CES -161 - Modelos Probabilísticos em Grafos

Prof. Paulo André Castro
pauloac@ita.br
www.comp.ita.br/~pauloac
Sala 110,
IEC-ITA

Paulo André Lima de Castro

- Bolsista de Produtividade Desen. Tec. e Extensão Inovadora do CNPq Nível 2.

- Engenheiro de Computação pelo Instituto Tecnológico de Aeronáutica (ITA, 1997), Mestre e Doutor pela Escola Politécnica da Universidade de São Paulo (Poli/USP 2009). Pós-doutorado na *City University of New York* (CUNY, 2013).
- Atualmente é professor do Instituto Tecnológico de Aeronáutica (ITA) e Chefe do Departamento de Metodologias de Computação da Divisão de Ciência da Computação do ITA.

- Participei de diversos projetos de Pesquisa e Desenvolvimento incluindo desenvolvimento de simuladores, avaliação de segurança da informação em sistemas computacionais e aplicação de técnicas inteligentes em sistemas distribuídos.

- Realizo pesquisas na área de Inteligência Artificial com ênfase em Sistemas multiagentes, atuando principalmente nos seguintes temas: agent-based finance, agentes autônomos e aplicações de técnicas inteligentes especialmente em economia e finanças

Ementa da disciplina

- **CES-161 - Modelos Probabilísticos em Grafos** - Introdução, conceitos e Raciocínio Probabilístico. Modelos de Markov. Introdução a Redes Bayesianas e Inferência Bayesiana. Análise de Decisão. Aplicações de Redes Bayesianas. Aprendizagem de Modelos Causais. Classificadores, Regressores, Avaliação de modelos de Machine learning. Validação cruzada e Overfittting. Cassificadores Bayesianos e métodos Ensemble. Knowledge Engineering em ambientes com incerteza. Aprendizado de Máquina no contexto financeiro. Cross-validation e Backtesting em Finanças.

- **Bibliografia**: Korb, K. Nicholson, A. **Bayesian Artificial Intelligence.** CRC Press. 2011. Witten, I., Frank, E. **Data Mining: Practical Machine learning Tools and Techniques.** 4a. ed. Elsevier. 2016. Prado, M.L. **Advances in Financial Machine Learning**. Wiley. 2018. RUSSEL, S.; NORVIG, P. **Inteligência Artificial: Uma abordagem moderna.** 3a. ed. Rio de Janeiro: Elsevier Editora, 2009.

- **Outras referências:** Pearl, Judea. **Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference.** Morgan Kaufmann, San Mateo, California. 1988.

Outlook

- Chap. I. Introduction
- Chap. 2. Rational Decisions
- Chap. 3. Decision Making with Bayesian Networks
- Chap. 4. Learning Probabilistic Models and Knowledge Engineering
- Chap. 5. Markov Decision Process
- Chap. 6. Reinforcement Learning
- Chap. 7. Artificial Intelligence and Machine Learning in Financial Environments

Avaliações

- 1 Prova na última semana

- 1 Projeto de construção de Modelo probabilístico em Grafo (Final da disciplina)

Definições de IA

Pensando como seres humanos	Pensando Racionalmente
"O novo e interessante esforço para fazer os computadores pensarem (…) máquinas com mentes, no sentido total e literal" (Haugeland, 1985) "[Automação de] atividades que associamos ao pensamento humano, atvidades como a tomada de decisões, a resolução de problemas, o aprendizado.." (Bellman, 1978)	"O estudo das faculdades mentais pelo uso de modelos computacionais" (Charniak e McDermoot, 1985) "O estudo das computações que tornam possível perceber, raciocinar e agir" (Winston, 1992)
Agindo como Seres Humanos	**Agindo Racionalmente**
"A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas" (Kurzweill, 1990) "O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas por pessoas" (Rich and Knight, 1991)	"**Inteligência Computacional** é o estudo do projeto de agentes inteligentes" (Poole et al. 1998) "IA…está relacionada a um desempenho inteligente de artefatos" (Nilsson, 1998)

Inteligência Artificial – Novo ?

- O termo Inteligência Artificial foi usado oficialmente pela primeira vez no verão de 1956, em um convite para um workshop de 2 meses organizado por John McCarthy, Marvin Minsky, Claude Shannon, e outros…

Artificial Intelligence -birth certificate

- “We propose that a 2 month, 10 man study of **artificial intelligence** be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.” **John McCarthy, Marvin Minsky, Claude Shannon et al.1956**

- Perhaps “computational rationality” would have been more precise and less threatening, but “AI” has stuck....
- **McCarthy** stated that he resisted the terms “computer” or “computational” in deference to Norbert Weiner, who was promoting analog cybernetic devices rather than digital computers

What about .... ?

