COMP9418: Advanced Topics in Statistical Machine Learning

Introduction

Instructor: Gustavo Batista

University of New South Wales

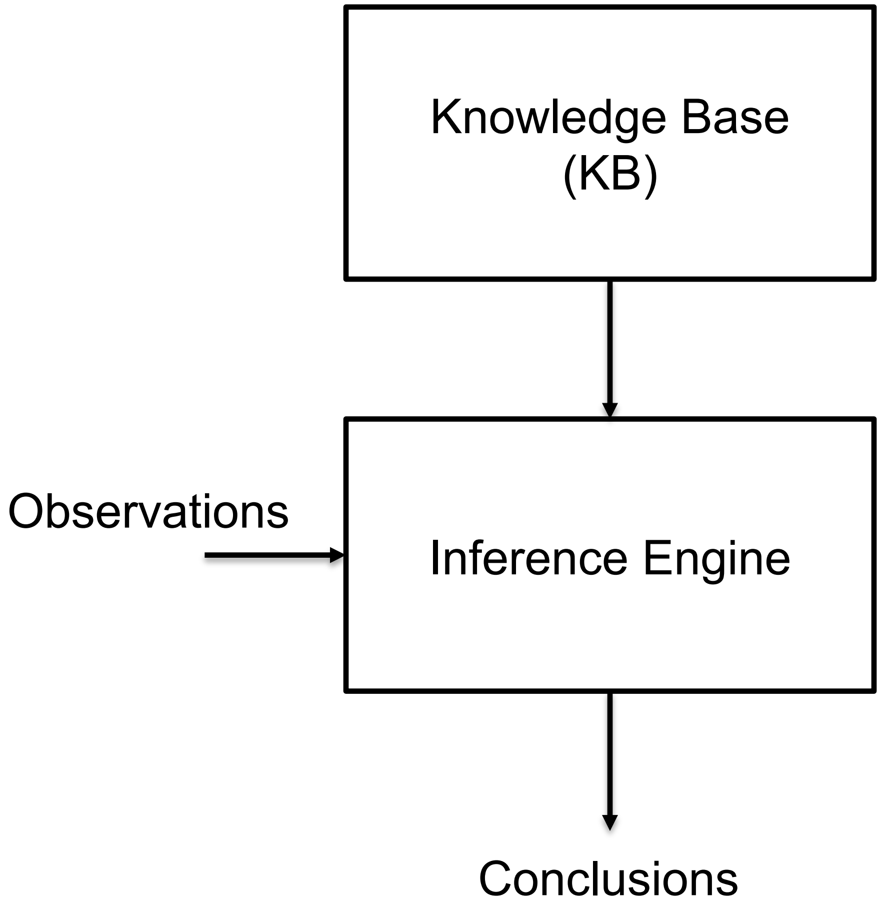
Introduction

- This lecture provides an overview on Probabilistic Graphical Models (PGMs)
  - We discuss the long pathway probabilistic reasoning had until its acceptance in AI
  - We provide a quick overview of Bayesian networks

