x_4) = p(x_5|x_1, x_3).$$

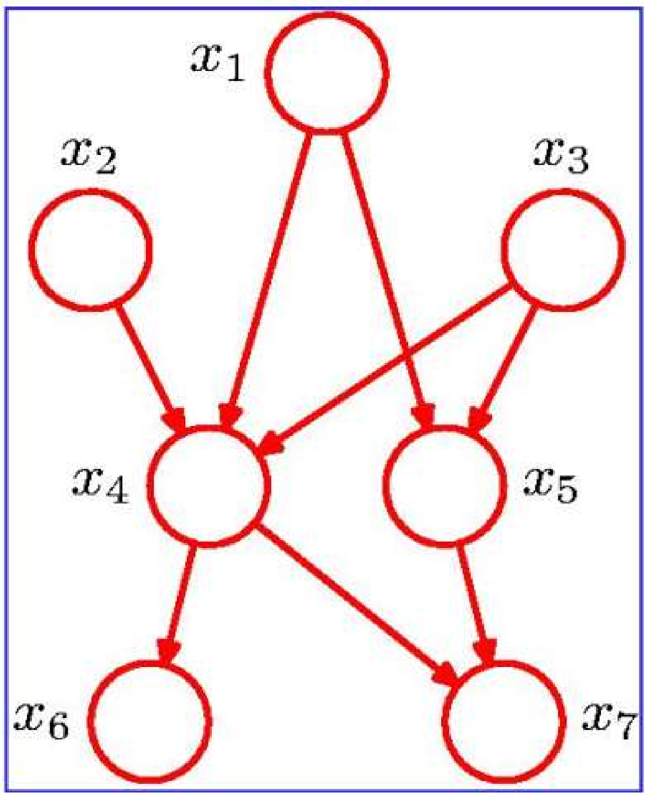


Joint PDF

- Therefore, for this graph

$$p(x_2|x_1) = p(x_2)$$

$$p(x_5|x_1, x_2, x_3, x_4) = p(x_5|x_1, x_3)$$



Joint distribution

- The joint PDF can be simplified as

$$p(x_1, x_2, \dots, x_K) = \prod_{k=1}^{K} \underbrace{p(x_k \mid P_{k\cdot})}_{\text{Factor}} .$$

BN principle : Key principle

Joint distribution

Conditional Factors.

- The joint PDF can be simplified as

$$p(x_1, x_2, \ldots, x_K) = \prod_{k=1}^{K} p(x_k | \mathcal{P}_k)$$

Joint PDF

- Therefore, joint PDF can be simplified as

Simplifies joint PDF

$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

$$= p(x_1) \times p(x_2) \times p(x_3) \times p(x_4 | x_1, x_2, x_3) \times p(x_5 | x_1, x_3) \times p(x_6 | x_4) \times p(x_7 | x_4 | x_5).$$

Joint PDF

- Therefore, joint PDF can be simplified as

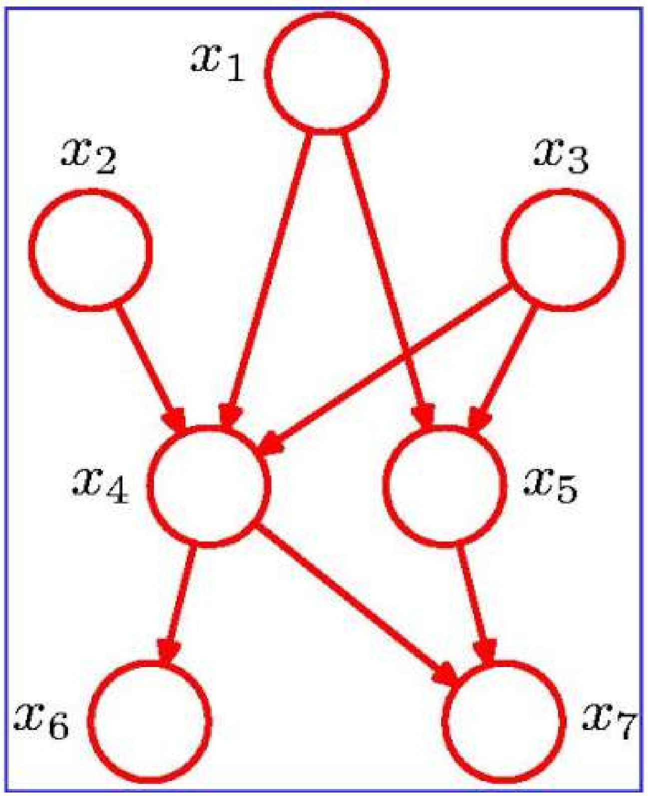
$$p(x_1, x_2, x_3, x_4, x_5, x_6)$$

=

Joint PDF

- Therefore, joint PDF can be simplified as

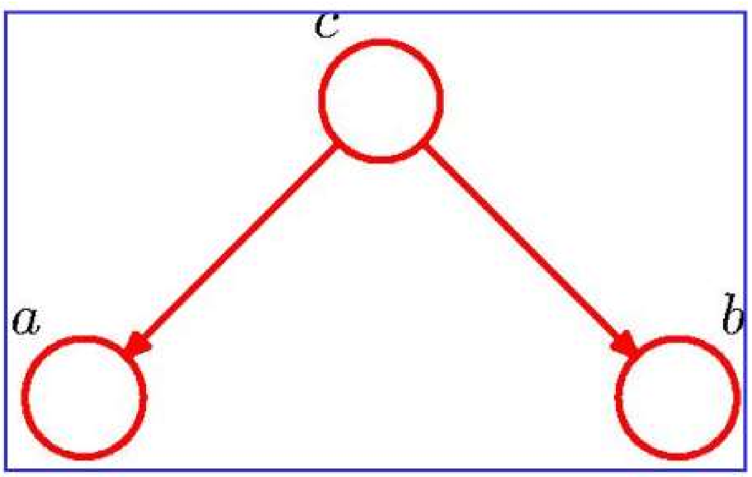
$$
\begin{gathered}
p(x_1, x_2, x_3, x_4, x_5, x_6) \\
= p(x_1) \times p(x_2) \times p(x_3) \times \\
p(x_4|x_1, x_2, x_3) \times p(x_5|x_1, x_3) \\
\times p(x_6|x_4) \times p(x_7|x_4, x_5)
\end{gathered}
$$



Simple examples

- Joint PDF is given as

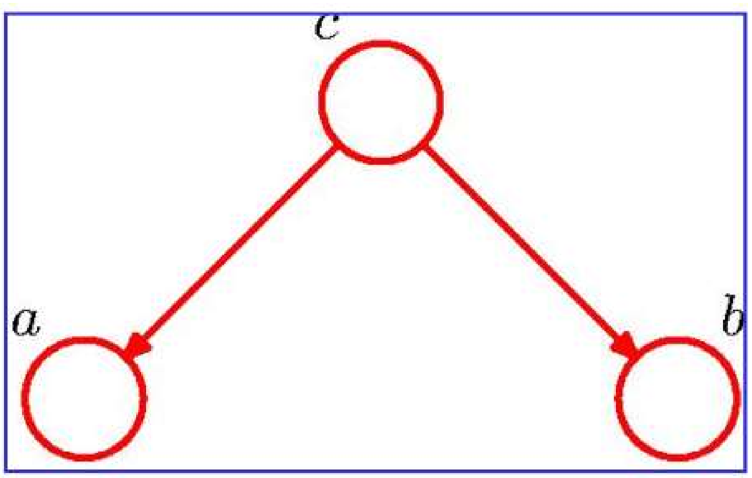
$p(a, b, c) =$ $p(c) \times p(a|c) \times p(b|c)$.



Simple examples

- Joint PDF is given as

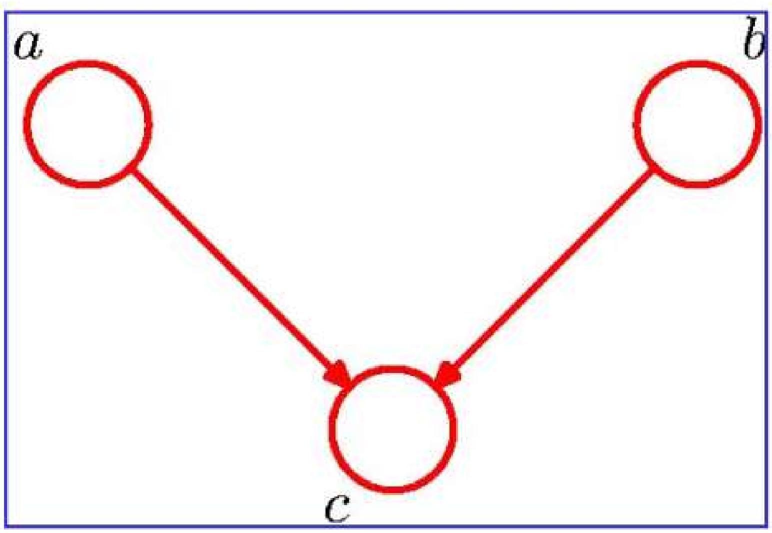
$$p(a,b,c) = p(c) \times p(a|c) \times p(b|c)$$



Simple examples

- Joint PDF is given as

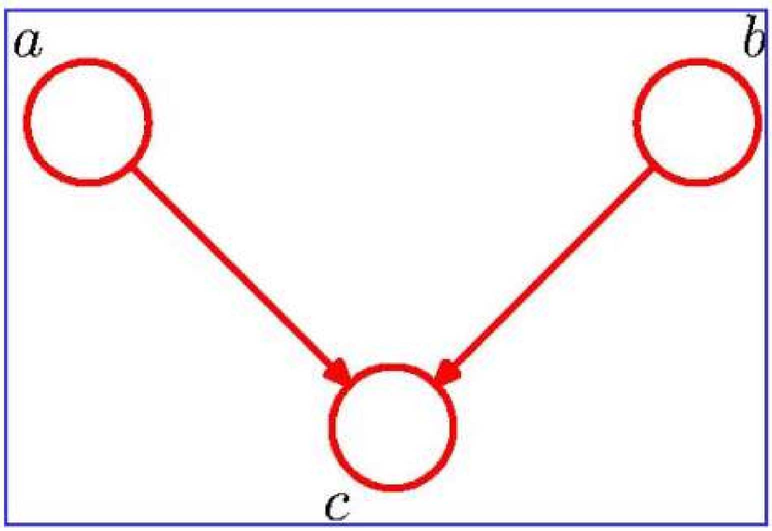
$p(a, b, c) = p(a) \times p(b) \times p(c|a,b)$.



Simple examples

- Joint PDF is given as

$$p(a, b, c) = p(a) \times p(b) \times p(c|a, b)$$

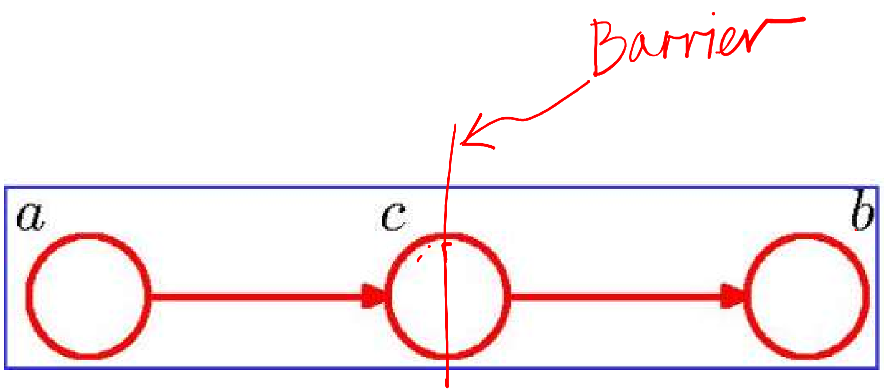


Simple examples

- Joint PDF is given as

$$p(a, b, c) = p(a) \cdot p(c|a)\, p(b|c)$$

- This is a Markov chain.



