Lecture 21: 18 April, 2022

Madhavan Mukund

https://www.cmi.ac.in/~madhavan

Data Mining and Machine Learning
January–May 2022

Conditional probabilities

- Boolean variables $x_1, x_2, \ldots, x_n$

Conditional probabilities

- Boolean variables $x_1, x_2, \ldots, x_n$
- Joint probabilities $P(v_1, v_2, \ldots, v_n)$
  - 2^n combinations of $x_1, x_2, \ldots, x_n$
  - $2^n - 1$ parameters

Conditional probabilities

- Boolean variables $x_1, x_2, \ldots, x_n$
- Joint probabilities $P(v_1, v_2, \ldots, v_n)$
  - 2^n combinations of $x_1, x_2, \ldots, x_n$
  - $2^n - 1$ parameters
- Naïve Bayes assumption — complete independence
  - $P(x_i = 1)$ for each x_i
  - n parameters

Conditional probabilities

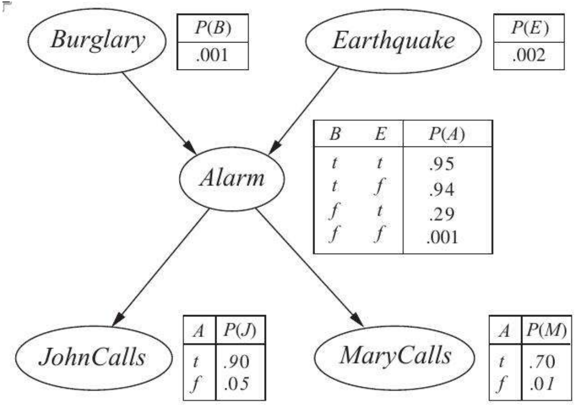
- Boolean variables $x_1, x_2, \ldots, x_n$
- Joint probabilities $P(v_1, v_2, \ldots, v_n)$
  - 2^n combinations of $x_1, x_2, \ldots, x_n$
  - $2^n - 1$ parameters
- Naïve Bayes assumption — complete independence
  - $P(x_i = 1)$ for each x_i
  - n parameters
- Can we strive for something in between?
  - "Local" dependencies between some variables

Probabilistic graphical models

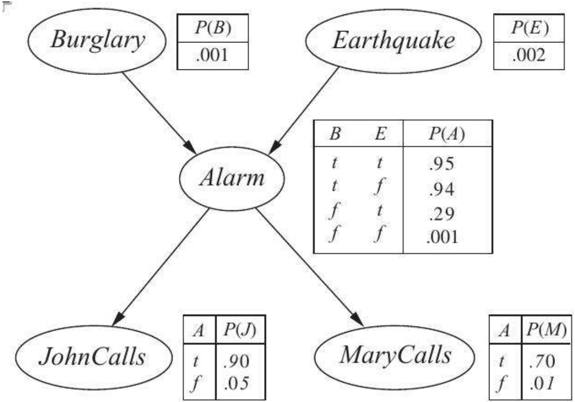
- Judea Pearl [Turing Award 2011]
- Represent local dependencies using directed graph

Probabilistic graphical models

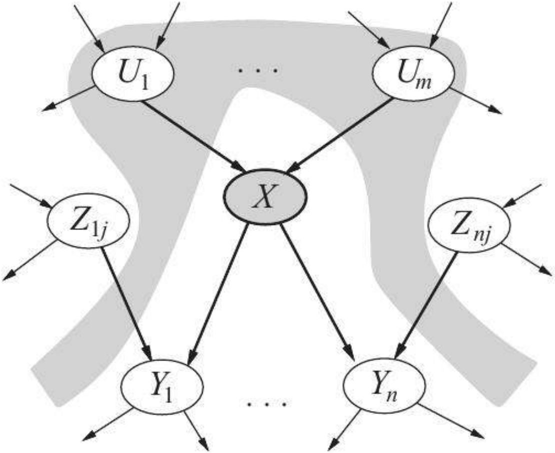
- Judea Pearl [Turing Award 2011]
- Represent local dependencies using directed graph
- Example: Burglar alarm
  - Pearl's house has a burglar alarm
  - Neighbours John and Mary call if they hear the alarm
  - John is prone to mistaking ambulances etc for the alarm
  - Mary listens to loud music and sometimes fails to hear the alarm
  - The alarm may also be triggered by an earthquake (California!)