- We conclude with a list of established PGMS that we will study in this course

Automated Reasoning

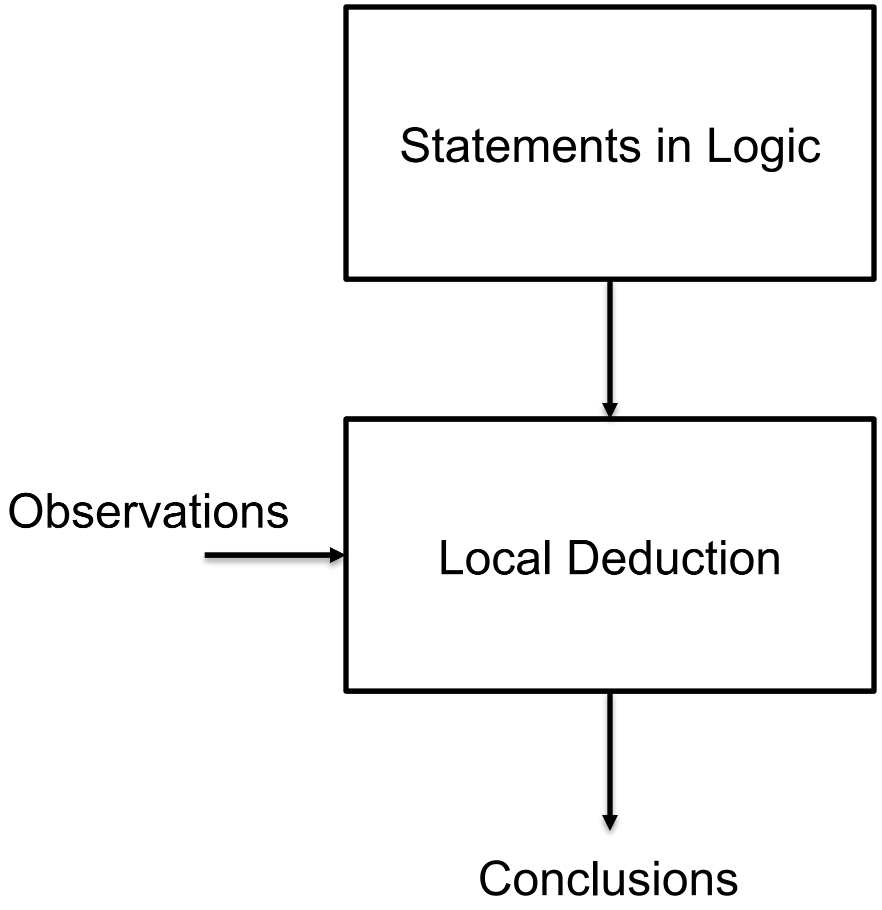
- Automated reasoning is an objective of Artificial Intelligence since its early days
- John McCarthy made an influential proposal that consists of
  - A knowledge base (KB) – encodes what we know about the world
  - Reasoner – acts on the KB to answer queries
- This proposal has an important contribution
  - The separation of the KB (what we know) from the reasoners (how we think)
  - The KB can be domain-specific (changes from applications), while the reasoner is general and fixed



J. McCarthy. Programs with common sense. In Proceedings of the Teddington Conference on the Mechanization of Thought Processes, 1959.

Knowledge-based Systems

- This proposal is the basis for a class of methods known as *knowledge-based* or *model-based systems*
- However, McCarthy's proposal committed with logic as the representation language of the KB
  - This was later revised by McCarthy
  - The main idea remains powerful in the context of others forms of reasoning, including probabilistic reasoning
- In probabilistic reasoning
  - The KB is a Graphical Model such as a Bayesian network
  - Inference engine is based on the laws of probability theory



Monotonic Logic

- Deductive logic is *monotonic*
  - It lacks the ability to dynamically assert and retract assumptions
  - However, such ability is typical of common-sense reasoning
- This problem leaded to the proposal of *non-monotonic logics*
  - Mechanism to manage assumptions
  - It turned out to be a very challenging problem, including conflicts about assumptions
- A different pathway relies on a more fundamental notion of *degree of belief*

If Δ logically implies α,
then Δ and Γ will also logically imply α

If a bird is normal, it will fly

Degree of Belief

- Degree of belief is a number assigned to a proposition
  - Instead assuming a bird is normal and concluding it can fly
  - We assign a degree of believe to its normality, say 99%
  - Use this to derive a corresponding degree of believe in the bird flying ability
- Degrees of belief have different interpretations
  - Notion of possibility used in fuzzy logic
  - Probability, focus of this course
- Degrees of belief are updatable
  - Upwards or downwards
  - Governed by the notion of *probability calculus*

If a bird is normal, it will fly

We can assume a bird is normal with probability 99%, and revise to, say, 20% after learning that its wing is wounded

Decision Theory

- After forming beliefs we usually want to make decisions
  - However, with degree of belief we have to make decisions without assuming any particular state
  - We need a *decision theory*, whose purpose is to convert degree of belief into definitive decisions
- Decision theory needs to bring some additional information
  - Costs of various decisions
  - Rewards or penalties associated with their outcomes
- Decision theory is an essential complement to the theory of probability reasoning

We want to capture a bird worth \$40. We have two methods. The first costs \$30 and is guaranteed to capture the bird, whether flying or not. The second costs \$10 and guarantees the capture of non-flying birds while it may capture a flying bird with 25% probability