Simple examples

- Joint PDF is given as

$$p(a, b, c)$$
$$= p(a) \times p(c|a) \times p(b|c)$$

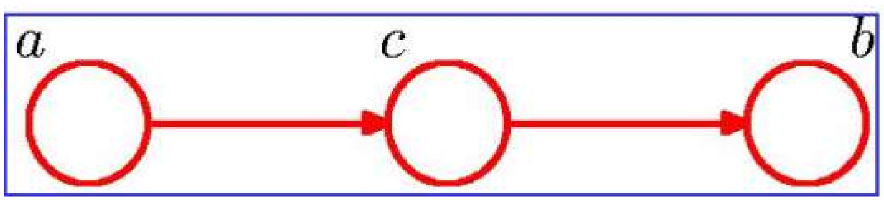
immediate past

history

$$p(x_k \mid \underbrace{x_1, x_2, \ldots, x_{k-1}}_{}) = p(x_k \mid \underline{x_{k-1}}).$$

- This is a Markov chain.

$$P_k = \{x_{k-1}\}.$$



BN Example

- Consider the BN example shown

Student Example :

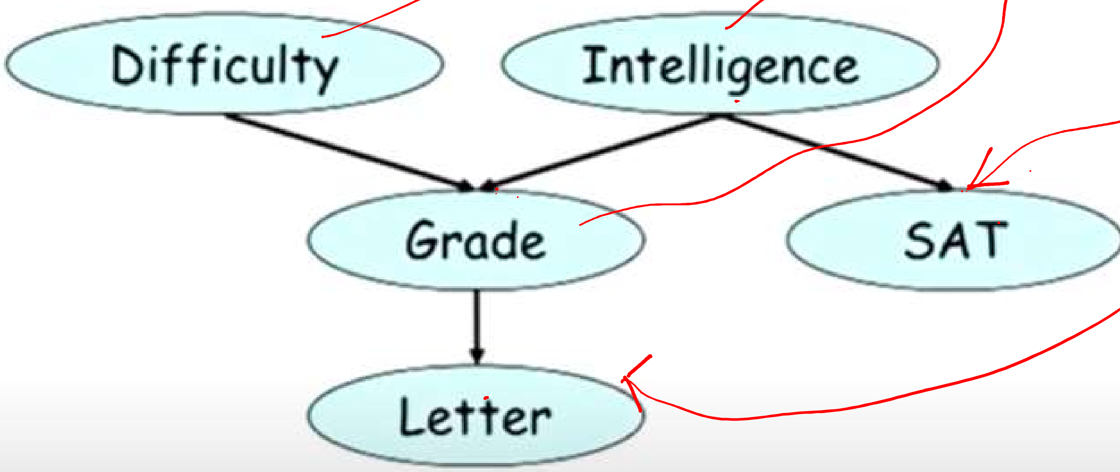
Course

Student

Grade in course

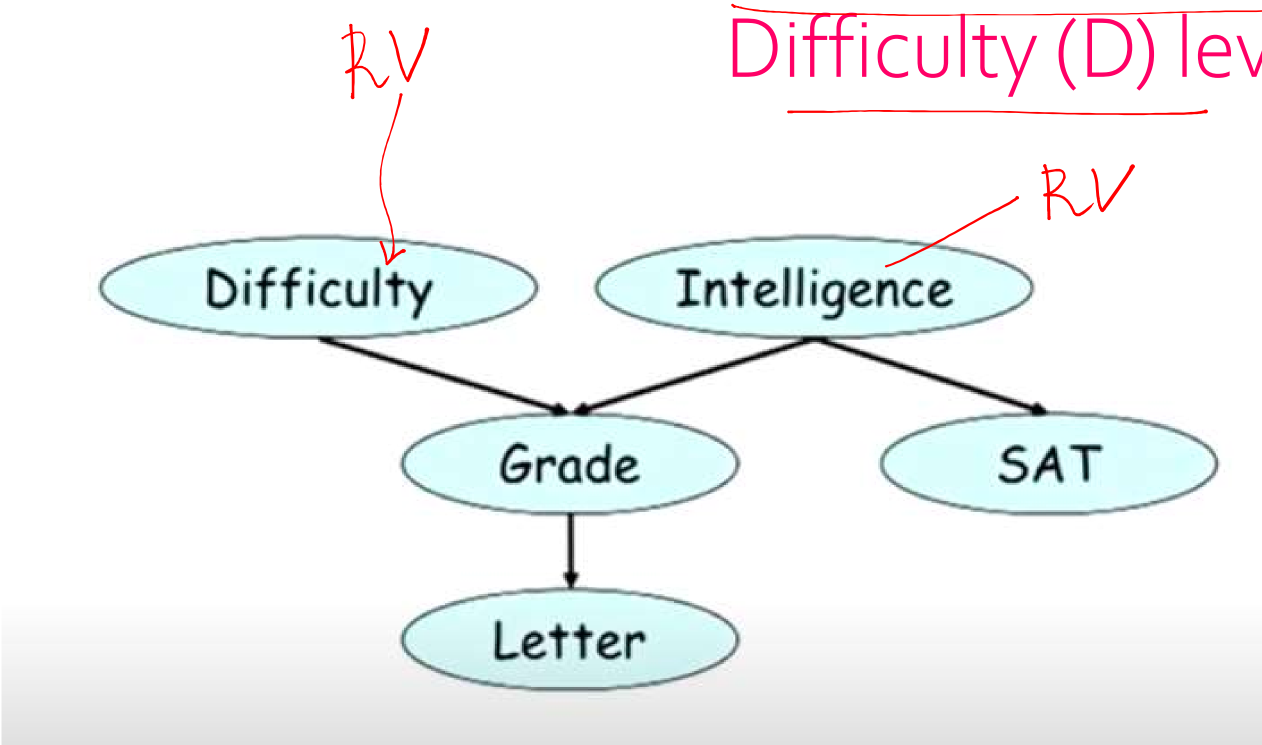
Reference Letters.

SAT Exam.



BN Example

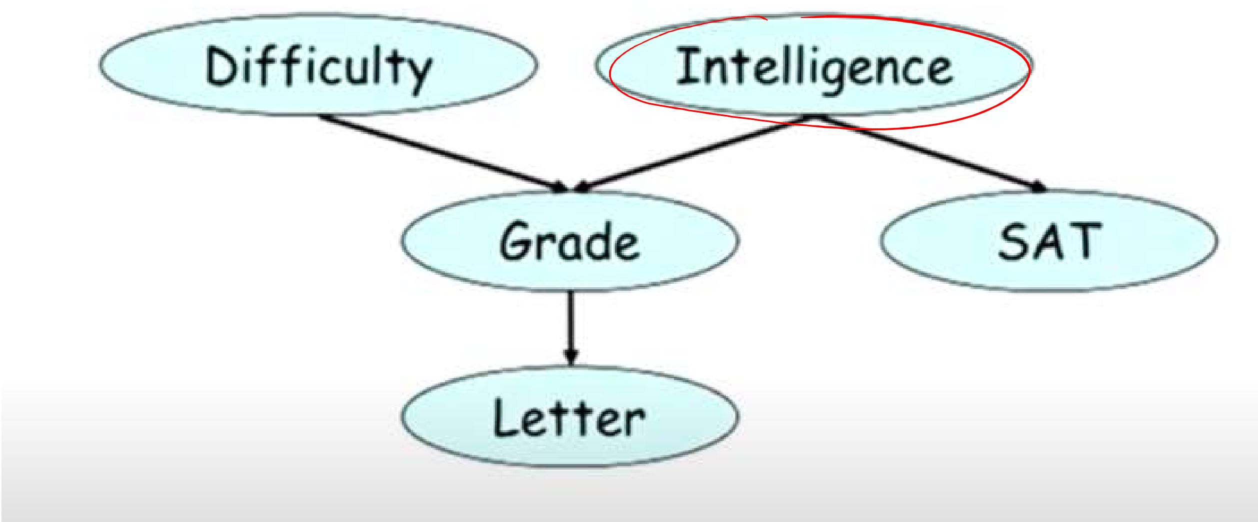
- Student has certain Intelligence (I) level and takes a course of certain Difficulty (D) level



BN Example

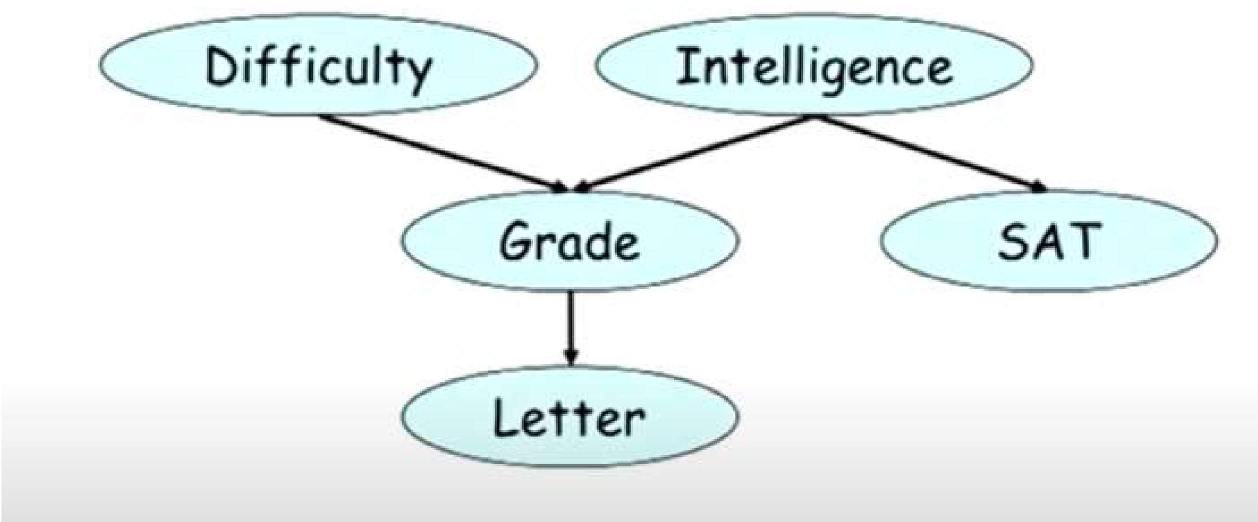
- Grade (G) in course is determined by D and I.

Difficulty
+ Intelligence.



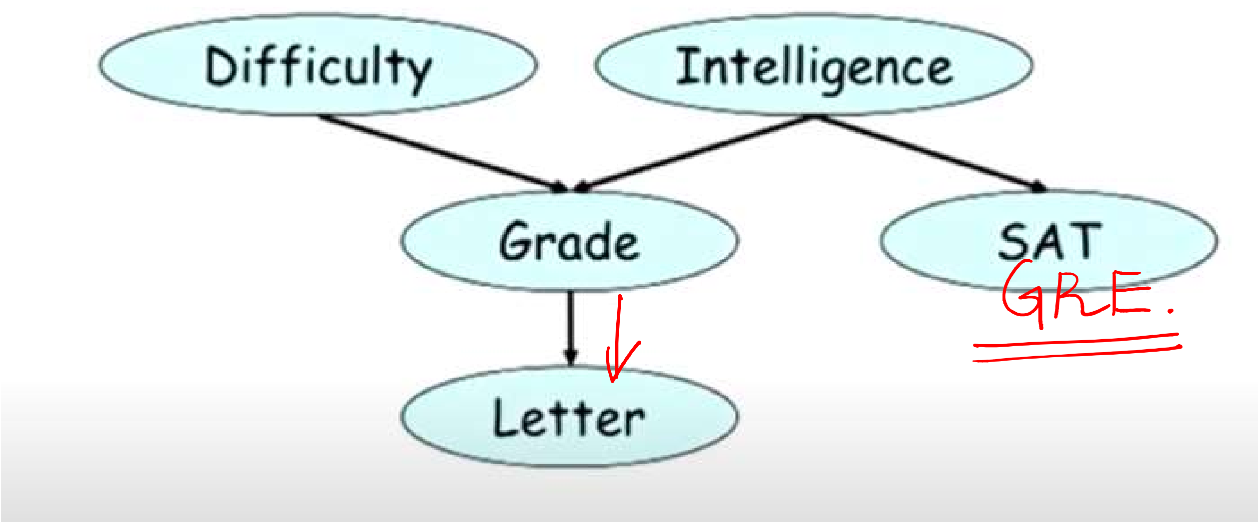
BN Example

- Quality of Letter (L) determined by Grade (G).



BN Example

- SAT score (S) exclusively determined by Intelligence (I).

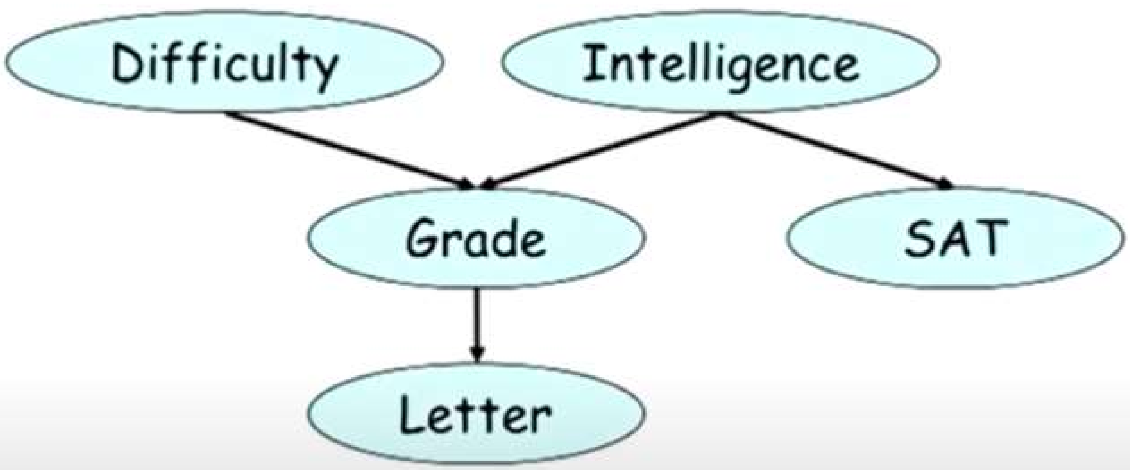


BN Example

- Joint PDF can be evaluated as

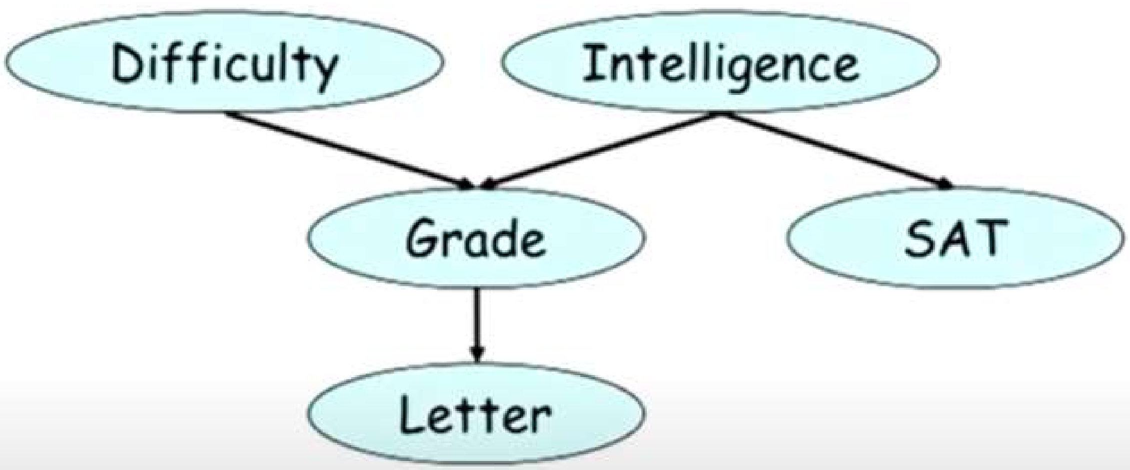
$$p(D, I, G, L, S) = p(D) \times p(I) \times p(G|D,I) \times p(L|G) \times p(S|I).$$

Joint PDF.