- Each node has a local (conditional) probability table

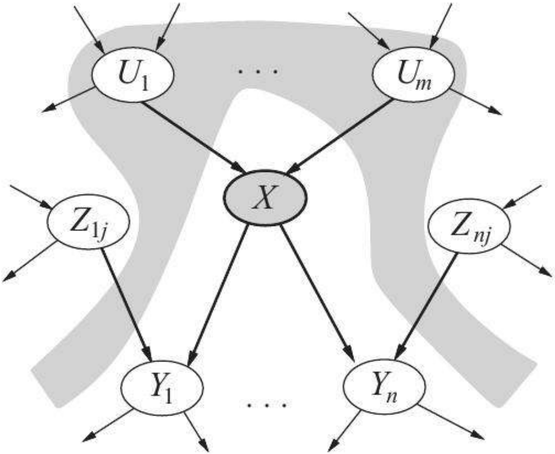


- Each node has a local (conditional) probability table
- Fundamental assumption:
  A node is conditionally independent of non-descendants, given its parents



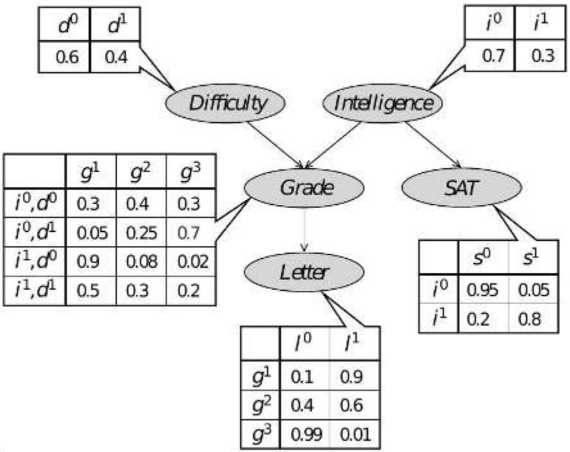
Probabilistic graphical models

- Each node has a local (conditional) probability table
- Fundamental assumption:
  A node is conditionally independent of non-descendants, given its parents
- Graph is a DAG, no cyclic dependencies



Student example

- Example due to Nir Friedman and Daphne Koller
- Student asks teacher for a reference letter
- Teacher has forgotten the student, so letter is entirely based on student's grade in the course



d^0	d^1
0.6	0.4

i^0	i^1
0.7	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

	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

Evaluating a network

- John and Mary call Pearl. What is the probability that there has been a burglary?

Evaluating a network

- John and Mary call Pearl. What is the probability that there has been a burglary?
- $P(b, m, j)$, where b: burglary, j: John calls, m: Mary calls

Evaluating a network

- John and Mary call Pearl. What is the probability that there has been a burglary?
- $P(b, m, j)$, where b: burglary, j: John calls, m: Mary calls
- $P(b, m, j) = \sum_{a=0}^{1} \sum_{e=0}^{1} P(b, j, m, a, e)$, where a: alarm rings, e: earthquake

Evaluating a network

- John and Mary call Pearl. What is the probability that there has been a burglary?
- $P(b, m, j)$, where b: burglary, j: John calls, m: Mary calls
- $P(b, m, j) = \sum_{a=0}^{1} \sum_{e=0}^{1} P(b, j, m, a, e)$, where a: alarm rings, e: earthquake
- Bayes Rule: $P(A, B) = P(A \mid B)P(B)$

Evaluating a network

- John and Mary call Pearl. What is the probability that there has been a burglary?
- $P(b, m, j)$, where b: burglary, j: John calls, m: Mary calls
- $P(b, m, j) = \sum_{a=0}^{1} \sum_{e=0}^{1} P(b, j, m, a, e)$, where a: alarm rings, e: earthquake
- Bayes Rule: $P(A, B) = P(A \mid B)P(B)$
- $P(x_1, x_2, \ldots, x_n) = P(x_1 \mid x_2, \ldots, x_n)P(x_2, x_3, \ldots, x_n)$