Probabilities Meaning

- Probabilities can be interpreted as
  - Objective frequencies
  - Subjective degrees of belief
- There is a classical controversy which interpretation should be used
- In this course, we will use both interpretations
  - This will not impact the techniques discussed
  - Both interpretations are governed by the same laws of probability

Probabilistic Reasoning

- Probability theory has been around for centuries
  - AI required its utilization at the scale and rate never attempted before
  - This created some key computational challenges that needed to be confronted by the first time
- Also, probabilistic methods had to compete with existing ones
  - The responses to these challenges composes most of the material of this course
  - We will review some of these challenges as a motivation to the utility and significance of the covered topics

Probabilistic Reasoning: Criticism

- Initial attempts to use degrees of belief in AI significant were received with criticism
- Cognitive
  - Humans do not use degrees of belief in reasoning
  - Important in early AI since imitating human cognition was highly valued
- Pragmatic
  - Availability of degrees of belief
  - At the time, KBs were obtained through expert elicitation
  - Robustness of the probabilistic reasoning
- Computational
  - Scale probabilistic reasoning can handle
  - Concerns with representing the joint probability distribution which grows exponentially in the number of variables

Probabilistic Reasoning: Second Chance

- Limitations in logic opened a new opportunity for probabilities in AI in the 80s.
  - Judea Pearl was one of the pioneers in the area
  - He advocated in favour of a numeric formalism
  - Developed methods for representation and computation with probabilities
- Pearl demonstrated the benefits for the probabilistic approach
  - $P(A) > P(A|B)$ – belief in A can decrease with observation of B
  - Development of Bayesian networks as response to representation and computational challenges
- Bayesian networks
  - Can represent exponentially sized probability distributions
  - Have algorithms such as polytree and jointree that can handle arbitrary networks