BN Example

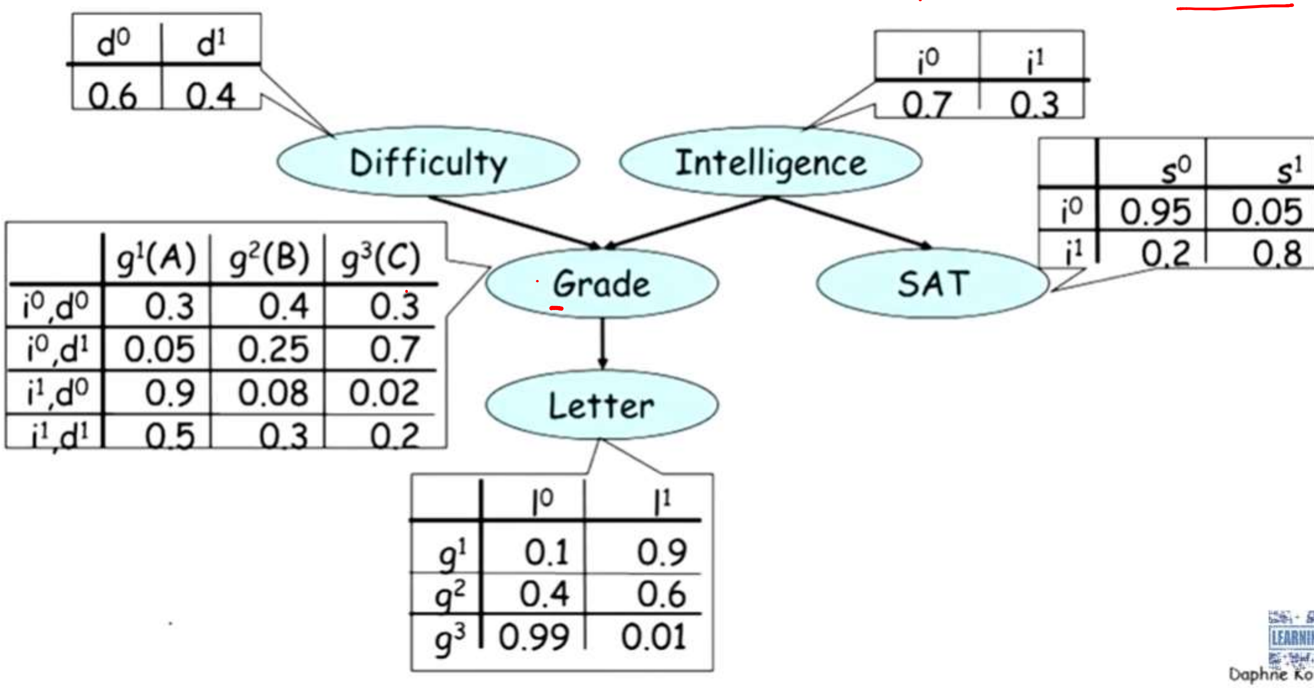
- Joint PDF can be evaluated as
  $$p(D, I, G, L, S) = p(D) \times p(I) \times p(G|D, I) \times p(L|G) \times p(S|I)$$



BN Example

- JPDF can be represented as CPDs

Conditional probability Density



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	$g^1(A)$	$g^2(B)$	$g^3(C)$
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

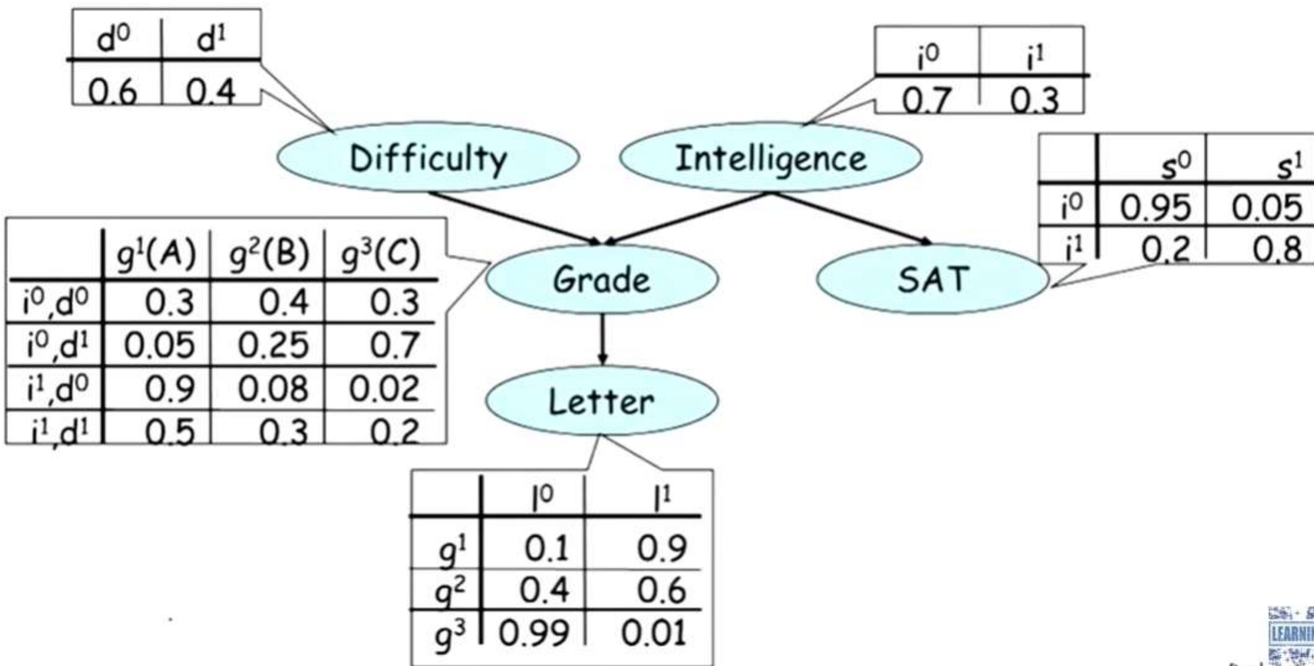
	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Daphne Koller

BN Example

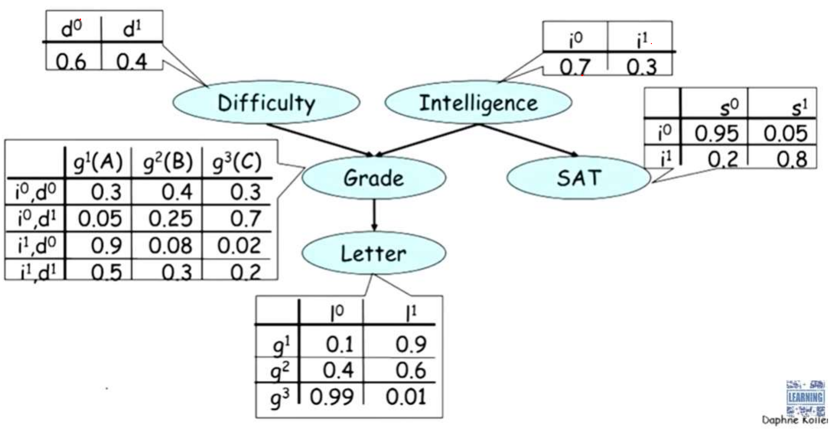
- JPDF can be represented as CPDs (Conditional Probability Distributions)



Daphne Koller

BN Example

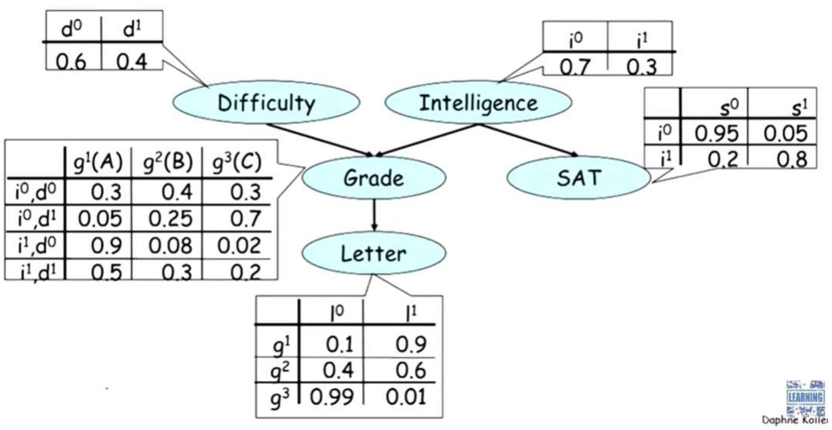
- $i^0 =$ Low level.
- $i^1 =$ High
- $d^0 =$ Easy
- $d^1 =$ Hard/tough



Daphne Koller

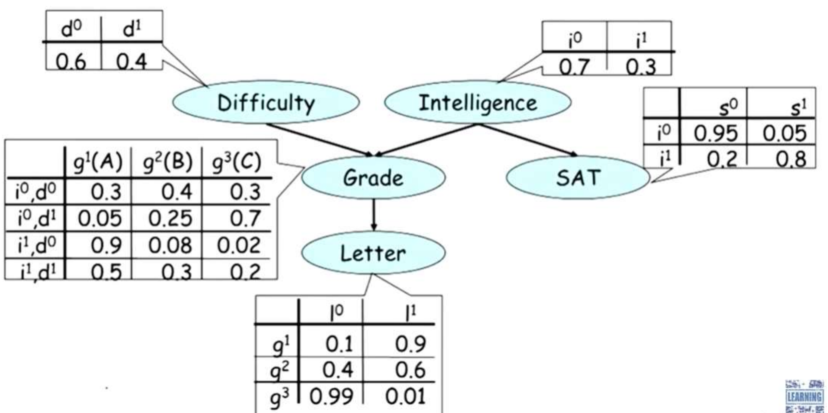
BN Example

- i^0 =Low intelligence
- i^1 =High intelligence
- d^0 =Easy
- d^1 =Difficult



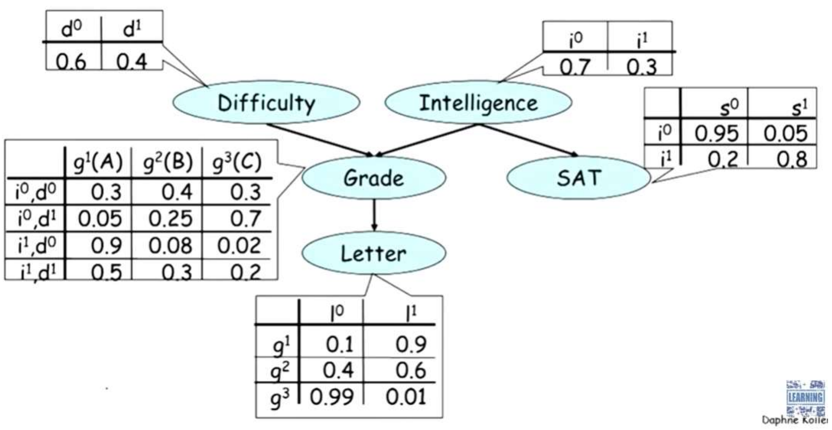
BN Example

- g^1 = A·Grade high
- g^2 = B. Grade Mediocre
- g^3 = C Grade Poor



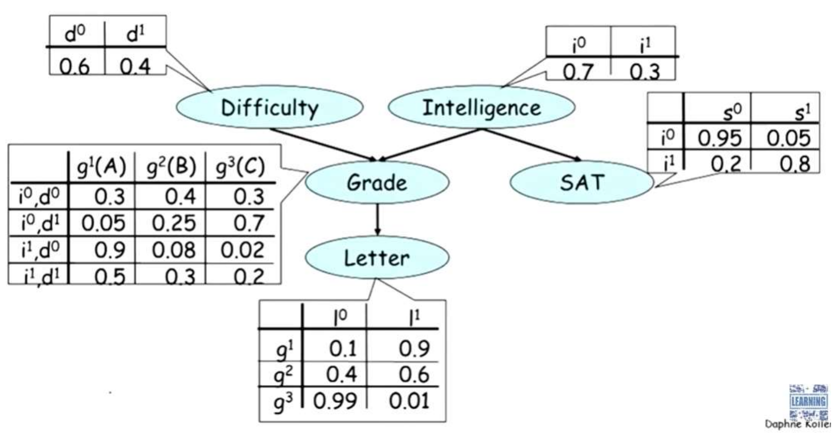
BN Example

- g^1 =A grade
- g^2 =B grade
- g^3 =C grade



BN Example

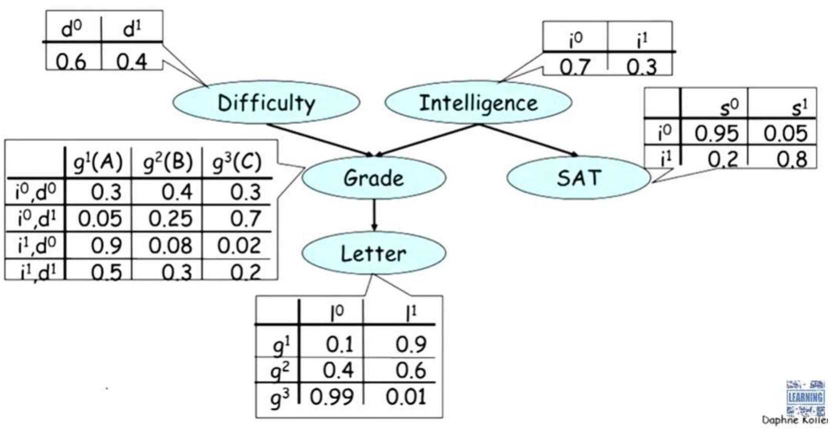
- $l^0 =$ Poor letter
- $l^1 =$ Good.
- $s^0 =$ Poor score
- $s^1 =$ Good score.