A B

$P(x_2 \mid x_3 - x_n) \cdot P(x_3 - x_n)$

$P(x_3 \mid x_4 - x_n) P(x_4 - \ldots$

Evaluating a network

- John and Mary call Pearl. What is the probability that there has been a burglary?
- $P(b, m, j)$, where b: burglary, j: John calls, m: Mary calls
- $P(b, m, j) = \sum_{a=0}^{1} \sum_{e=0}^{1} P(b, j, m, a, e)$, where a: alarm rings, e: earthquake
- Bayes Rule: $P(A, B) = P(A \mid B)P(B)$
- $P(x_1, x_2, \ldots, x_n) = P(x_1 \mid x_2, \ldots, x_n)P(x_2, x_3, \ldots, x_n)$
- Applied recursively, this gives us the chain rule
  $P(x_1, x_2, \ldots, x_n) = P(x_1 \mid x_2, \ldots, x_n)P(x_2 \mid x_3, \ldots, x_n) \cdots P(x_{n-1} \mid x_n)P(x_n)$

Evaluating a network

- $P(x_1, x_2, \ldots, x_n) = P(x_1 \mid x_2, \ldots, x_n) P(x_2 \mid x_3, \ldots, x_n) \cdots P(x_{n-1} \mid x_n) P(x_n)$

Evaluating a network

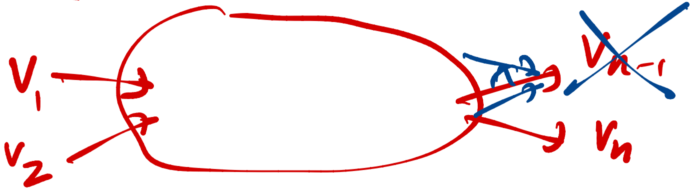
- $P(x_1, x_2, \ldots, x_n) = P(x_1 \mid x_2, \ldots, x_n) P(x_2 \mid x_3, \ldots, x_n) \cdots P(x_{n-1} \mid x_n) P(x_n)$
- Can choose any ordering of $x_1, x_2, \ldots, x_n$

Evaluating a network

- $P(x_1, x_2, \ldots, x_n) = P(x_1 \mid x_2, \ldots, x_n) P(x_2 \mid x_3, \ldots, x_n) \cdots P(x_{n-1} \mid x_n) P(x_n)$
- Can choose any ordering of $x_1, x_2, \ldots, x_n$
- Use topological ordering in a Bayesian network

DAG must have a node with no incoming (outgoing) edges

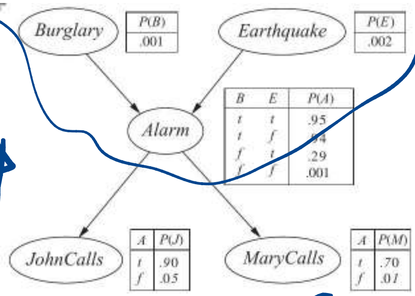
Suppose not



V_{n-1}

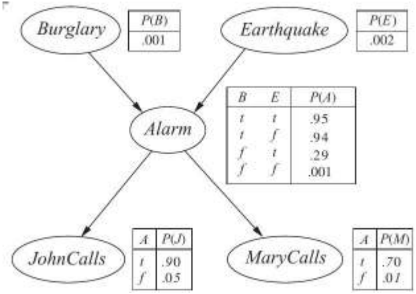
Evaluating a network

- $P(x_1, x_2, \ldots, x_n) = P(x_1 \mid x_2, \ldots, x_n) P(x_2 \mid x_3, \ldots, x_n) \cdots P(x_{n-1} \mid x_n) P(x_n)$
- Can choose any ordering of $x_1, x_2, \ldots, x_n$
- Use topological ordering in a Bayesian network
- $P(m, j, a, b, e) =$
  $P(m \mid j, a, b, e) P(j \mid a, b, e) P(a \mid b, e) P(b \mid e) P(e)$



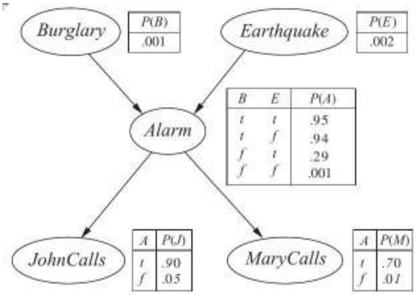
Evaluating a network

- $P(x_1, x_2, \ldots, x_n) = P(x_1 \mid x_2, \ldots, x_n) P(x_2 \mid x_3, \ldots, x_n) \cdots P(x_{n-1} \mid x_n) P(x_n)$
- Can choose any ordering of $x_1, x_2, \ldots, x_n$
- Use topological ordering in a Bayesian network
- $P(m, j, a, b, e) =$
  $P(m \mid j, a, b, e) P(j \mid a, b, e) P(a \mid b, e) P(b \mid e) P(e)$
  $= P(m \mid a) P(j \mid a) P(a \mid b, e) P(b) P(e)$