- Deep learning
  - “Deep learning is a subset of a more general field of artificial intelligence called *machine learning*” Buduma, N. The fundamentals of deep learning.
- Machine Learning
  - Construção de software “...que pode melhorar seu próprio comportamento através do estudo diligente de suas próprias experiências” (Russel, Norvig, 2013)
- Data mining
  - Finding patterns in data that provide insight or enable fast and accurate decision making (Witten,2016) Data Mining: Practical Machine learning)
- Big data
  - Capturing and managing lots of information (computer systems)
  - Analyzing these masses of new data (data mining)

Machine learning

- Definitions of "learning" from dictionary:

  To get knowledge of by study, experience, or being taught } Difficult to measure

- Operational definition:

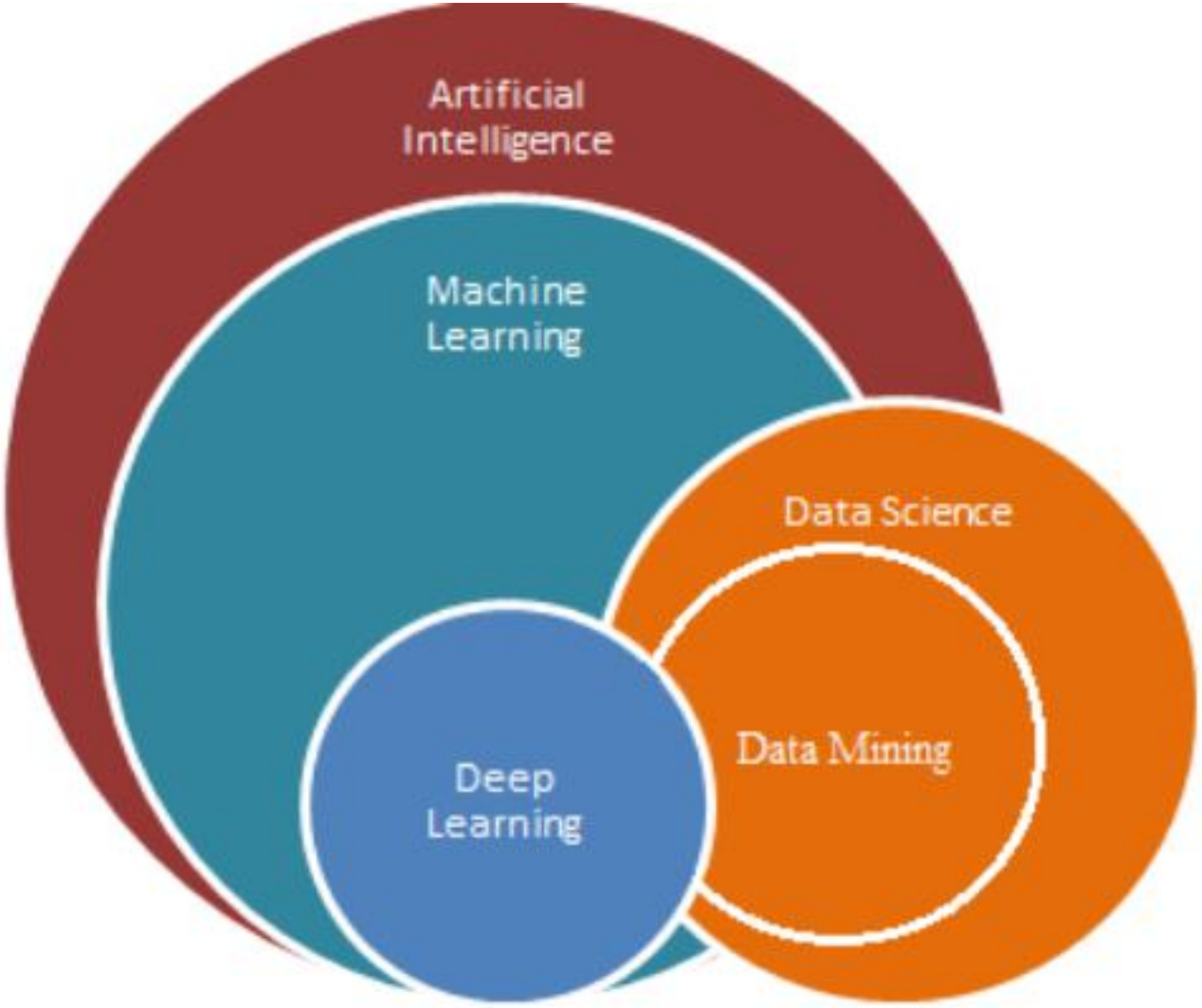
  Things learn when they change their behavior in a way that makes them perform better in the future.

Machine Learning vs Statistcs

"In truth, you should not look for a dividing line between machine learning and statistics because there is a continuum— and a multidimensional one at that—of data analysis techniques" Witten, Data Mining: Practical Machine learning.

In fact, it could be also stated to Machine Learning vs Data mining vs Statistics

A "Reasonable" Graph Representation of Intersections of Related Areas to AI



E Modelos Probabilísticos em Grafos?