Daphne Koller

BN Example

- l^0 =Poor letter
- l^1 =Good letter
- s^0 =Poor SAT score
- s^1 =Good SAT score



BN Example

- Each entry is conditional probability of column element given row element

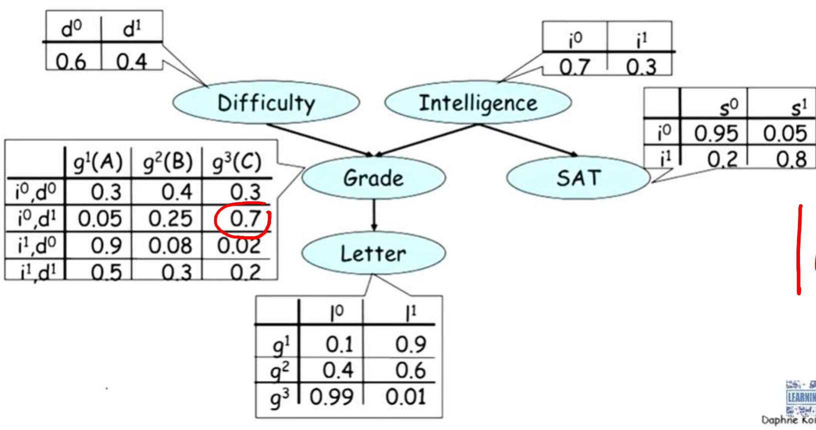
Poor grade

- $p(g^3 | i^0, d^1) =$ 0.7

low

Hard.

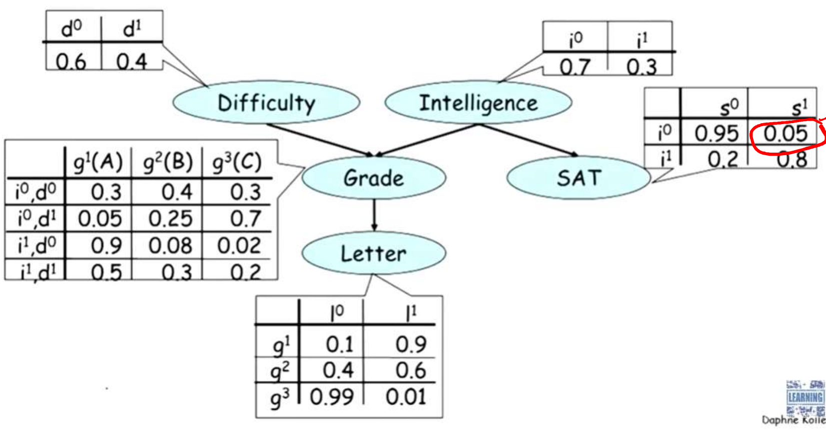
CPD table.



BN Example

- Each entry is conditional probability of column element given row element
- $p(g^3|i^0, d^1) = 0.7$

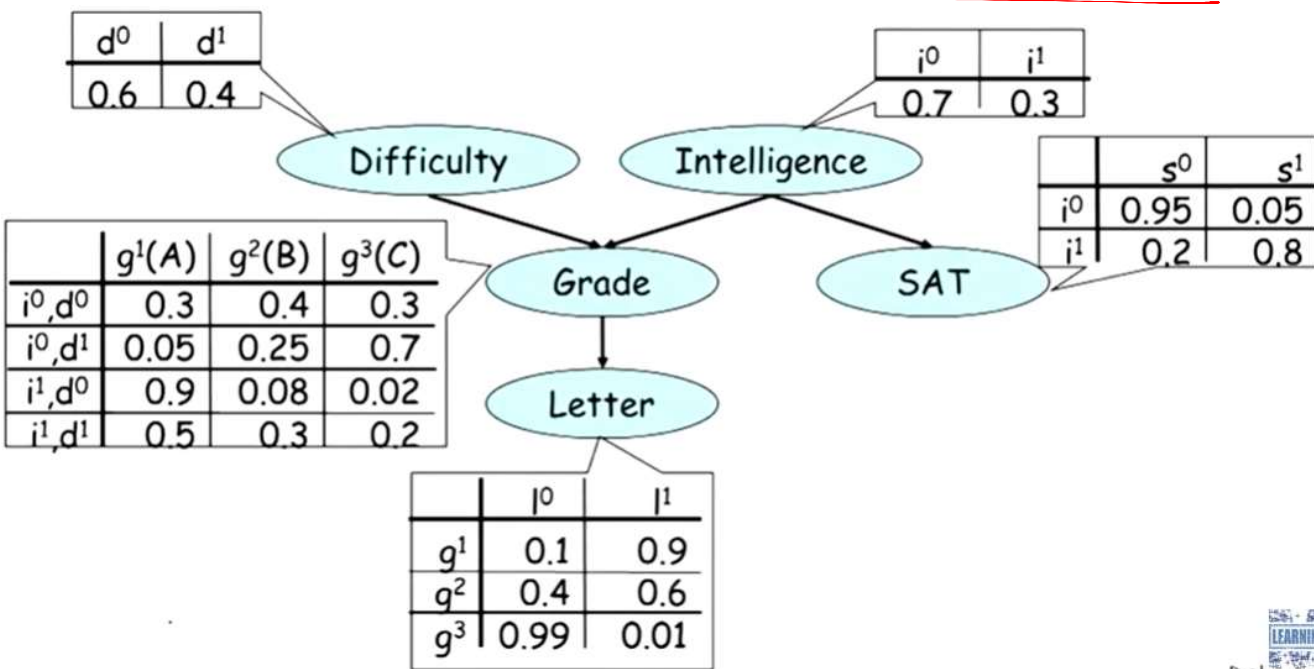
$p(s^1|i^0)$.



BN Example

CPDs enable Very compact Representation

- Enables very compact representation!



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	$g^1(A)$	$g^2(B)$	$g^3(C)$
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

BN Computation

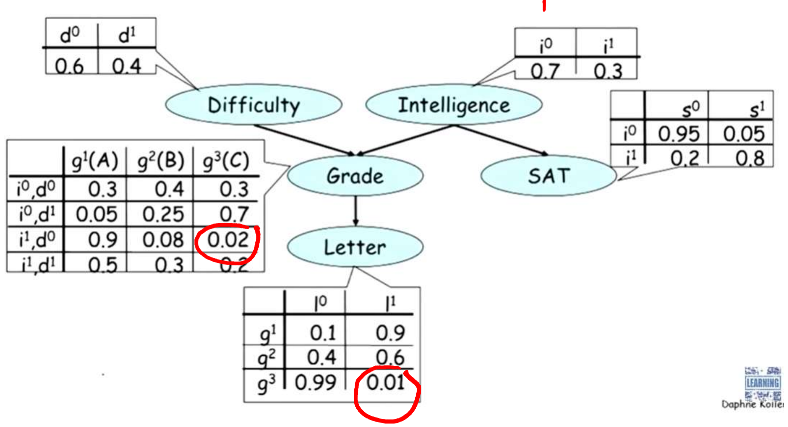
- $p(d^0, i^1, g^3, s^1, l^1)$ can be evaluated as

$$p(d^0) \times p(i^1) \times p(g^3 \mid d^0, i^1) \times p(l^1 \mid g^3) \times p(s^1 \mid i^1)$$

Value of joint PDF.

$$= 0.6 \times 0.3 \times 0.02 \times 0.01 \times 0.8$$

Product of factors.

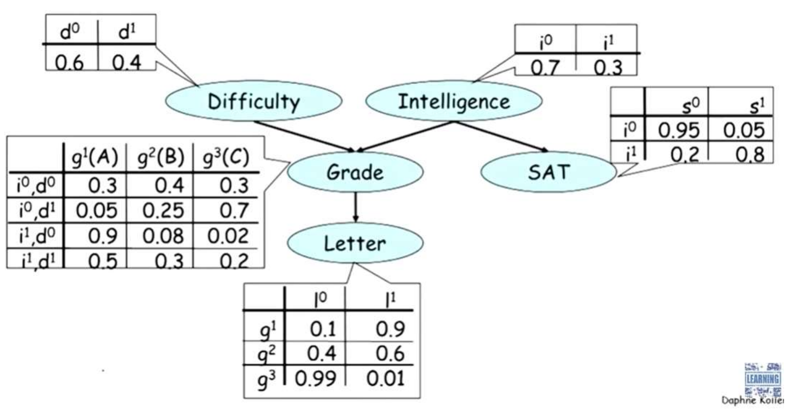


BN Computation

- $p(d^0, i^1, g^3, s^1, l^1)$ can be evaluated as

Evaluated
Easily using CPD tables
at each node in BN.

$$p(d^0) \times p(i^1) \times p(g^3 | d^0, i^1) \times p(s^1 | i^1) \times p(l^1 | g^3)$$
$$= 0.6 \times 0.3 \times 0.02 \times 0.8 \times 0.01$$



BN Inference

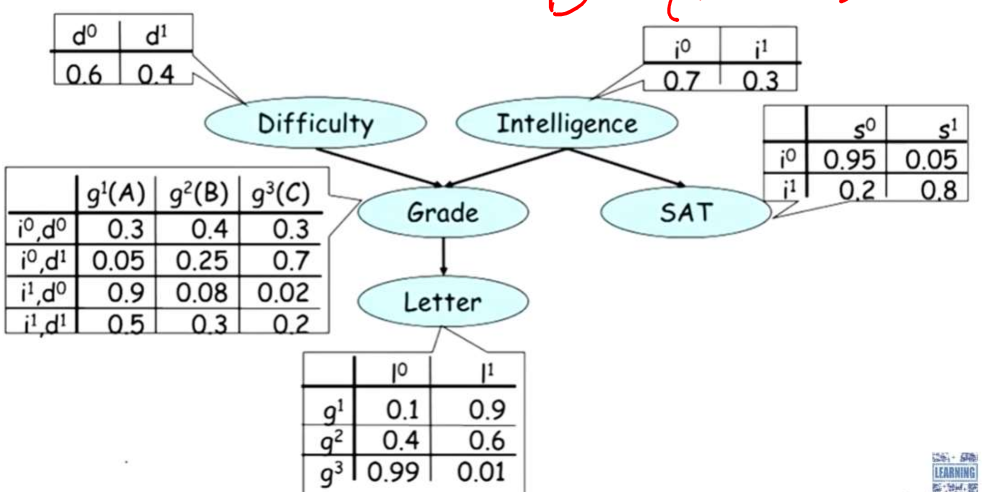
$P(A|B) = \frac{P(A \cap B)}{P(B)}$

- How to evaluate $p(i^1|s^1, l^0)$?

$$p(i^1|s^1, l^0) = \frac{p(i^1, s^1, l^0)}{p(s^1, l^0)}$$

$A = \{i^1\}$

$B = \{s^1, l^0\}$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	$g^1(A)$	$g^2(B)$	$g^3(C)$
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

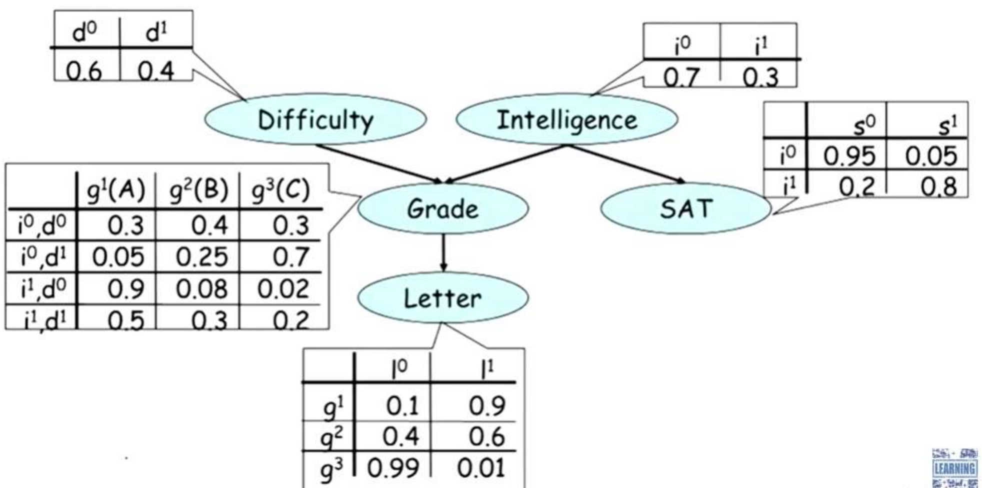
	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Daphne Koller