Evaluating a network

- $P(x_1, x_2, \ldots, x_n) = P(x_1 \mid x_2, \ldots, x_n)P(x_2 \mid x_3, \ldots, x_n) \cdots P(x_{n-1} \mid x_n)P(x_n)$
- Can choose any ordering of $x_1, x_2, \ldots, x_n$
- Use topological ordering in a Bayesian network
- $P(m, j, a, b, e) =$
  $P(m \mid j, a, b, e)P(j \mid a, b, e)P(a \mid b, e)P(b \mid e)P(e)$
  $= P(m \mid a)P(j \mid a)P(a \mid b, e)P(b)P(e)$
- $P(m, j, b) =$
  $$\sum_{a=0}^{1}\sum_{e=0}^{1} P(m \mid a)P(j \mid a)P(a \mid b, e)P(b)P(e)$$

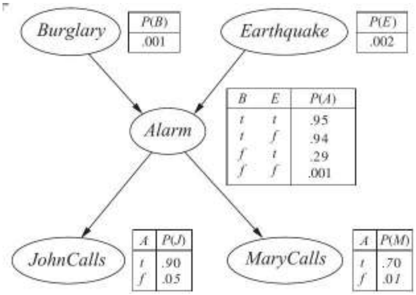


$P(b)(t_1 + t_2 + t_3 + t_4)$

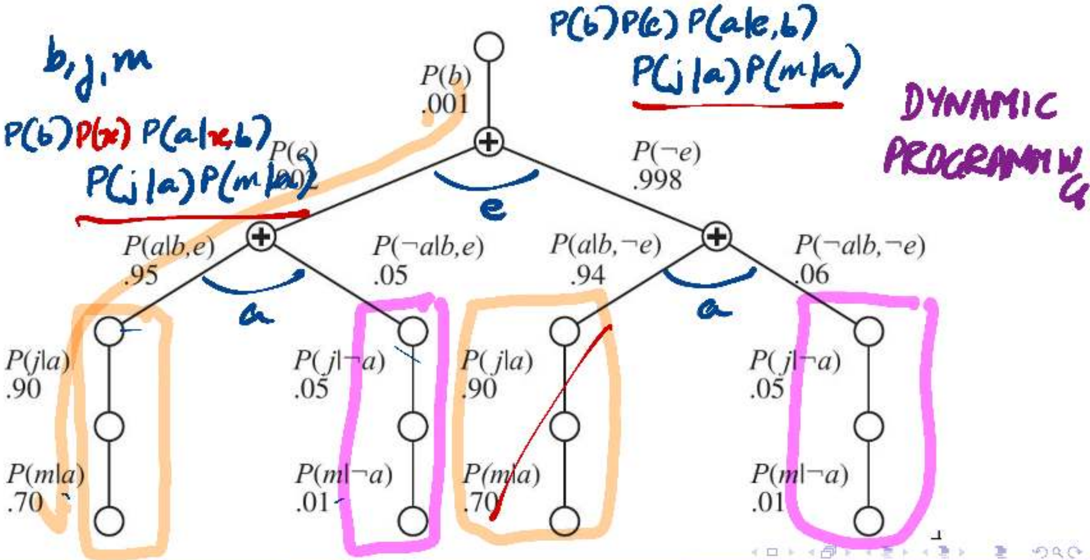
$t_1 P(b) + t_2 P(b) \ldots + t_4 P(b)$

Evaluating a network

- $P(x_1, x_2, \ldots, x_n) = P(x_1 \mid x_2, \ldots, x_n) P(x_2 \mid x_3, \ldots, x_n) \cdots P(x_{n-1} \mid x_n) P(x_n)$
- Can choose any ordering of $x_1, x_2, \ldots, x_n$
- Use topological ordering in a Bayesian network
- $P(m, j, a, b, e) =$
  $P(m \mid j, a, b, e) P(j \mid a, b, e) P(a \mid b, e) P(b \mid e) P(e)$
  $= P(m \mid a) P(j \mid a) P(a \mid b, e) P(b) P(e)$
- $P(m, j, b) =$
  $$\sum_{a=0}^{1} \sum_{e=0}^{1} P(m \mid a) P(j \mid a) P(a \mid b, e) P(b) P(e)$$
- $P(m, j, b) = P(b) \sum_{e=0}^{1} P(e) \sum_{a=0}^{1} P(m \mid a) P(j \mid a) P(a \mid b, e)$