- Desafios enfrentados por IA
  - Resolução de Problemas
  - Conhecimento: Raciocínio e planejamento
  - **Incerteza: Conhecimento e Raciocínio**
  - **Aprendizado**
  - Comunicação, percepção e Ação

A IA como um Campo Multidisciplinar

Can a Machine Think?, Turing, A. (1950) Section 1: The imitation game

Teste de Turing

Alan Turing

Defina máquina e pensar….

"O computador passará no teste se um interrogador humano, depois de propor algumas perguntas por escrito, não conseguir descobrir se as respostas escritas vêm de uma pessoa ou não".

Um computador precisaria ter as seguintes capacidades:

- Processamento de linguagem natural (comunicação);
- Representação de conhecimento (armazenar o que sabe);
- Raciocínio automatizado (tirar conclusões a partir das perguntas);
- Aprendizado de máquina (adaptar-se à novas circunstâncias).

Teste de Turing Total:

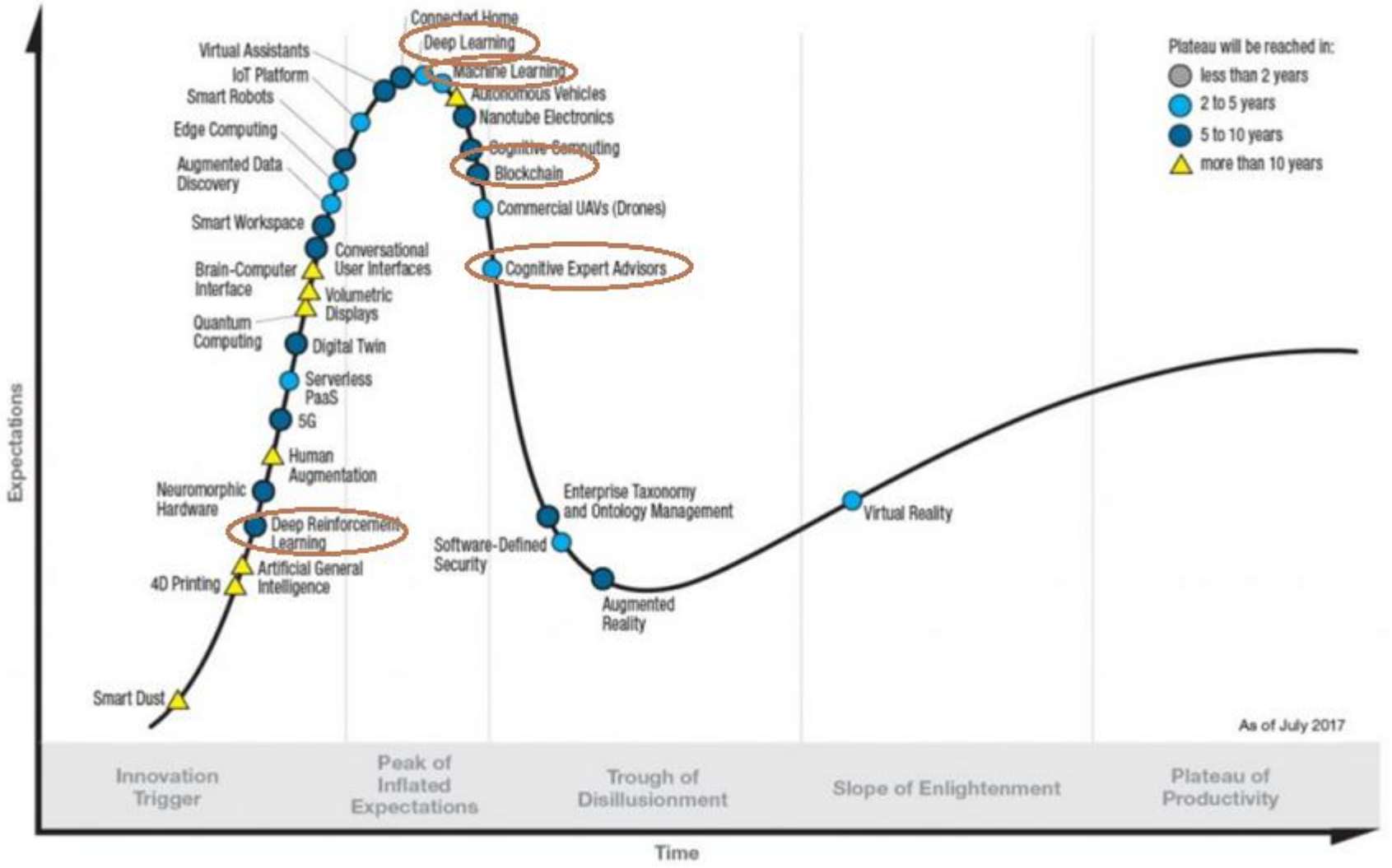
- *Visão computacional (para perceber objetos);*
- *Robótica (movimentar-se e manipular objetos)*
- *Aparência correta....*

Can a Machine Think?, Turing, A. (1950)

- Objeções:
  - The Theological Objection
  - The "Heads in the Sand" Objection
  - The Mathematical Objection

  - The Argument from Consciousness
  - And others....