BN Inference

- How to evaluate $p(i^1|s^1, l^0)$?

$$p(i^1|s^1, l^0) = \frac{p(i^1, s^1, l^0)}{p(s^1, l^0)}$$



BN Inference

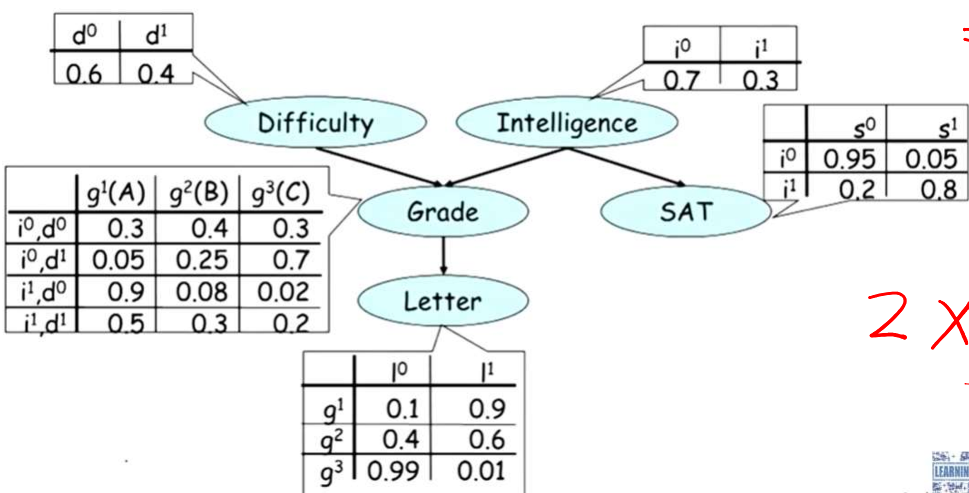
$$p(A) = \sum_j p(A \cap B_j)$$

$$p(i^1, s^1, l^0) = \overbrace{\sum_{D,G} p(D, i^1, G, l^0, s^1)}^{2 \times 3 = 6}.$$

$$p(s^1, l^0) = \sum_{I,D,G} p(D, I, G, l^0, s^1).$$

$$= p(d^0, i^0, g^1, l^0, s^1) + p(d^1, i^0, g^1, l^0, s^1) + p(d^0, i^1, g^1, l^0, s^1) + \dots$$

$$2 \times 2 \times 3 = 12$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	$g^1(A)$	$g^2(B)$	$g^3(C)$
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

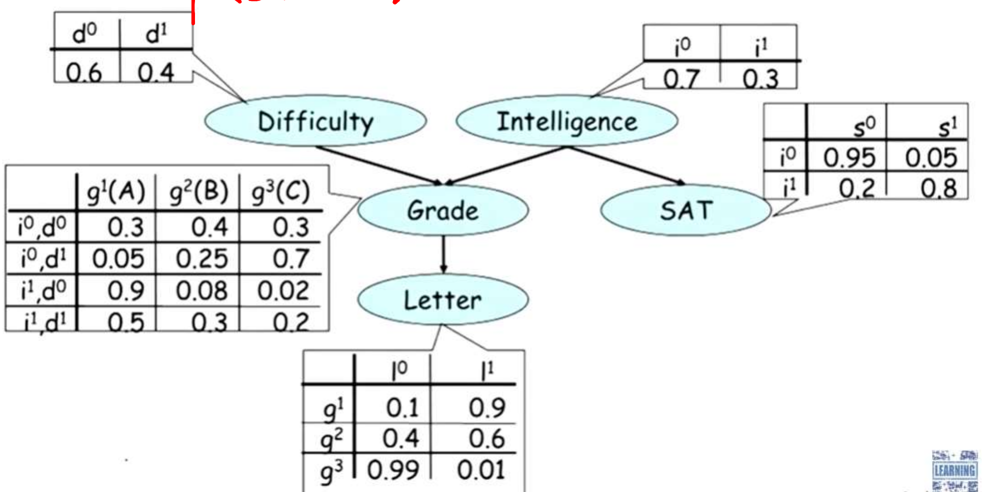
	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

BN Inference

$$p(i^1, s^1, l^0) = \sum_{D,G} p(D, i^1, G, l^0, s^1)$$

$$p(s^1, l^0) = \sum_{I,D,G} p(D, I, G, l^0, s^1)$$

$$p(i^1 | s^1, l^0) = \frac{p(i^1, s^1, l^0)}{p(s^1, l^0)}$$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1(A)	g^2(B)	g^3(C)
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

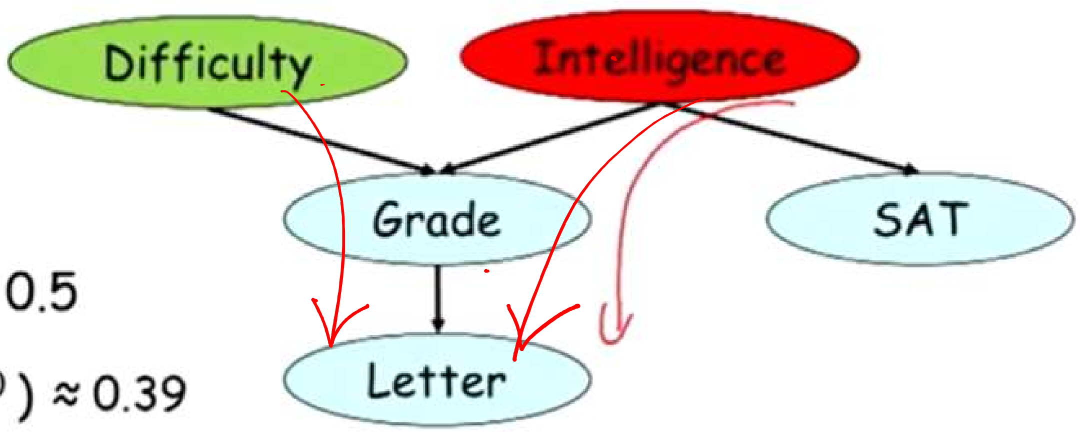
	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

Daphne Koller

Causal reasoning

- Cause Difficulty/ Intelligence explain the evidence Letter.
- Hence, Causal reasoning

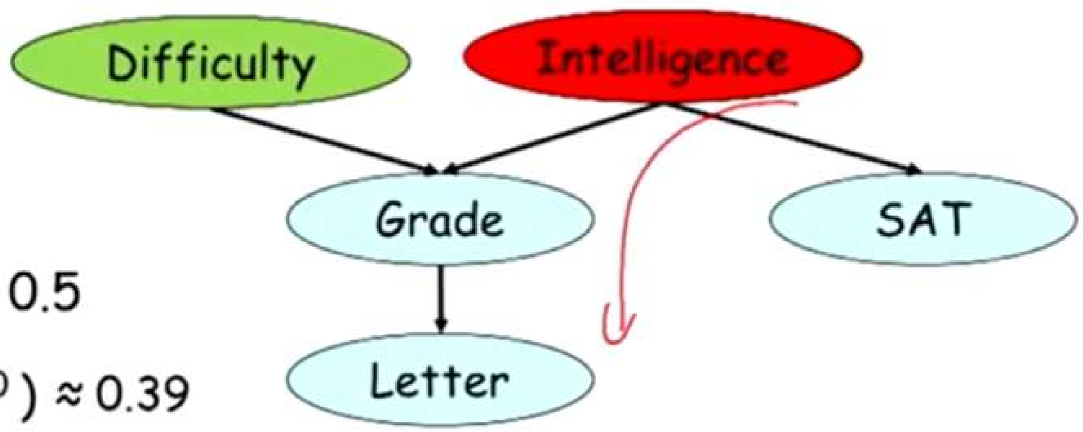


$P(l^1) \approx 0.5$

$P(l^1 \mid i^0) \approx 0.39$

Causal reasoning

- Cause Difficulty/ Intelligence explain the evidence Letter.
- Hence, termed causal reasoning