Judea Pearl
2011 Turing Award

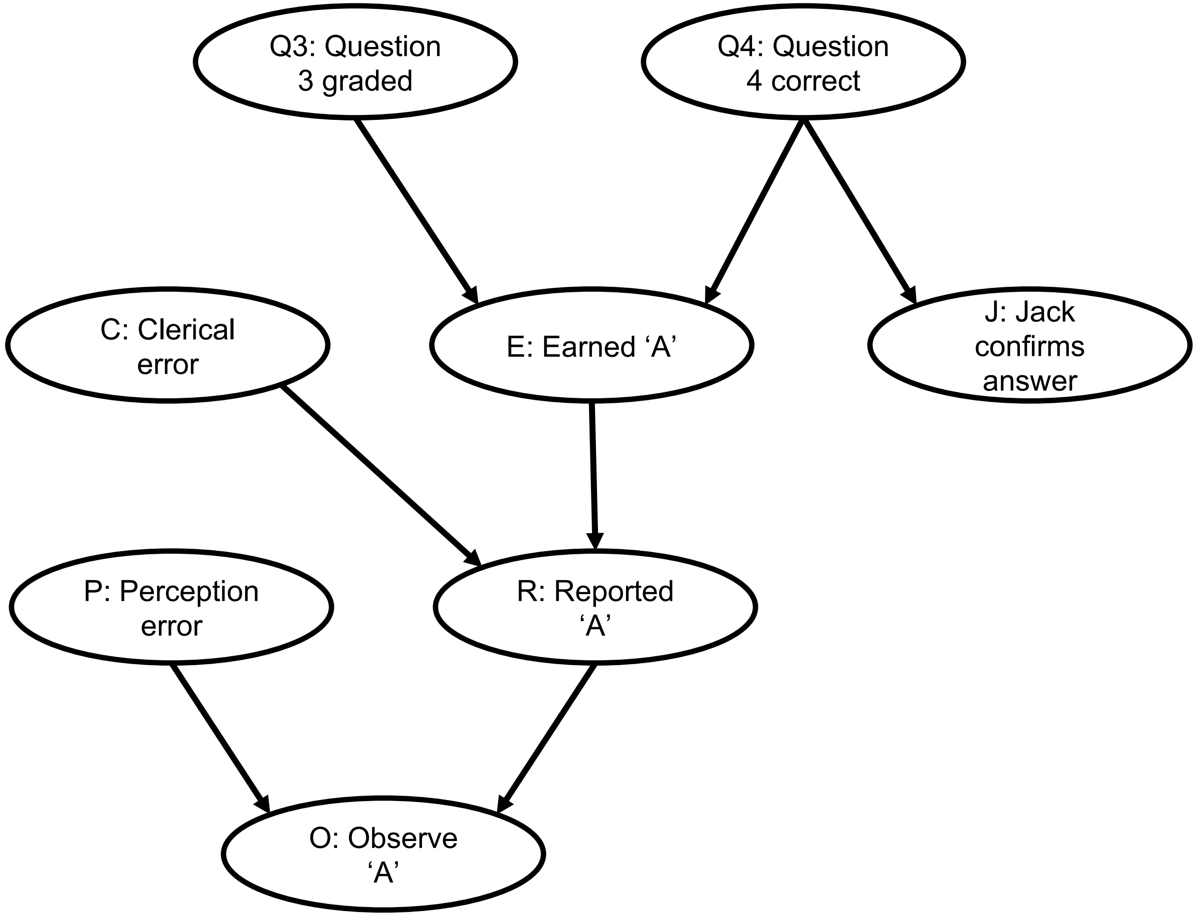
Bayesian Networks

- Consider a student, Drew, who finished the final exam for physics class and received a B grade instead of an expected A

Let me first check that I am looking at the grade of my physics class instead of some other class. Hmm! It is indeed physics. Is it possible the professor made a mistake in entering the grade? I don't think so… I have taken a few classes with him, and he has proven to be quite careful and thorough. Well, perhaps he did not grade my Question 3, as I wrote the answer on the back of the page in the middle of a big mess. I think I will need to check with him on this . . . I just hope I did not miss Question 4; it was somewhat difficult and I am not too sure about my answer there. Let me check with Jack on this, as he knows the material quite well. Ah! Jack seems to have gotten the same answer I got. I think it is Question 3 after all . . . I'd better see the professor soon to make sure he graded this one.