Soluções e Gartner Hype Cycle



Agentes

- Um agente é tudo que pode ser considerado capaz de *perceber seu* ***ambiente*** por meio de ***sensores*** e de *agir sobre esse ambiente* por intermédio de ***atuadores***.
- **Exemplos:** agente animal, agente robótico, agente de software, termostatos….

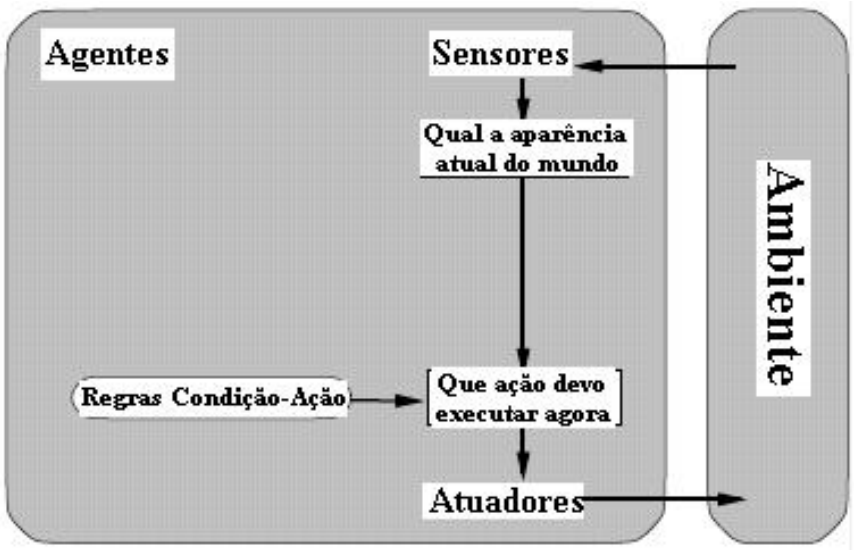
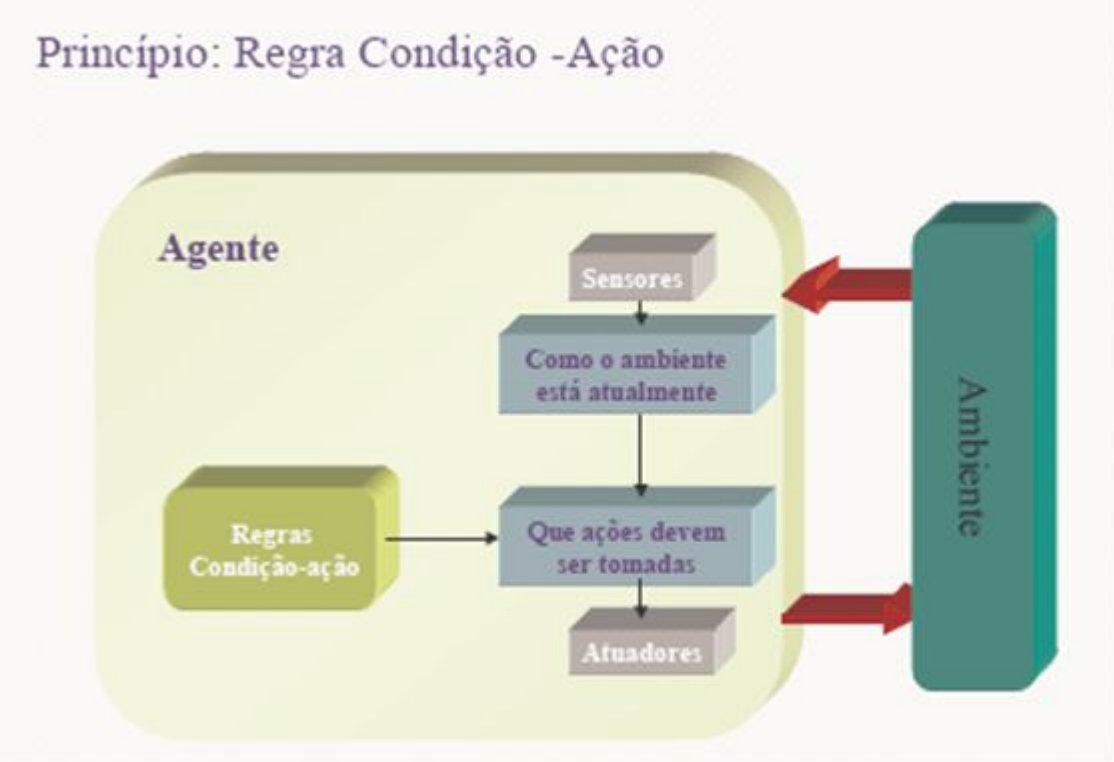


Diagrama esquemático de um agente reativo simples.