$P(l^1) \approx 0.5$

$P(l^1 \mid i^0) \approx 0.39$

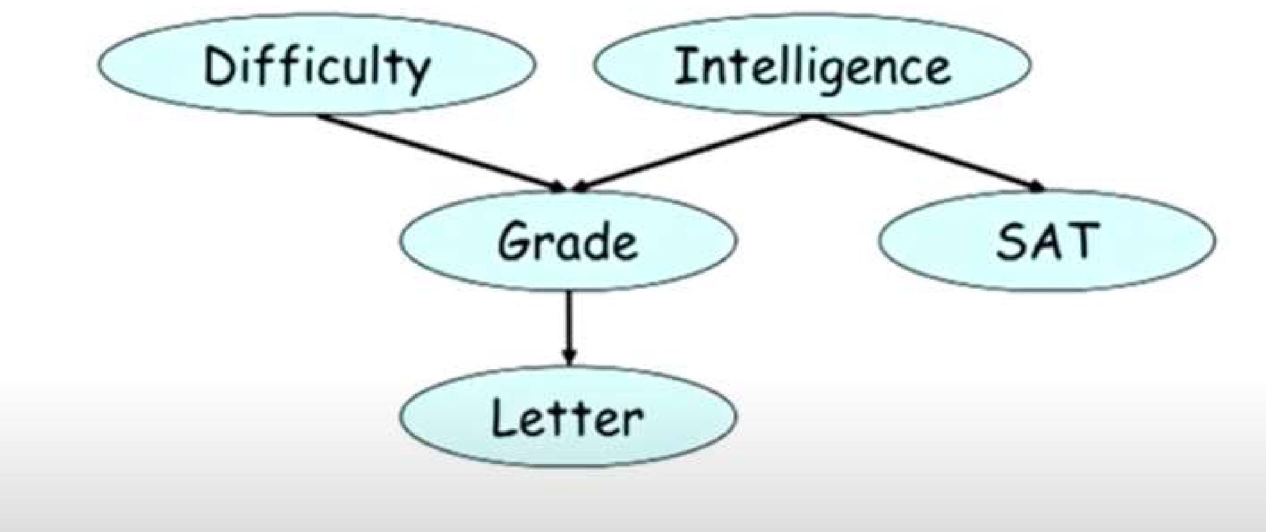
Causal reasoning

- $p(l^1) \approx$ ____0.5____

good letter

Causal reasoning

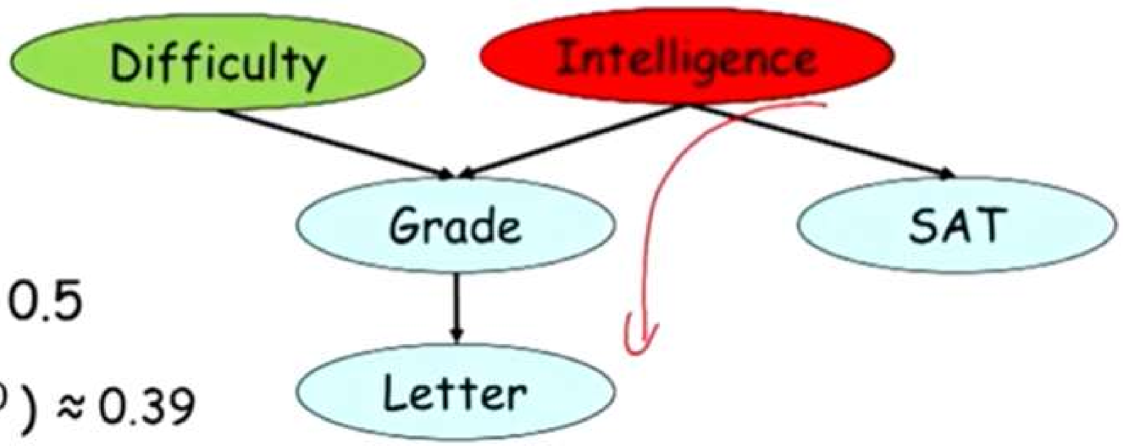
- $p(l^1) \approx 0.5$



Causal reasoning

- $p(l^1) \approx 0.5$
- $p(l^1|i^0) \approx$ ___0.39___

low level intelligence

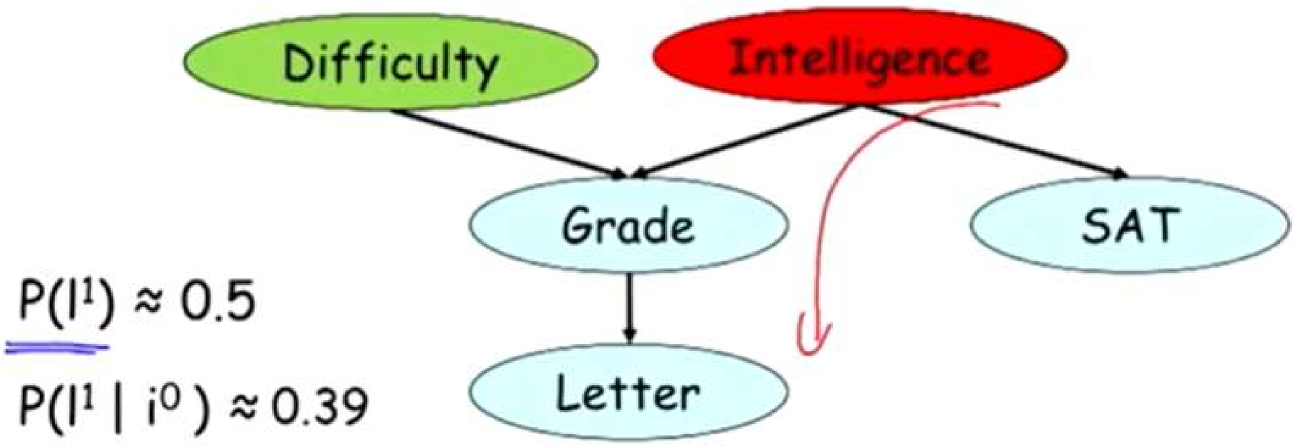


$P(l^1) \approx 0.5$

$P(l^1 \mid i^0) \approx 0.39$

Causal reasoning

- $p(l^1) \approx 0.5$
- $p(l^1 | i^0) \approx 0.39$

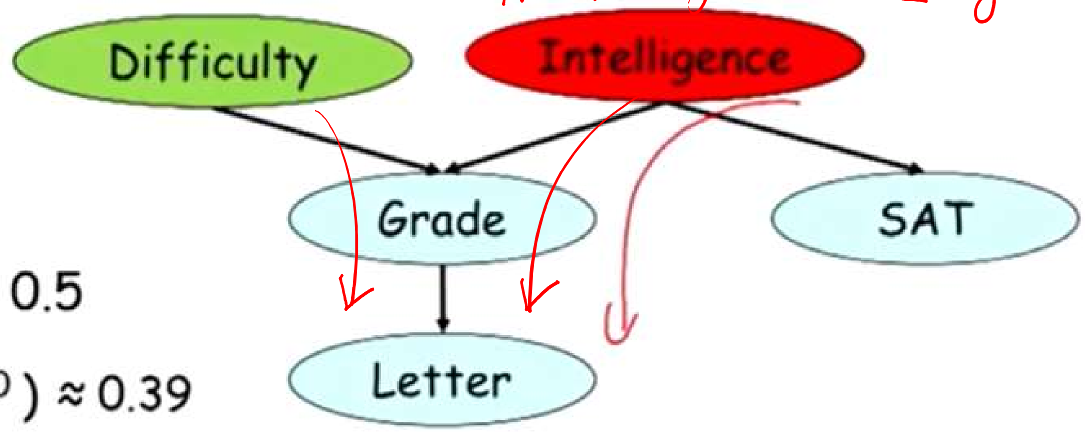


Causal reasoning

- $p(l^1) \approx 0.5$
- $p(l^1|i^0) \approx 0.39$
- $p(l^1|i^0, d^0) \approx$ 0.51.

low intelligence

Easy course

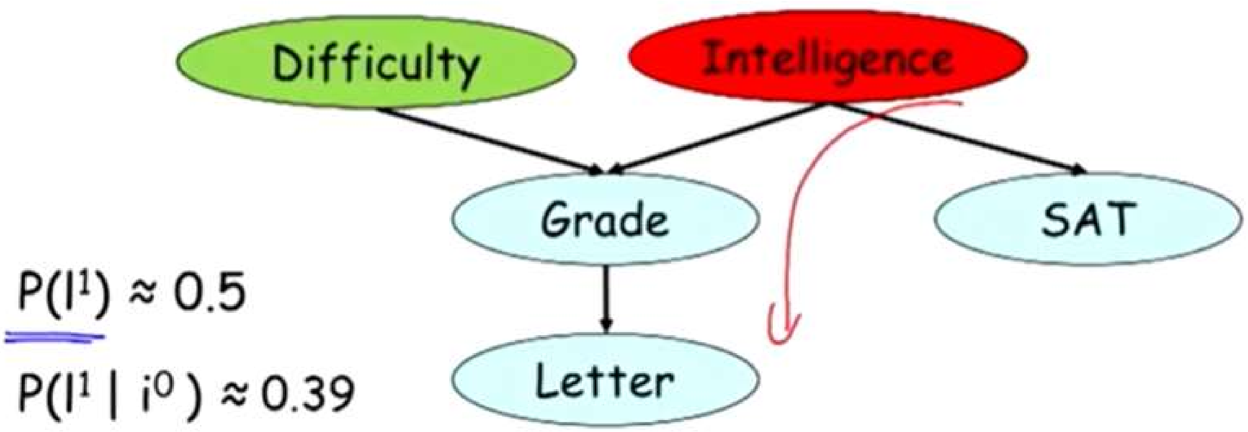


$P(l^1) \approx 0.5$

$P(l^1 \mid i^0) \approx 0.39$

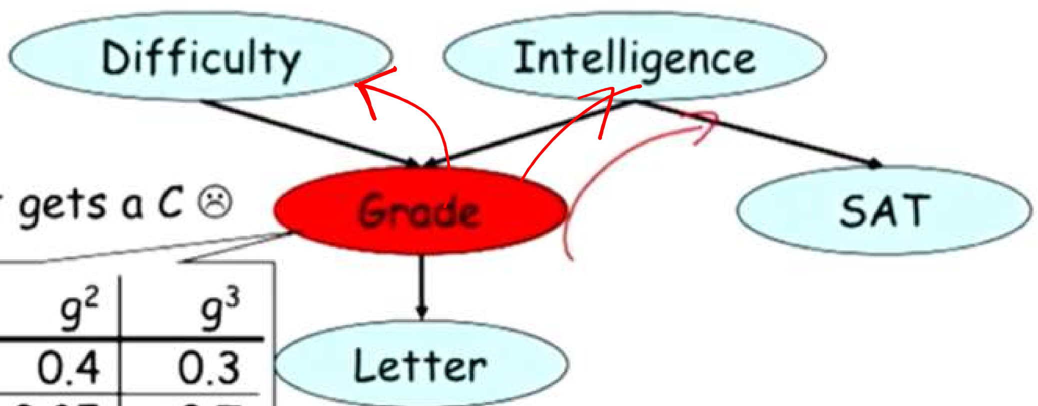
Causal reasoning

- $p(l^1) \approx 0.5$
- $p(l^1|i^0) \approx 0.39$
- $p(l^1|i^0, d^0) \approx 0.51$



Evidential Reasoning

- Evidence Grade explains the causes Difficulty/ Intelligence Letter.
- Hence, Evidential Reasoning.

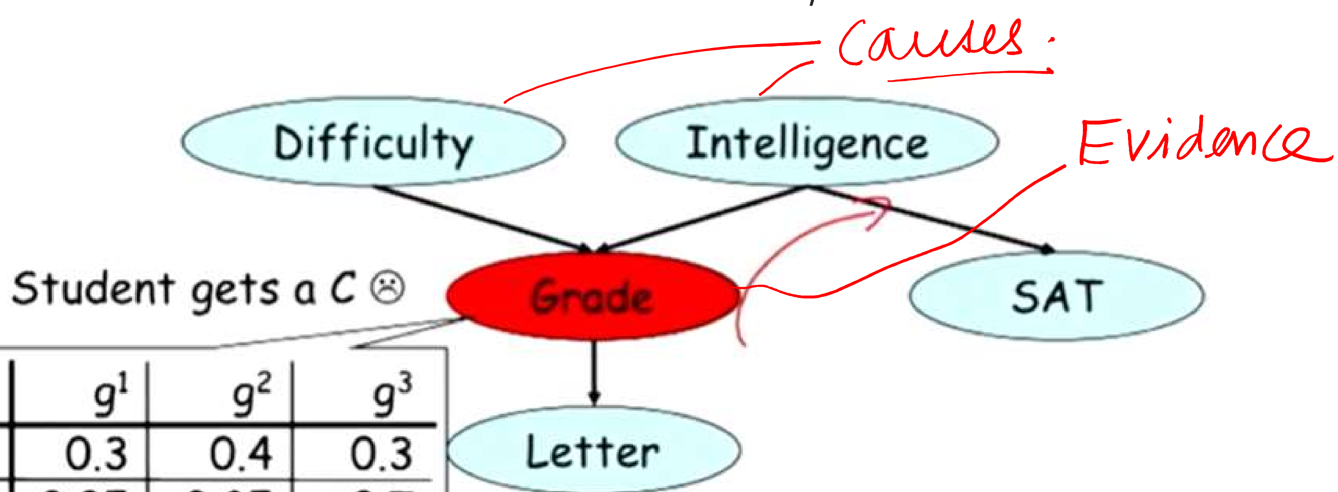


Student gets a C ☹

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

Evidential Reasoning

- Evidence Grade explains the causes Difficulty/ Intelligence Letter.
- Hence, Evidential Reasoning.



	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

Evidential Reasoning

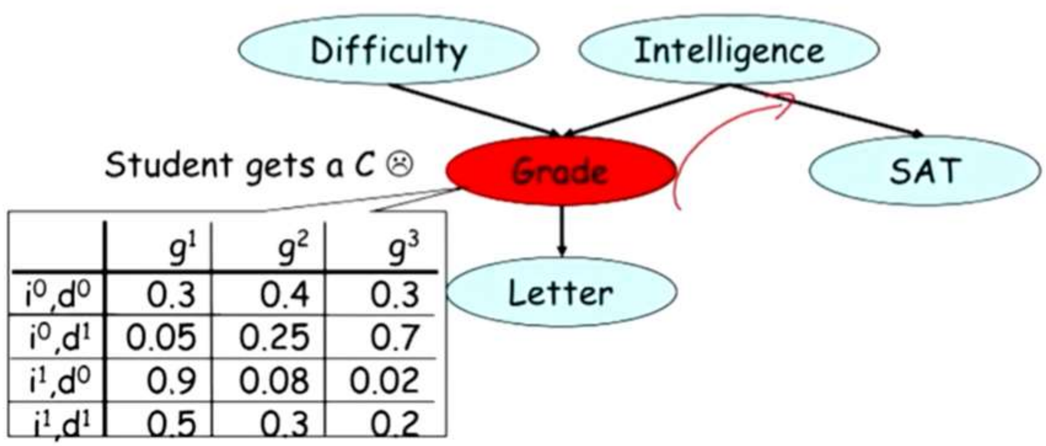
- $p(d^1) = $ 0.4
- $p(i^1) = $ 0.3

$P(d^1) = 0.4$

$P(d^1 \mid g^3) \approx$

$P(i^1) = 0.3$

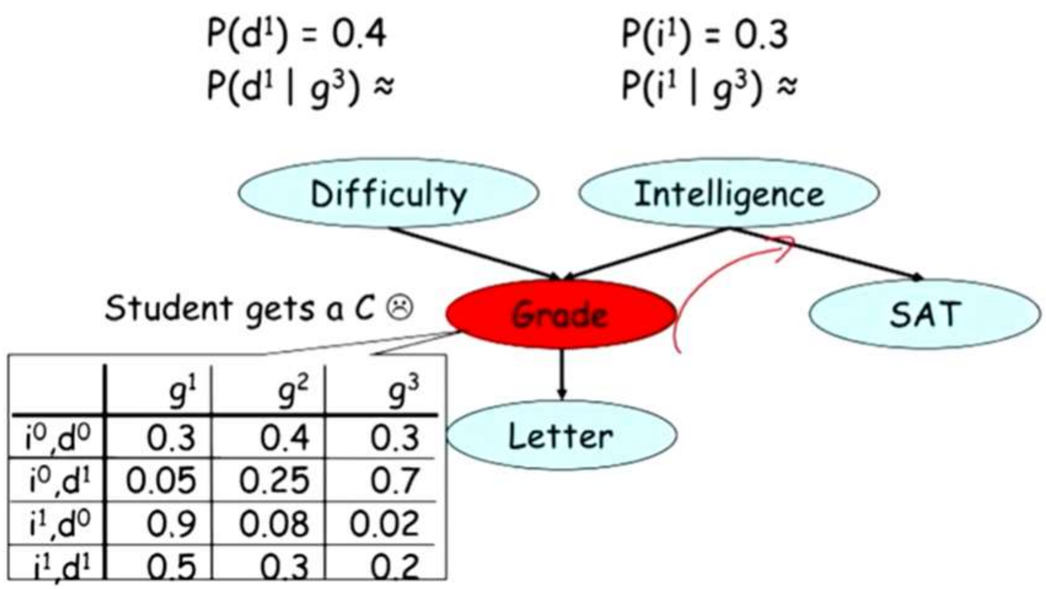
$P(i^1 \mid g^3) \approx$



	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

Evidential Reasoning

- $p(d^1) = 0.4$
- $p(i^1) = 0.3$



	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

Evidential Reasoning

$g^3 (C)$: Poor grade!