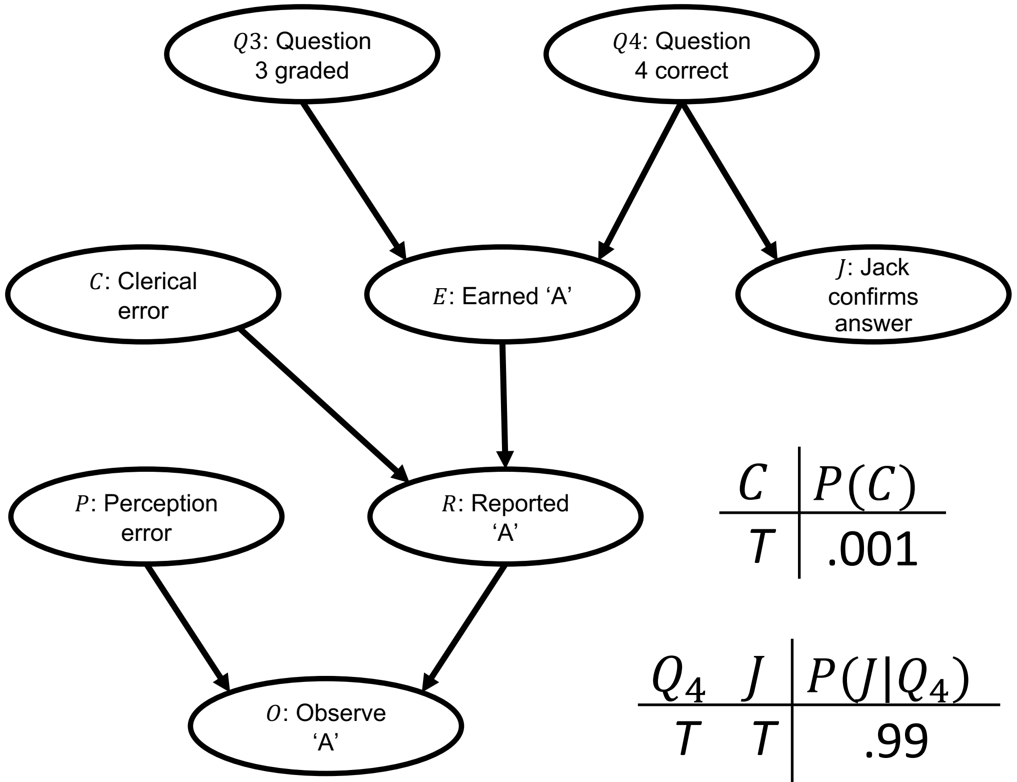
Bayesian Networks: Structure

- Bayesian networks have a structural component
  - Variables represent the relevant primitive propositions
  - Edges convey information about the dependencies between variables
- Frequently, we think of edges as causal influences
  - Causation is a very valuable guide in constructing these networks
  - Although, Bayesian networks can have an interpretation completely independent of causation



Bayesian Networks: Probabilities

- Probabilities quantify the relationships between variables and their parents
  - It is local information
  - Variable E: probabilities only reference E and its direct causes Q_3 and Q_4
  - Variable C: only reference this variable since it has no causes
- We never specify a quantitative relationship unless they have an edge
  - Other probabilities are computed by inference algorithms



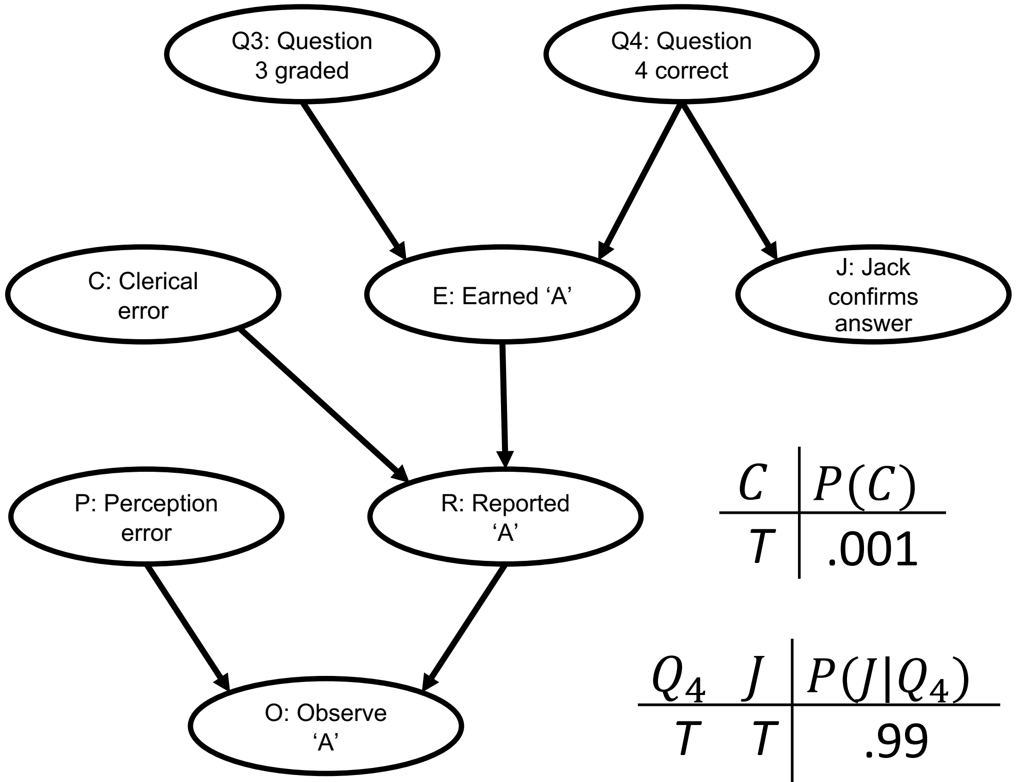
C	$P(C)$
T	.001

Q_4	J	$P(J\mid Q_4)$
T	T	.99
F	T	.20

Q_3	Q_4	E	$P(E\mid Q_3, Q_4)$
T	T	T	1
T	F	T	0
F	T	T	0
F	F	T	0

Bayesian Networks: Representation

- Bayesian networks are attractive as representation tool
  - It is guaranteed to define a unique probability distribution over the variables
  - They are modular in the sense consistency and completeness are ensured using local tests only to variables and direct causes
  - They are compact as it allows to specify an exponentially sized probability distribution using a polynomial number of probabilities (assuming the number of direct causes are small)