Agentes Reativos Simples



Agentes Reativos Baseados em Modelo

Agentes que seguem o mundo

•Mantêm um estado interno que será combinado com as novas percepções.

Agentes Baseados em Objetivos

Agentes Baseados na Utilidade

Agentes Baseados em Aprendizado

Agentes Baseados em Conhecimento

- São agentes que
- *- Conhecem seu mundo através de uma **Base de Conhecimento;***
- *- Raciocinam sobre suas possíveis ações através de uma **Máquina de Inferência.***

- Eles Sabem:
- *- O estado atual do mundo (propriedades relevantes);*
- *- Como o mundo evolui;*
- *- Como identificar estados desejáveis do mundo;*
- *- Como avaliar o resultado das ações;*
- *- Conhecimento sobre conhecimento (meta-conhecimento);*

Agentes Baseados em Conhecimento (Definições Importantes)

- Dados:
- *- Cadeias numéricas ou alfanuméricas que não possuem significado associado;*
- *- Podem ser fatos ou figuras a processar.*
- Informação:
- *- Dados organizados;*
- *- Significam alguma coisa para quem os recebe.*
- Conhecimento:
- *- Representa objetos (entidades) de algum domínio, com suas propriedades e relações.*
- Meta-conhecimento:
- *- Conhecimento sobre o conhecimento disponível.*
- ***Ex:** Regras sobre "como" manipular as regras sobre conhecimento que estão em uma base.*

Agentes Baseados em Conhecimento (Definições Importantes)

- Sistemas Baseados em Conhecimento:
- - *Têm uma Base de Conhecimento e uma Máquina de Inferência associada;*
- - *Formalizam e implementam parte dos* ***agentes.***
- Qual a diferença entre Agentes e Sistemas Baseados em Conhecimento (SBC)?
- - Agentes interagem com o ambiente onde estão imersos através dos SENSORES e ATUADORES;
- Base de Conhecimento:
- - Contém sentenças em uma linguagem de representação de conhecimento;
- - Representações de fatos e regras;
- - Conhecimento em forma "tratável" pelo computador.
- Exemplo: Computador é um aparelho eletrônico.
- DX50 é um computador.
- Mecanismo (Máquina) de Inferência:
- - Responsável por inferir, a partir do conhecimento da base, novos fatos ou hipóteses intermediárias/temporárias.
- Logo: DX50 é um aparelho eletrônico.

Ambiente: Onde os agentes vivem e atuam

- Propriedades dos Ambientes
- Observável x Parcialmente Observável
- Determinístico x Estocástico
- Episódico x Seqüencial
- Estático x Dinâmico
- Discreto x Contínuo
- Agente Único x Multiagente

Uncertainty (Partially observed or stochastic) environments?

1. **Ignorance.** The limits of our knowledge lead us to be uncertain about many things. Does our poker opponent have a flush or is she bluffing?
2. **Physical randomness or indeterminism.** Even if we know everything that we might care to investigate about a coin and how we impart spin to it when we toss it, there will remain an inescapable degree of uncertainty about whether it will land heads or tails when we toss it. A die-hard determinist might claim otherwise, that some unimagined amount of detailed investigation might some-day reveal which way the coin will fall; but such a view is for the foreseeable future a mere act of scientistic faith. We are all practical indeterminists.
3. **Vagueness.** Many of the predicates we employ appear to be vague. It is often unclear whether to classify a dog as a spaniel or not, a human as brave or not, a thought as knowledge or opinion.

Example 1: Breast Cancer

Suponha que a probabilidade de uma mulher tenha 1% de chance de ter cancer. Em uma clínica, há um teste de cancer com 20% de falso positivo e 10% de falso negativo, i.e. 10% das mulheres com cancer terão um resultado negativo. Logo, 90% (das mulheres com câncer) terão um resultado positivo.Uma paciente da clínica teve um resultado positivo de cancer. Qual a probabilidade dela ter cancer realmente?

Como há apenas 20% de chance falso positivo, então seria 80%,certo?

Não! P(Cancer | Pos) não é igual a 1- P(Pos| Not cancer)

$$
\begin{aligned}
P(Cancer|Pos) &= \frac{P(Pos|Cancer)P(Cancer)}{P(Pos)} \\
&= \frac{P(Pos|Cancer)P(Cancer)}{P(Pos|Cancer)P(Cancer)+P(Pos|\neg Cancer)P(\neg Cancer)} \\
&= \frac{0.9\times 0.01}{0.9\times 0.01+0.2\times 0.99} \\
&= \frac{0.009}{0.009+0.198} \\
&\approx 0.043
\end{aligned}
$$

Example 2: People vs Collins

In 1964 an interracial couple was convicted of robbery in Los Angeles, largely on the grounds that they matched a highly improbable profile, a profile which fit witness reports (Sullivan, Sullivan). In particular, the two robbers were reported to be

- A man with a mustache
- Who was black and had a beard
- And a woman with a ponytail
- Who was blonde
- The couple was interracial
- And were driving a yellow car

The prosecution suggested that these characteristics had the following probabilities of being observed at random in the LA area:

1. A man with a mustache 1/4
2. Who was black and had a beard 1/10
3. And a woman with a ponytail 1/10
4. Who was blonde 1/3
5. The couple was interracial 1/1000
6. And were driving a yellow car 1/10

Example 2: People vs Collins – cont.