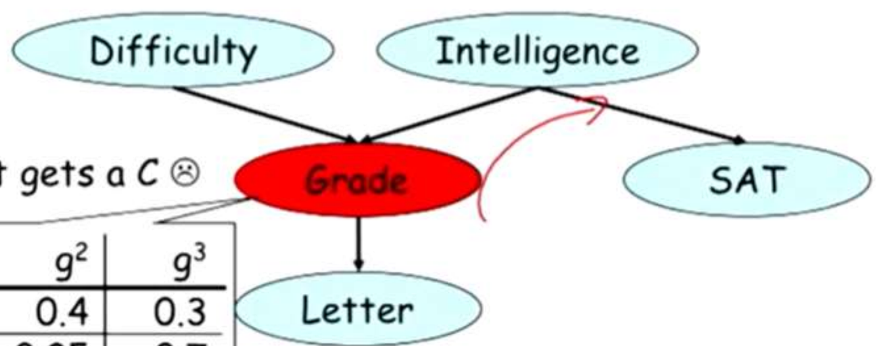
- $p(d^1) = 0.4, p(d^1|g^3) = \underline{\ 0.63\ }$
- $p(i^1) = 0.3, p(i^1|g^3) = \underline{\ 0.08\ }$

Prior probabilities

Aposteriori probability.

$P(d^1) = 0.4$
$P(d^1 \mid g^3) \approx 0.63$

$P(i^1) = 0.3$
$P(i^1 \mid g^3) \approx 0.08$

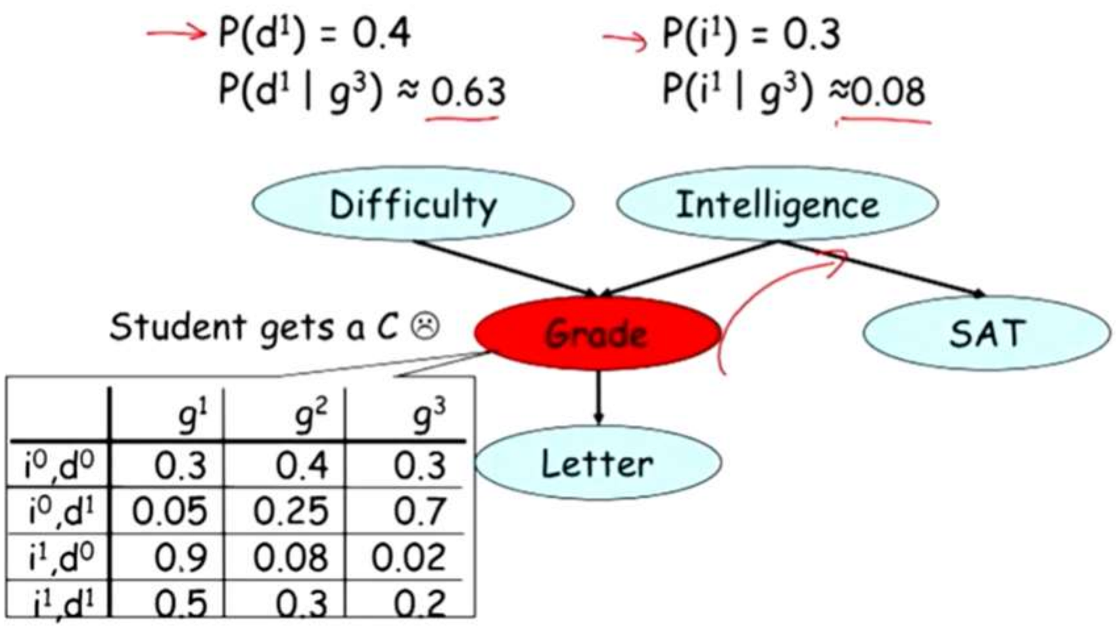


Student gets a C ☹

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

Evidential Reasoning

- $p(d^1) = 0.4, p(d^1|g^3) = 0.63$
- $p(i^1) = 0.3, p(i^1|g^3) = 0.08$

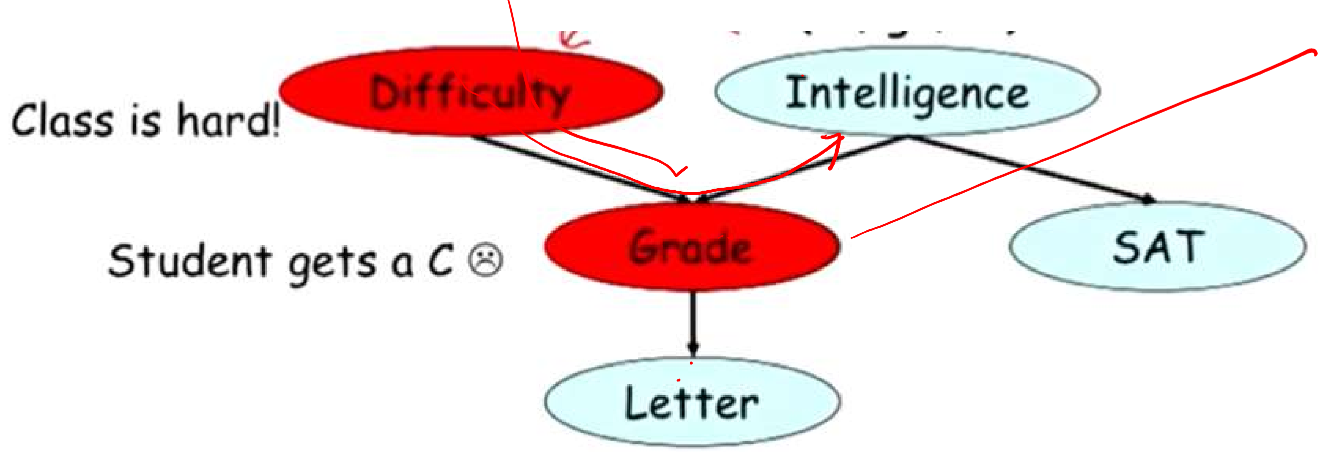


	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

Intercausal Reasoning

- One cause explains other cause.
- Difficulty explains intelligence.
- Example: student gets g^3 = C grade

V Shaped Structure.



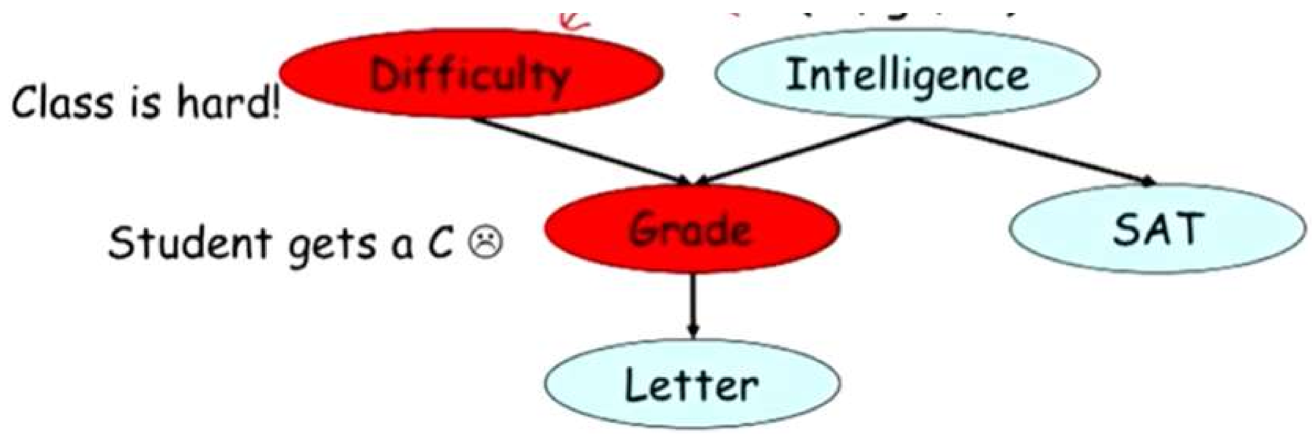
Activate V structure.