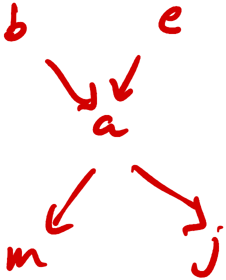


Evaluation tree



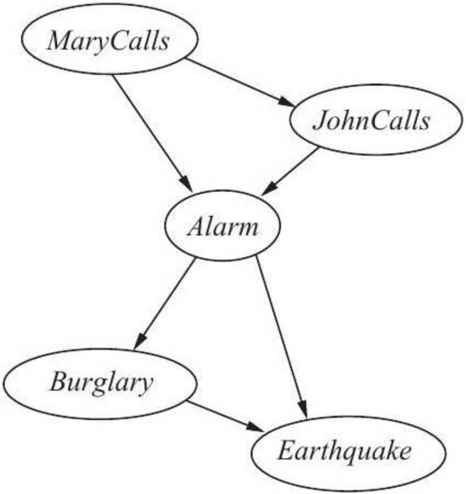
Designing the Bayesian network

- Need to choose node ordering wisely to get a compact Bayesian network



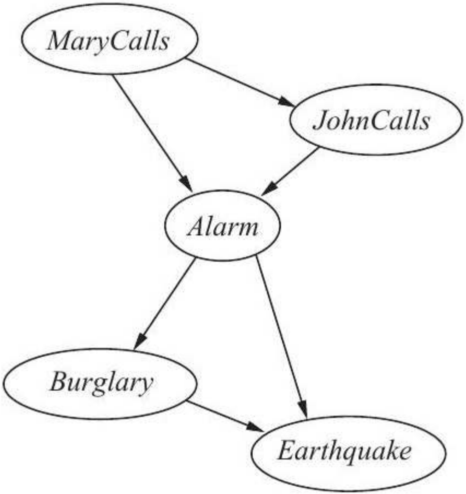
Designing the Bayesian network

- Need to choose node ordering wisely to get a compact Bayesian network
- Ordering *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake* produces this network



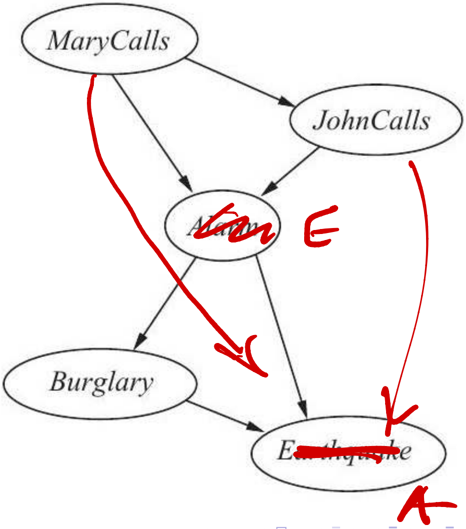
Designing the Bayesian network

- Need to choose node ordering wisely to get a compact Bayesian network
- Ordering *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake* produces this network
- Ordering *MaryCalls*, *JohnCalls*, *Earthquake*, *Burglary*, *Alarm* is even worse



Designing the Bayesian network

- Need to choose node ordering wisely to get a compact Bayesian network
- Ordering *MaryCalls*, *JohnCalls*, *Alarm*, *Burglary*, *Earthquake* produces this network
- Ordering *MaryCalls*, *JohnCalls*, *Earthquake*, *Burglary*, *Alarm* is even worse
- Causal model (causes to effects) works better than diagnostic model (effects to causes)



Complexity of exact inference

- Exact inference of Bayesian networks is NP-complete

Complexity of exact inference

- Exact inference of Bayesian networks is NP-complete
- Boolean formula in Conjunctive Normal Form (CNF)
  - Boolean variables $\{x_1, x_2, \ldots, x_n\}$
  - A literal ℓ_i is either x_i or $\neg x_i$ — not (x_i)