- The prosecution called an instructor of math from a State university who apparently testified that the "product rule" could be applied. So, the probability of the evidence(e) be collected for an non guilty couple (h) would be:

$$P(e|\neg h) = \prod_i P(e_i|\neg h) = 1/12000000$$

- The prosecution stated that given the evidence the probabiliyt of the couple were innocent was no more than 1/12.000.000. The jury convicted them.
- Is the probability estimate correct?
- No. The product rule does not apply in this case!!
- P(h|e) is not equal to 1-P(e| *not* h)
- What is the probability of the couple being guilty?

Example 2: People vs Collins – cont.

- The pieces of evidence are NOT independent!!!

If, for example, we know of the occupants of a car that one is black and the other has blonde hair, what then is the probability that the occupants are an interracial couple? Clearly not 1/1000! If we know of a man that he has a mustache, is the probability of having a beard unchanged? These claims are preposterous, and it is simply shameful that a judge, prosecutor and defence attorney could not recognize how preposterous they are — let alone the mathematics "expert" who testified to them. Since e_2 implies e_1, while e_2, e_3, e_4 jointly imply e_5 (to a fair approximation), a far better estimate for $P(e|\neg h)$ is $P(e_2|\neg h)P(e_3|\neg h)P(e_4|\neg h)P(e_6|\neg h) = 1/3000$.

- Furthermore, P(h|e) is not equal to 1-P(e| ***not*** h), but:

- And by Sum-out.. $$P(h|e) = \frac{P(e|h)P(h)}{P(e)}$$

- P(e|h) ? $$P(h|e) = \frac{P(e|h)P(h)}{P(e|h)P(h) + P(e|\neg h)P(\neg h)}$$

Example 2: People vs Collins – cont.

Now if the couple in question *were* guilty, what are the chances the evidence accumulated would have been observed? That's a rather hard question to answer, but feeling generous towards the prosecution, let us simplify and say 1. That is, let us accept that $P(e|h) = 1$. Plugging in our assumptions we have thus far:

$$P(h|e) = \frac{P(h)}{P(h) + P(\neg h)/3000}$$

We are missing the crucial prior probability of a random couple being guilty of the robbery. Note that we cannot here use the prior probability of, for example, an interracial couple being guilty, since the fact that they are interracial is a piece of the evidence. The most plausible approach to generating a prior of the needed type is to count the number of couples in the LA area and give them an equal prior probability.

- Let's say there are 1,625,000 eligible males and as many female in Los Angeles area...so:

$$P(h|e) = \frac{1/1625000}{1/1625000 + (1 - 1/1625000)/3000} \approx 0.002$$

Revisão de Conceitos Básicos de Probabilidade

P(A | K) – probabilidade condicional ou posterior.
Crença em *A*, dado o corpo de informação *K.*

P(A) – probabilidade *a priori*: Crença em A, na falta de informação adicional proveniente de K.

Uma Variável aleatória tem um domínio (conjunto de valores)e associada a cada um a probabilidade de ocorrência daquele valor. Essa função é chamada de distribuição de Probabilidade.
Exemplo:
Variável Tempo = {Sol, Chuva, Nublado}
P*(Tempo) – é uma distribuição de probabilidade*
P*(Tempo) = <0,7;0,2;0,1>*

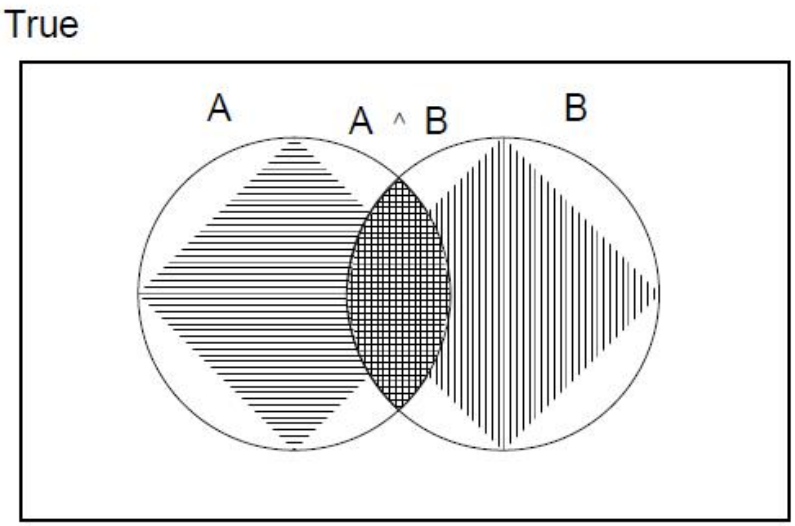
P*(Tempo=sol) = 0.7*
P*(Tempo=chuva) = 0.2*
P*(Tempo=nublado) = 0.1*

No caso contínuo, usa-se o termo função de densidade de probabilidade. Vamos focar no caso discreto.