Intercausal Reasoning

- $p(i^1|g^3) =$ ___0.08___

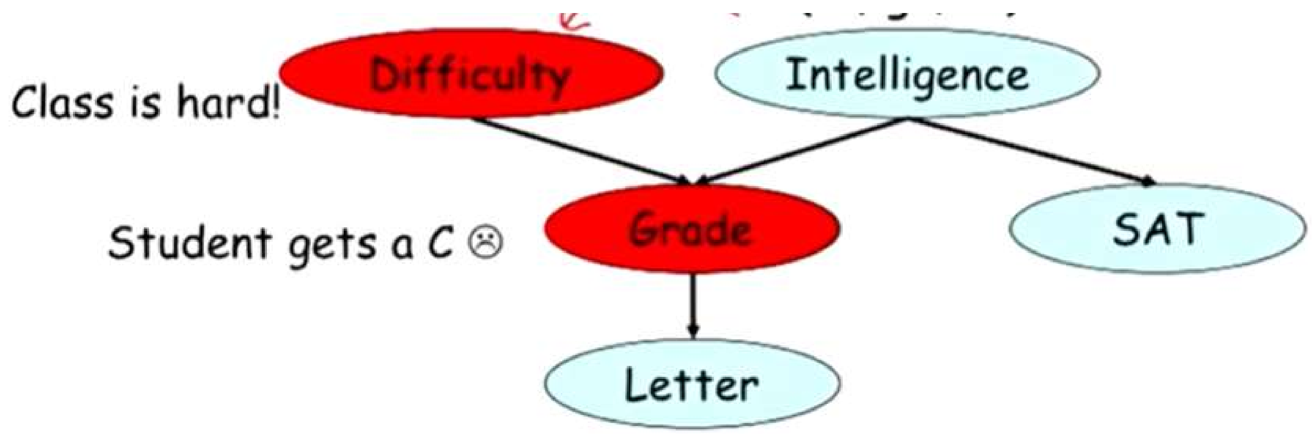
high IQ

C grade



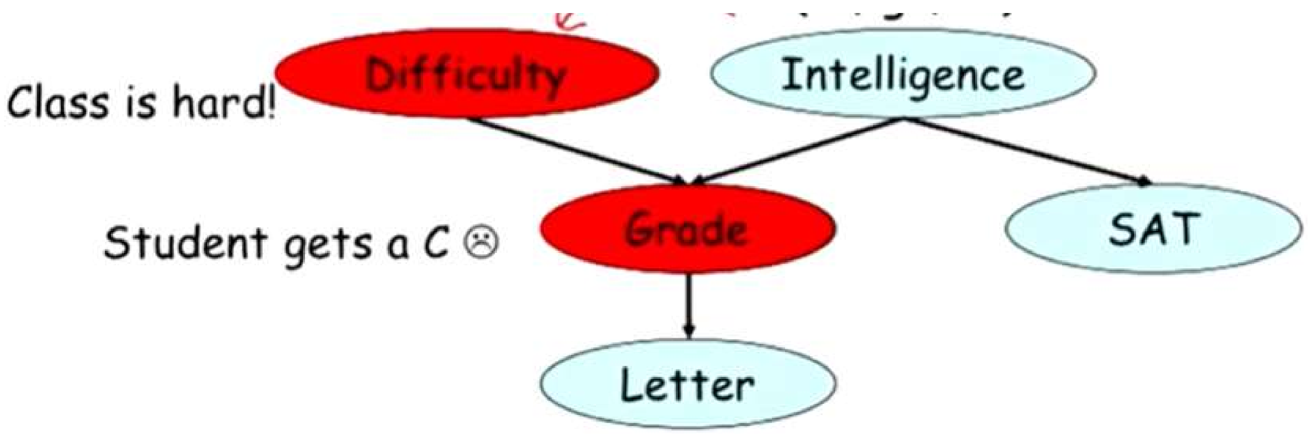
Intercausal Reasoning

- $p(i^1|g^3) = 0.08$



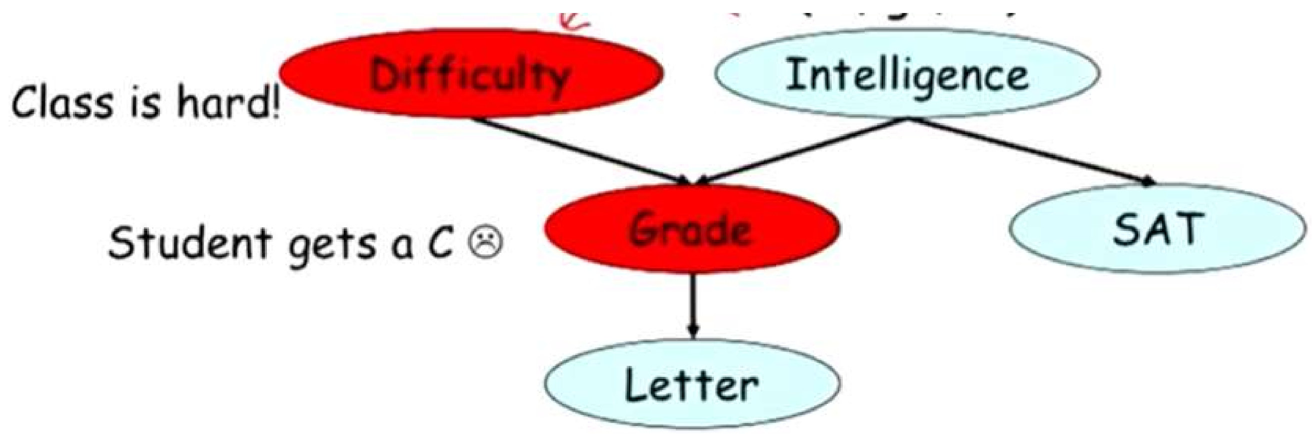
Intercausal Reasoning

- $p(i^1|g^3) = 0.08$
- $p(i^1|g^3, d^1) \approx$ 0.11



Intercausal Reasoning

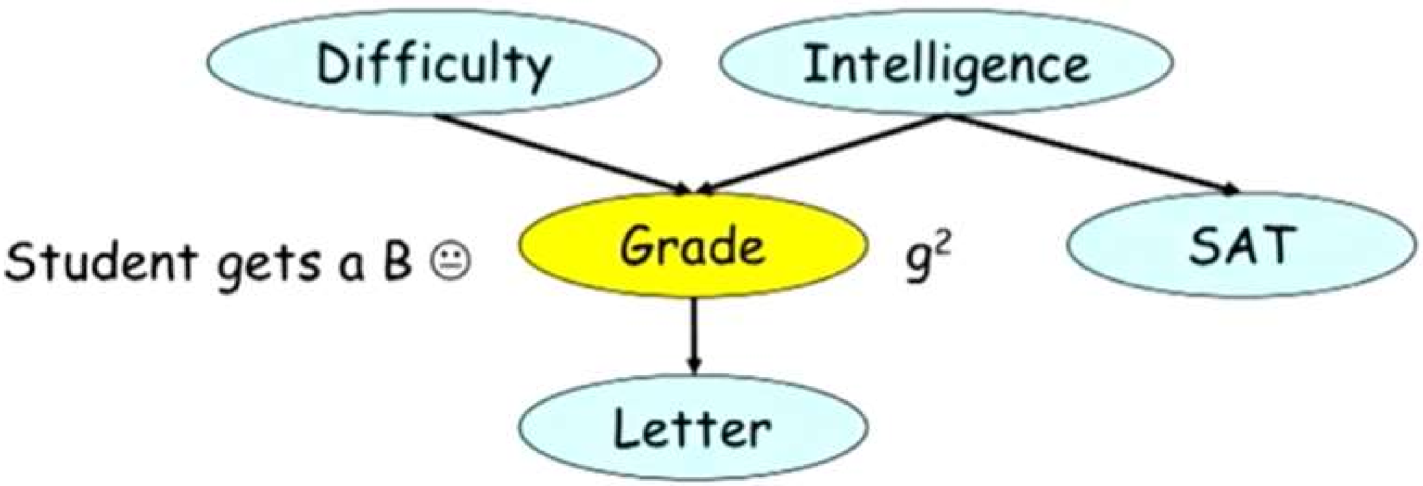
- $p(i^1|g^3) = 0.08$
- $p(i^1|g^3, d^1) \approx 0.11$



Intercausal Reasoning

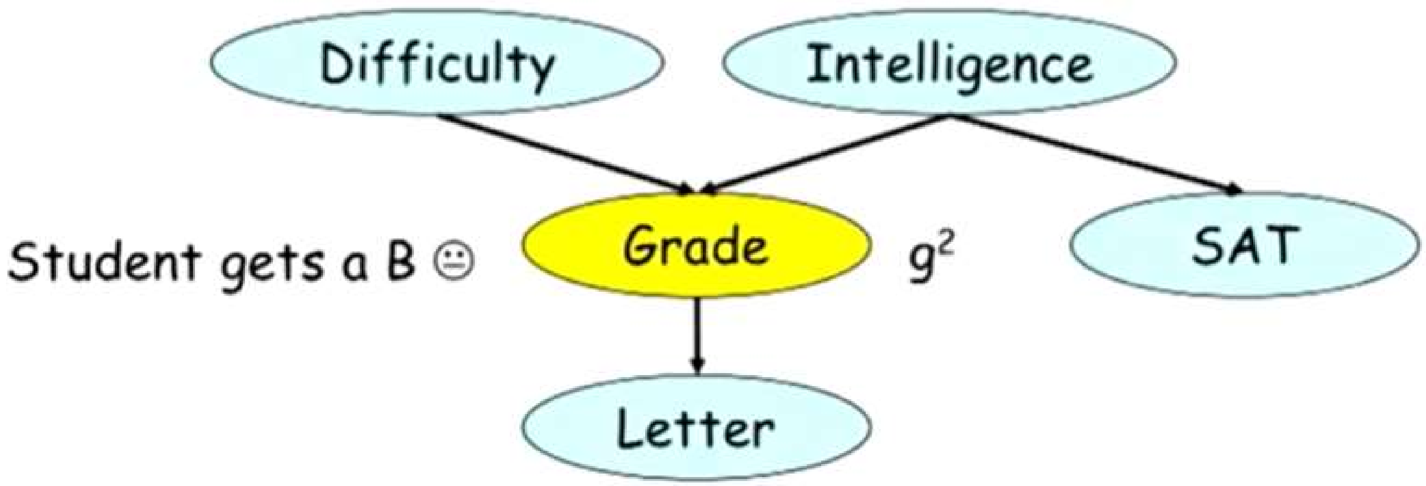
- Student gets B grade
- $p(i^1|g^2) = $ ___0.175___

B Grade



Intercausal Reasoning

- Student gets B grade
- $p(i^1|g^2) = 0.175$

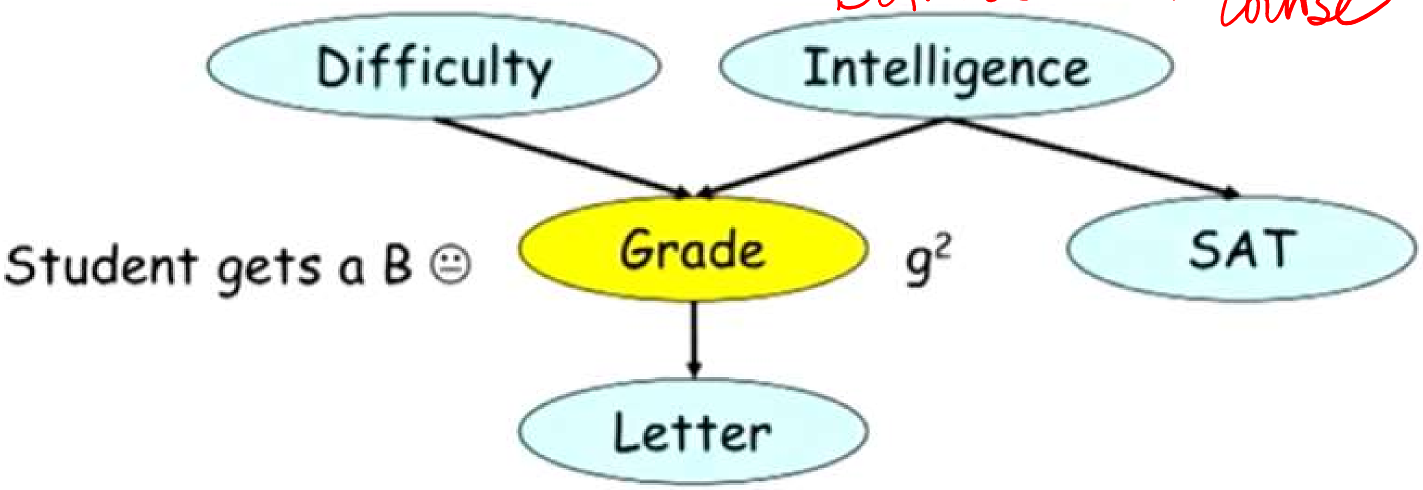


Intercausal Reasoning

- $p(i^1 | g^2) = 0.175$
- $p(i^1 | g^2, d^1) \approx$ 0.34

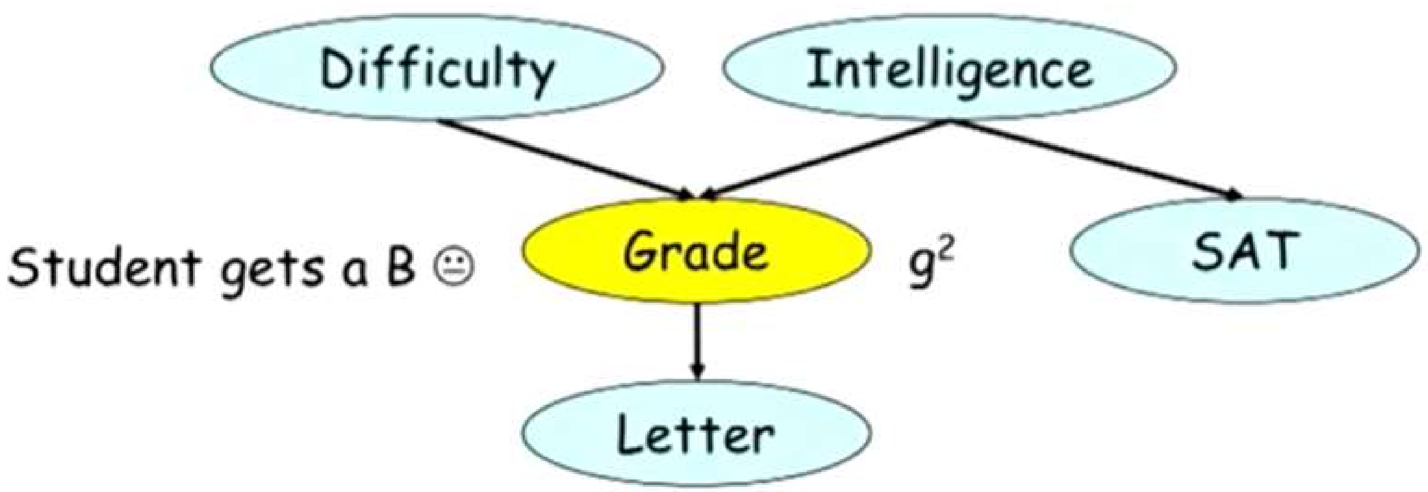
B Grade

Difficult course



Intercausal Reasoning

- $p(i^1|g^2) = 0.175$
- $p(i^1|g^2, d^1) \approx 0.34$



Bayesian networks

- Bayesian networks are powerful tools for Reasoning and Inference

Causal

Evidential

intercausal Reasoning

Bayesian networks

- Bayesian networks are powerful tools for reasoning and inference

Bayesian networks

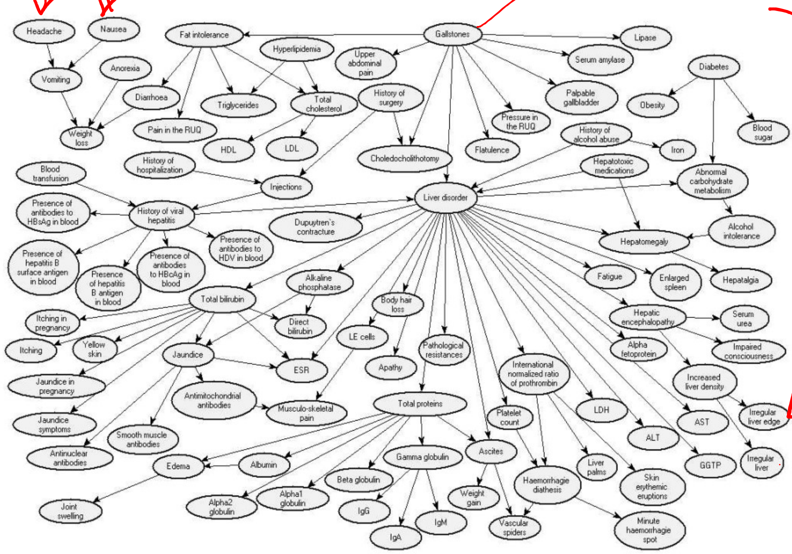
- BNs are used in several applications such as _Medical Diagnosis_
- Example below: BN for diagnosis of liver disorders.

Headache

Nausea

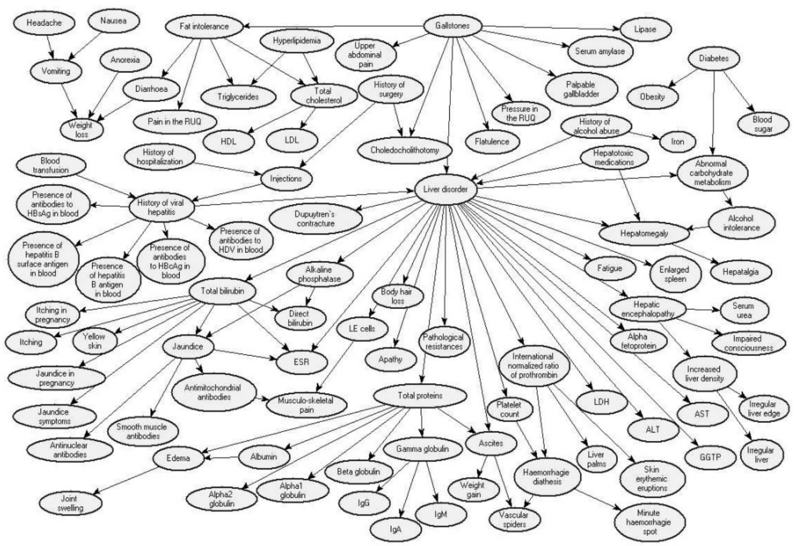
Gallstones.

BN for liver disorders.



Bayesian networks

- BNs are used in several applications such as Medical Diagnosis
- Example below: BN for diagnosis of liver disorders.



Instructors may use this white area (14.5 cm / 25.4 cm) for the text. Three options provided below for the font size.

Font: Avenir (Book), Size: 32, Colour: Dark Grey

Font: Avenir (Book), Size: 28, Colour: Dark Grey

Font: Avenir (Book), Size: 24, Colour: Dark Grey

Do not use the space below.