Complexity of exact inference

- Exact inference of Bayesian networks is NP-complete
- Boolean formula in Conjunctive Normal Form (CNF)
  - Boolean variables $\{x_1, x_2, \ldots, x_n\}$
  - A literal ℓ_i is either x_i or $\neg x_i$
  - A clause is a disjunction of literals, $\ell_{j_1} \vee \ell_{j_2} \vee \cdots \vee \ell_{j_k}$

$x_3 \vee \neg x_7 \vee x_9 \vee x_{11}$

Complexity of exact inference

- Exact inference of Bayesian networks is NP-complete
- Boolean formula in Conjunctive Normal Form (CNF)
  - Boolean variables $\{x_1, x_2, \ldots, x_n\}$
  - A literal ℓ_i is either x_i or $\neg x_i$
  - A clause is a disjunction of literals, $\ell_{j_1} \vee \ell_{j_2} \vee \cdots \vee \ell_{j_k}$
  - A CNF formula is a conjunction of clauses, $C_1 \wedge C_2 \wedge \cdots \vee C_m$

$$(x_1 \vee x_2) \wedge (\neg x_1 \vee x_3) \wedge (\neg x_3 \vee \neg x_1)$$

Complexity of exact inference

- Exact inference of Bayesian networks is NP-complete
- Boolean formula in Conjunctive Normal Form (CNF)
  - Boolean variables $\{x_1, x_2, \dots, x_n\}$
  - A literal ℓ_i is either x_i or $\neg x_i$
  - A clause is a disjunction of literals, $\ell_{j_1} \vee \ell_{j_2} \vee \cdots \vee \ell_{j_k}$
  - A CNF formula is a conjunction of clauses, $C_1 \wedge C_2 \wedge \cdots \vee C_m$
- SAT — given a formula in CNF, is there an assignment to variables that makes the formula true?

Complexity of exact inference

- Exact inference of Bayesian networks is NP-complete
- Boolean formula in Conjunctive Normal Form (CNF)
  - Boolean variables $\{x_1, x_2, \ldots, x_n\}$
  - A literal ℓ_i is either x_i or $\neg x_i$
  - A clause is a disjunction of literals, $\ell_{j_1} \vee \ell_{j_2} \vee \cdots \vee \ell_{j_k}$
  - A CNF formula is a conjunction of clauses, $C_1 \wedge C_2 \wedge \cdots \vee C_m$
- SAT — given a formula in CNF, is there an assignment to variables that makes the formula true?
- 3-SAT — SAT where each clause has exactly 3 literals

$x_1 \vee \neg x_4 \vee x_9$

3 literals

Complexity of exact inference

- Exact inference of Bayesian networks is NP-complete
- Boolean formula in Conjunctive Normal Form (CNF)
  - Boolean variables $\{x_1, x_2, \ldots, x_n\}$
  - A literal ℓ_i is either x_i or $\neg x_i$
  - A clause is a disjunction of literals, $\ell_{j_1} \vee \ell_{j_2} \vee \cdots \vee \ell_{j_k}$
  - A CNF formula is a conjunction of clauses, $C_1 \wedge C_2 \wedge \cdots \vee C_m$
- SAT — given a formula in CNF, is there an assignment to variables that makes the formula true?
- 3-SAT — SAT where each clause has exactly 3 literals
- Both SAT and 3-SAT are NP-complete
  - No known efficient algorithm — try all possible valuations

Complexity of exact inference

- Exact inference of Bayesian networks is NP-complete
- Boolean formula in Conjunctive Normal Form (CNF)
  - Boolean variables $\{x_1, x_2, \ldots, x_n\}$
  - A literal ℓ_i is either x_i or $\neg x_i$
  - A clause is a disjunction of literals, $\ell_{j_1} \vee \ell_{j_2} \vee \cdots \vee \ell_{j_k}$
  - A CNF formula is a conjunction of clauses, $C_1 \wedge C_2 \wedge \cdots \vee C_m$
- SAT — given a formula in CNF, is there an assignment to variables that makes the formula true?
- 3-SAT — SAT where each clause has exactly 3 literals
- Both SAT and 3-SAT are NP-complete
  - No known efficient algorithm — try all possible valuations