Axiomas da Probabilidade

- Para qualqueis proposições A e B

1. $0 \leq P(A) \leq 1$
2. $P(True) = 1$ and $P(False) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



Probabilidade condicional

Probabilidade condicional ou posterior, e.g., P(cárie|dordedente) = 0.8

i.e., dado que dordedente é tudo que conheço, a chance de cárie (vista por mim) é de 80%.

P(Cárie | Dordedente) = Vetor de 2 elementos cada um com dois elementos. Por Exemplo: P(Cárie | Dordedente) = <<0,8;0,2>;<0,01;0,99>>

Se sabemos mais, e.g., *cárie é também observada*, então

P(cárie|dordedente, *cárie*) = 1

OBS:

1) A crença menos específica permanece válida, mas pode ficar inútil.

2) A nova evidência pode ser inútil:

P(cárie|dordedente, Corinthians derrotado) = P(cárie|dordedente) = 0.8

NOTE A IMPORTÂNCIA DO CONHECIMENTO DO DOMÍNIO PARA QUALQUER PROCESSO DE INFERÊNCIA.

O Axioma Básico da Prob. condicional

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

- Ou:

$$\mathrm{P}(A,B) = \mathrm{P}(A \mid B)\,\mathrm{P}(B)$$

Corolário:

$$\mathrm{P}(A) = \sum_i \mathrm{P}(A, B_i)$$

$$\mathrm{P}(A) = \sum_i \mathrm{P}(A \mid B_i)\,\mathrm{P}(B_i)$$

Regra da Cadeia

Regra da Cadeia:

$$\mathrm{P}(E_1, E_2, \ldots E_n) = \mathrm{P}(E_n \mid E_{n-1}, \ldots E_2, E_1) \ldots \mathrm{P}(E_2 \mid E_1) P(E_1)$$

Prova:

$$\mathbf{P}(X_1, \ldots, X_n) = \mathbf{P}(X_1, \ldots, X_{n-1})\, \mathbf{P}(X_n \mid X_1, \ldots, X_{n-1})$$

$$= \mathbf{P}(X_1, \ldots, X_{n-2})\, \mathbf{P}(X_{n-1} \mid X_1, \ldots, X_{n-2})\, \mathbf{P}(X_n \mid X_1, \ldots, X_{n-1})$$

$$= \ldots$$

$$= \prod \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})$$

Inversão Bayesiana (Regra de Bayes)

$$P(H|e) = \frac{P(e|H)P(H)}{P(e)}$$

P(H|e): Probabilidade posterior

P(H): Probabilidade a priori

Por quê a fórmula é importante?

Muitas vezes P(e|H) é fácil de calcular, ao contrário de P(H|e)?

Exemplo.No cassino, um croupier fala 12! Ele jogou os dados ou estava comandando um jogo de roleta?

P*(12|dados)*, **P***(12|roleta)*: fácil de modelar. **P***(dados)*, **P***(roleta)*: fácil, basta ver número de mesas de dado ou roleta no cassino. **P***(dados|12)*, **P***(roleta|12)*: não é tão fácil estimar . . .

Cause and Effect

- We usually observe an effect and try to identify its cause

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

- However, it is usually easier to determine P(Effect| Cause) than P(Cause | Effect)

Another example: Meningitis

- Let's assume 0.8 of people with Meningitis present stiff neck (S), probability of Meningitis is 1 in 10000 and Stiff neck prob. is 0.1

For assessing diagnostic probability from causal probability:

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

E.g., let M be meningitis, S be stiff neck:

$$P(M|S) = \frac{P(S|M)P(M)}{P(S)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

Note: posterior probability of meningitis still very small!

Full joint distributions

A complete probability model specifies every entry in the joint distribution for all the variables $\mathbf{X} = X_1, \ldots, X_n$
I.e., a probability for each possible world $X_1 = x_1, \ldots, X_n = x_n$

E.g., suppose $Toothache$ and $Cavity$ are the random variables:

	$Toothache = true$	$Toothache = false$
$Cavity = true$	0.04	0.06
$Cavity = false$	0.01	0.89

Possible worlds are mutually exclusive $\Rightarrow$ $P(w_1 \wedge w_2) = 0$
Possible worlds are exhaustive $\Rightarrow$ $w_1 \vee \cdots \vee w_n$ is $True$
hence $\sum_i P(w_i) = 1$

Full joint distributions - 2

1) For any proposition ϕ defined on the random variables
$\phi(w_i)$ is true or false