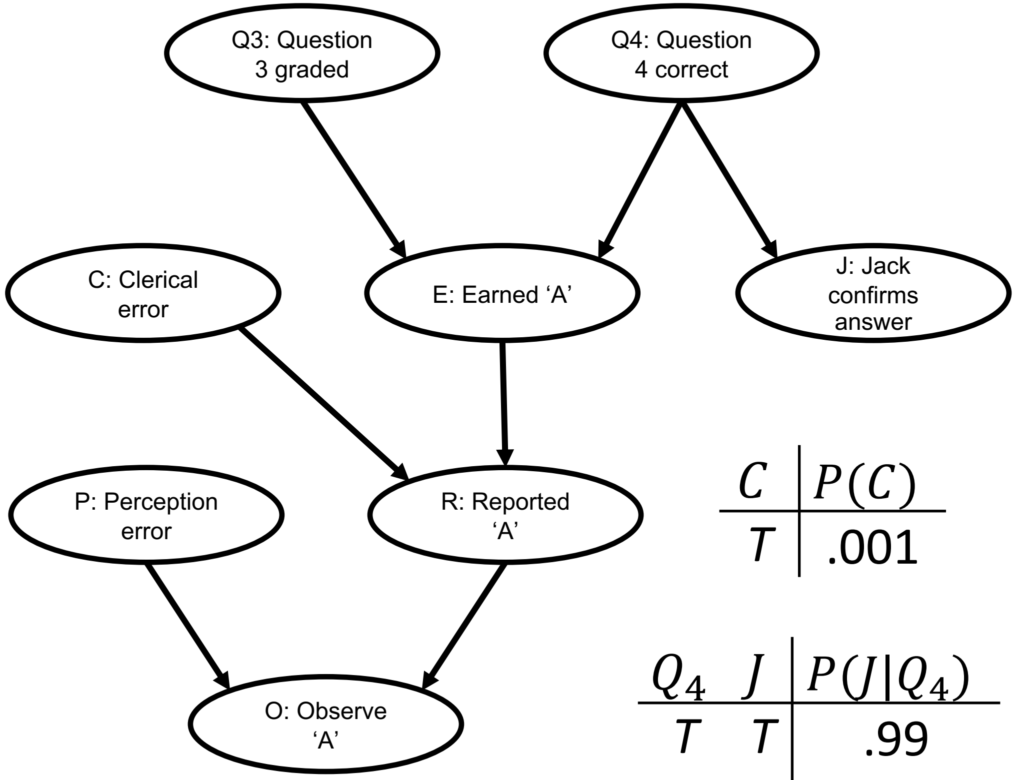
C	$P(C)$
T	.001

Q_4	J	$P(J\|Q_4)$
T	T	.99
F	T	.20

Q_3	Q_4	E	$P(E\|Q_3,Q_4)$
T	T	T	1
T	F	T	0
F	T	T	0
F	F	T	0

Bayesian Networks: Modelling

- There are three major approaches for modelling a Bayes network
  - One builds the network from their own knowledge or eliciting from others
  - The structure comes from the problem specification
  - Learning from data, in this case, we can learn the probabilities, the structure or both
- Learning is an inductive process
  - Machine learning approach
  - Two major approaches: maximum likelihood and the Bayesian approach (likelihood + prior)