NP-Complete problem $\xrightarrow{\text{transform}}$ My Problem

Solution $\xleftarrow{\text{implies}}$ Solution

$f \xrightarrow{\text{quick}}$ Network N_f

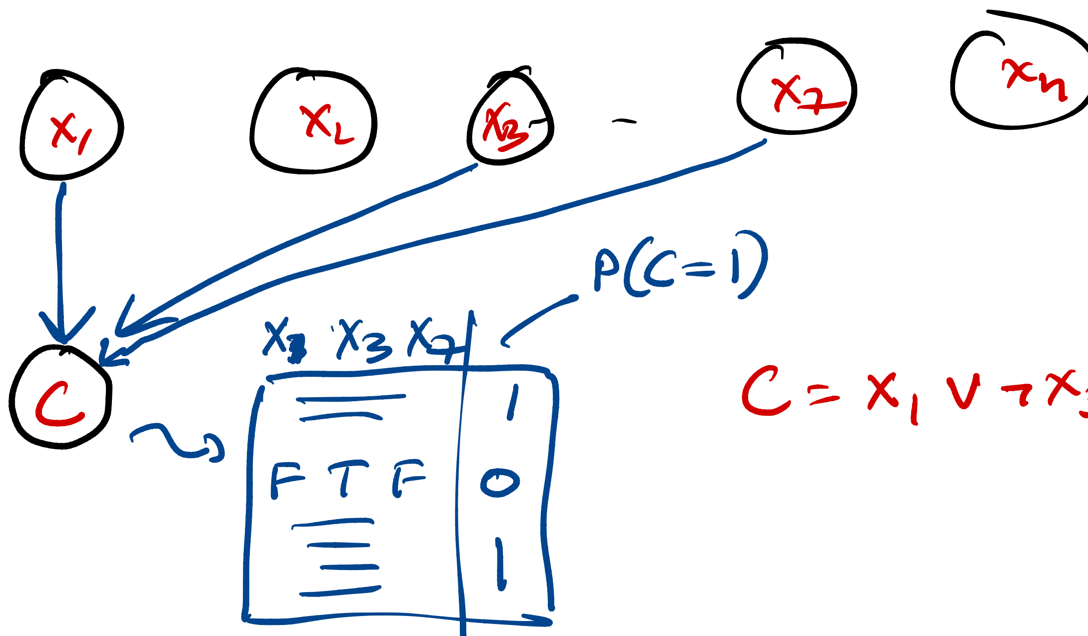
X $\Downarrow$ quick! $\qquad\qquad$ $\downarrow$ quick X

Solution for f $\xleftarrow{\text{quick}}$ solve N_f

Transforming 3-CNF to Bayesian network inference

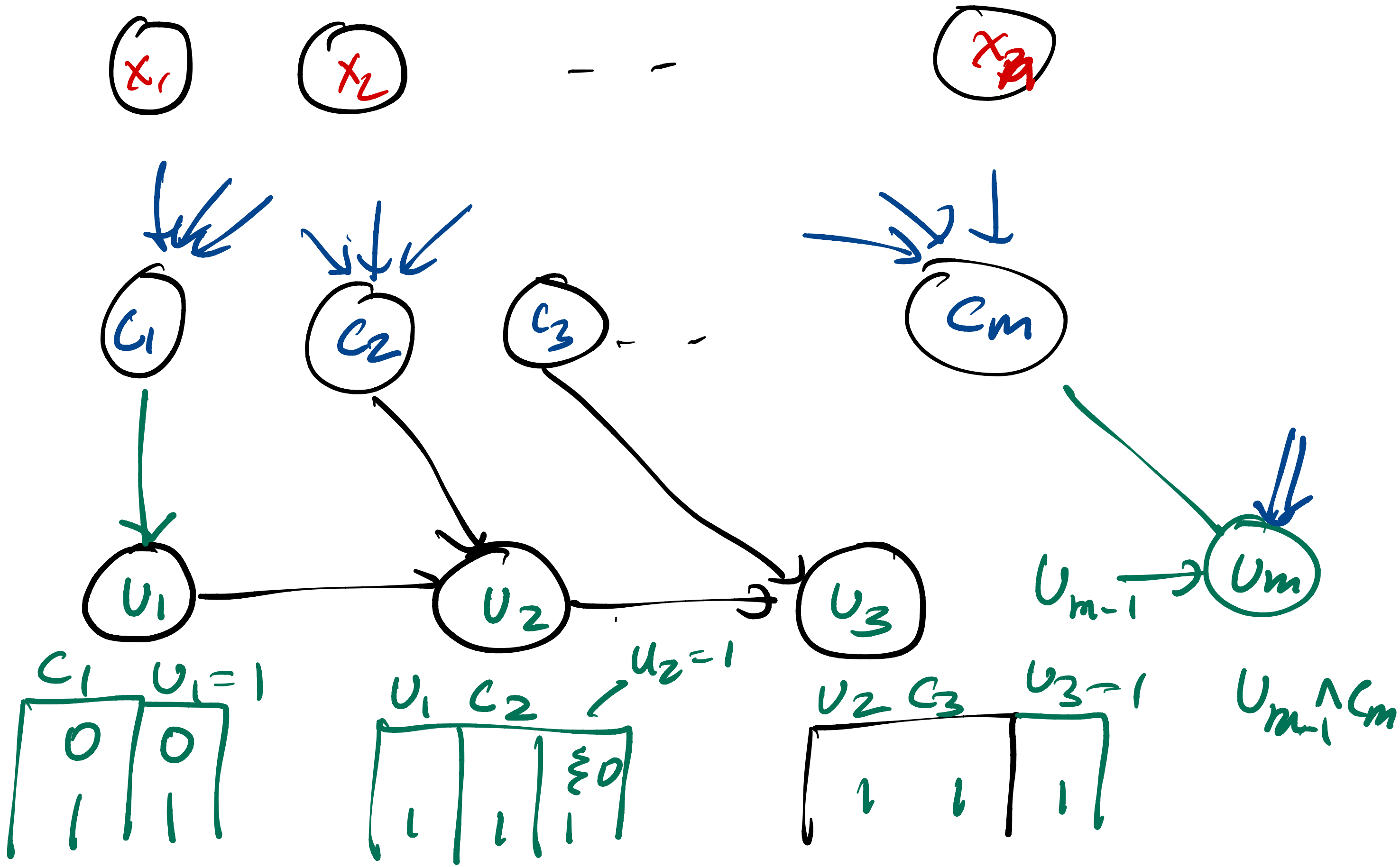
Variables $x_1, x_2, \ldots, x_n$

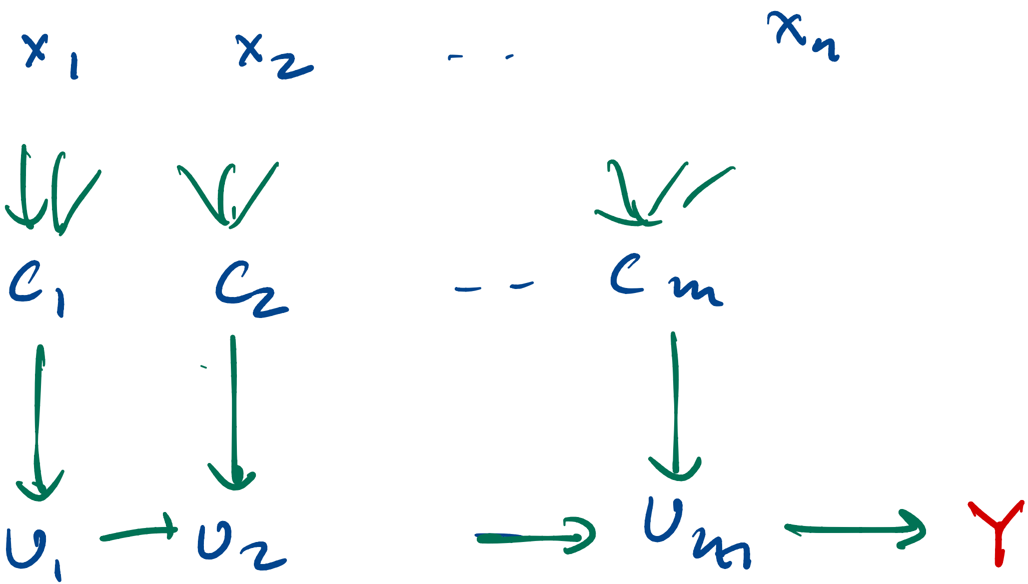
Clause $l_1 \vee l_2 \vee l_3$



$$C = x_1 \vee \neg x_3 \vee x_7$$

$C_1 \wedge C_2 \ldots \wedge C_m$





$Y = 1$ if $u_m = 1$

$P(Y=1) > 0$?