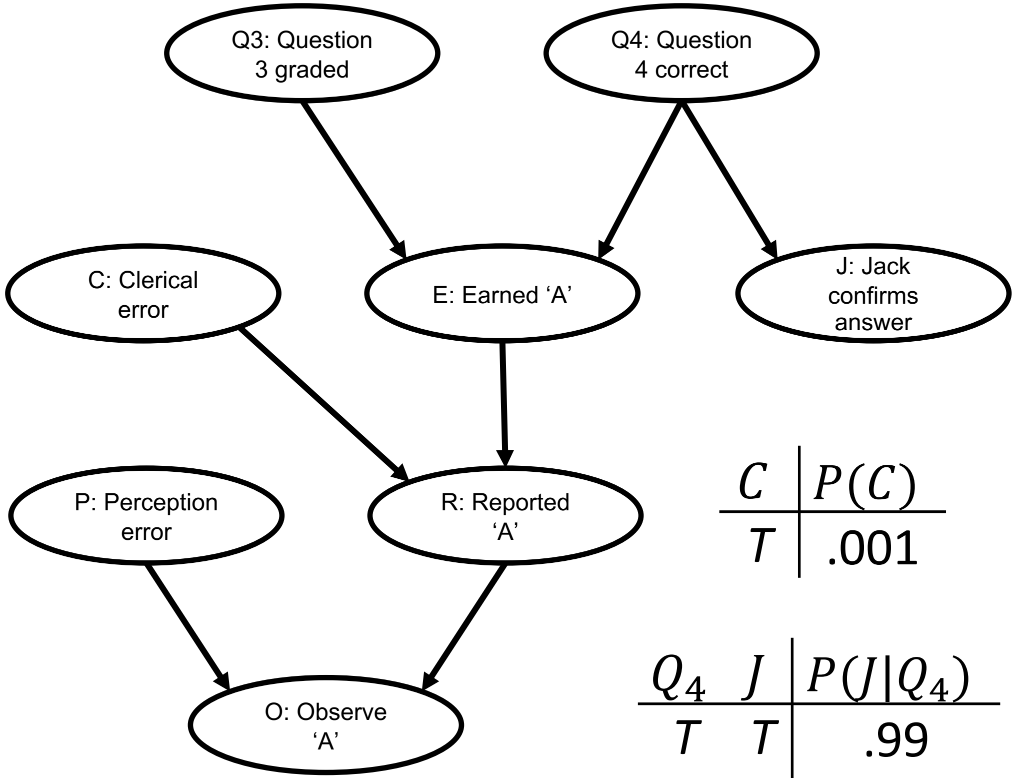
C	$P(C)$
T	.001

Q_4	J	$P(J\|Q_4)$
T	T	.99
F	T	.20

Q_3	Q_4	E	$P(E\|Q_3,Q_4)$
T	T	T	1
T	F	T	0
F	T	T	0
F	F	T	0

Bayesian Networks: Reasoning

- The Bayesian network assigns a unique probability to each proposition
  - The network only specify some of these probabilities
- However, consider the following
  - $P(E = true)$: The probability that Drew earned an A grade
  - $P(Q_3 = false|E = false)$: The probability that Q3 was graded, given that Drew did not earn an A grade
  - $P(Q_4 = true|E = true)$: The probability that Jack obtained the same answer as Drew on Q4, given Drew earned an A grade
- None of these probabilities are part of the network
  - Yet, the network is guaranteed to imply a unique value for each of these probabilities



C	$P(C)$
T	.001

Q_4	J	$P(J|Q_4)$
T	T	.99
F	T	.20

Q_3	Q_4	E	$P(E|Q_3, Q_4)$
T	T	T	1
T	F	T	0
F	T	T	0
F	F	T	0

Probabilistic Graphical Models

- Bayesian networks are one example of graphical models
- Graphical models can be classified by 3 properties
  - Directed or undirected
  - Static or dynamic
  - Probabilistic or decisional
- Directed or undirected
  - Undirected represent symmetric relations
- Static and dynamic
  - Dynamic models represent variables across different times
- Probabilistic or decisional
  - Decisional models include random and utility variables

PGM*	Directed/ Undirected	Static/ Dynamic	Probabilistic/ Decisional
Bayesian classifiers	D	S	P
Markov chains	D	D	P
Hidden Markov models	D	D	P
Markov random fields	U	S	P
Bayesian networks	D	S	P
Dynamic Bayesian networks	D	D	P
Influence diagrams	D	S	D
Markov decision processes (MDPs)	D	D	D
Partially observable MDPs	D	D	D

*Sucar, L. Probabilistic Graphical Models – Principles and Applications. Springer, 2015.

Conclusion

- Probabilistic Graphical Models have became very popular in the last years
  - It represents a joint research from Statistics (probabilistic reasoning) and Computer Science (graphs)
  - Both components are essential to model problem with uncertainty, large number of variables that require efficient algorithms
- Tasks
  - Read chapter 1 from the textbook (Darwiche)