2) ϕ is equivalent to the disjunction of w_is where $\phi(w_i)$ is true

Hence $P(\phi) = \Sigma_{\{w_i:\ \phi(w_i)\}} P(w_i)$

I.e., the unconditional probability of any proposition is computable as the sum of entries from the full joint distribution

Conditional probabilities can be computed in the same way as a ratio:

$$P(\phi|\xi) = \frac{P(\phi \wedge \xi)}{P(\xi)}$$

E.g.,

$$P(Cavity|Toothache) = \frac{P(Cavity \wedge Toothache)}{P(Toothache)} = \frac{0.04}{0.04 + 0.01} = 0.8$$

Calculating the probability of the evidence

- Suppose we wish to computer the probability of the observed evidence, let's say P(B=b) and A has possible values a_1, ...a_m . We can apply Bayes' rule for each value of A:

$$P(A=a_1|B=b) = P(B=b|A=a_1)P(A=a_1)/P(B=b)$$

...

$$P(A=a_m|B=b) = P(B=b|A=a_m)P(A=a_m)/P(B=b)$$

- Adding these up:

$$\Sigma_i P(A=a_i|B=b) = \Sigma_i P(B=b|A=a_i)P(A=a_i)/P(B=b)$$

- And noting that $\Sigma_i P(A=a_i|B=b) = 1$, then:

$$P(B=b) = \Sigma_i P(B=b|A=a_i)P(A=a_i)$$

Calculating the probability of the evidence - 2

- Since $P(B=b) = \sum_i P(B=b|A=a_i)P(A=a_i)$

- P(B=b) is a normalization factor regards i that we can denote α.
- In vectorial notation, we can write:

$$\mathbf{P}(A|B=b) = \alpha \mathbf{P}(B=b|A)\mathbf{P}(A)$$

Inference from Full joint distributions

Typically, we are interested in
the posterior joint distribution of the query variables $\mathbf{Y}$
given specific values $\mathbf{e}$ for the evidence variables $\mathbf{E}$

Let the hidden variables be $\mathbf{H} = \mathbf{X} - \mathbf{Y} - \mathbf{E}$

Then the required summation of joint entries is done by summing out the hidden variables:

$$\mathbf{P}(\mathbf{Y}|\mathbf{E}=\mathbf{e}) = \alpha\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e}) = \alpha\Sigma_{\mathbf{h}}\mathbf{P}(\mathbf{Y},\mathbf{E}=\mathbf{e},\mathbf{H}=\mathbf{h})$$

The terms in the summation are joint entries because $\mathbf{Y}$, $\mathbf{E}$, and $\mathbf{H}$ together exhaust the set of random variables

Obvious problems:
1) Worst-case time complexity $O(d^n)$ where d is the largest arity
2) Space complexity $O(d^n)$ to store the joint distribution
3) How to find the numbers for $O(d^n)$ entries???

d - number of possible elements of variable, n - number of variables

Inference from Full joint distributions - 2

- Inference from Full joint distributions could estimate any conditional probability even when involving hidden variables
- But, it would require a large amount of space to store it and even more data to build such full joint distribution

- Bayesian Network make it easier to build and store distributions

Introdução a Redes Bayesianas

Prof. Paulo André Castro
pauloac@ita.br
www.comp.ita.br/~pauloac
Sala 110,
IEC-ITA

Sumário

- Interpretação de Probabilidades
- Redes Bayesianas ou Redes de crença
- Inferência probabilística
- Aprendizado em método probabilísticos
- Métodos simplificados: Bayes ingênuo e Noisy-OR

Interpretations of Probabilities

There have been two main contending views about how to understand probability. One asserts that probabilities are fundamentally dispositional properties of non-deterministic physical systems, the classical such systems being gambling devices, such as dice. This view is particularly associated with **frequentism**,

Popper's observation (1959) that the frequency interpretation, precise though it was, fails to accommodate our intuition that probabilities of singular events exist and are meaningful.

- Do we need to toss a coin infinity (or many times) to make statements about the probability of it landing head in one specific toss?

- The alternative view of probability is to think of probabilities as reporting our subjective **degrees of belief**. This view was expressed by Thomas Bayes (1958) and Pierre Simon de Laplace (1951)

Principal Principle and Conditionalization

Principal Principle whenever you learn that the physical probability of an outcome is r, set your subjective probability for that outcome to r. This is really just common sense: you may think that the probability of a friend shaving his head is 0.01, but if you learn that he will do so if and only if a fair coin yet to be flipped lands heads, you'll revise your opinion accordingly.

Definition Conditionalization *After applying Bayes' theorem to obtain* $P(h|e)$ *adopt that as your posterior degree of belief in* h — *or,* $Bel(h) = P(h|e)$.

Rede Bayesiana ou Rede de Crença (Belief Network)

A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

Syntax:

- a set of nodes, one per variable
- a directed, acyclic graph (link $\approx$ "directly influences")
- a conditional distribution for each node given its parents:
  $\mathbf{P}(X_i|Parents(X_i))$

In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

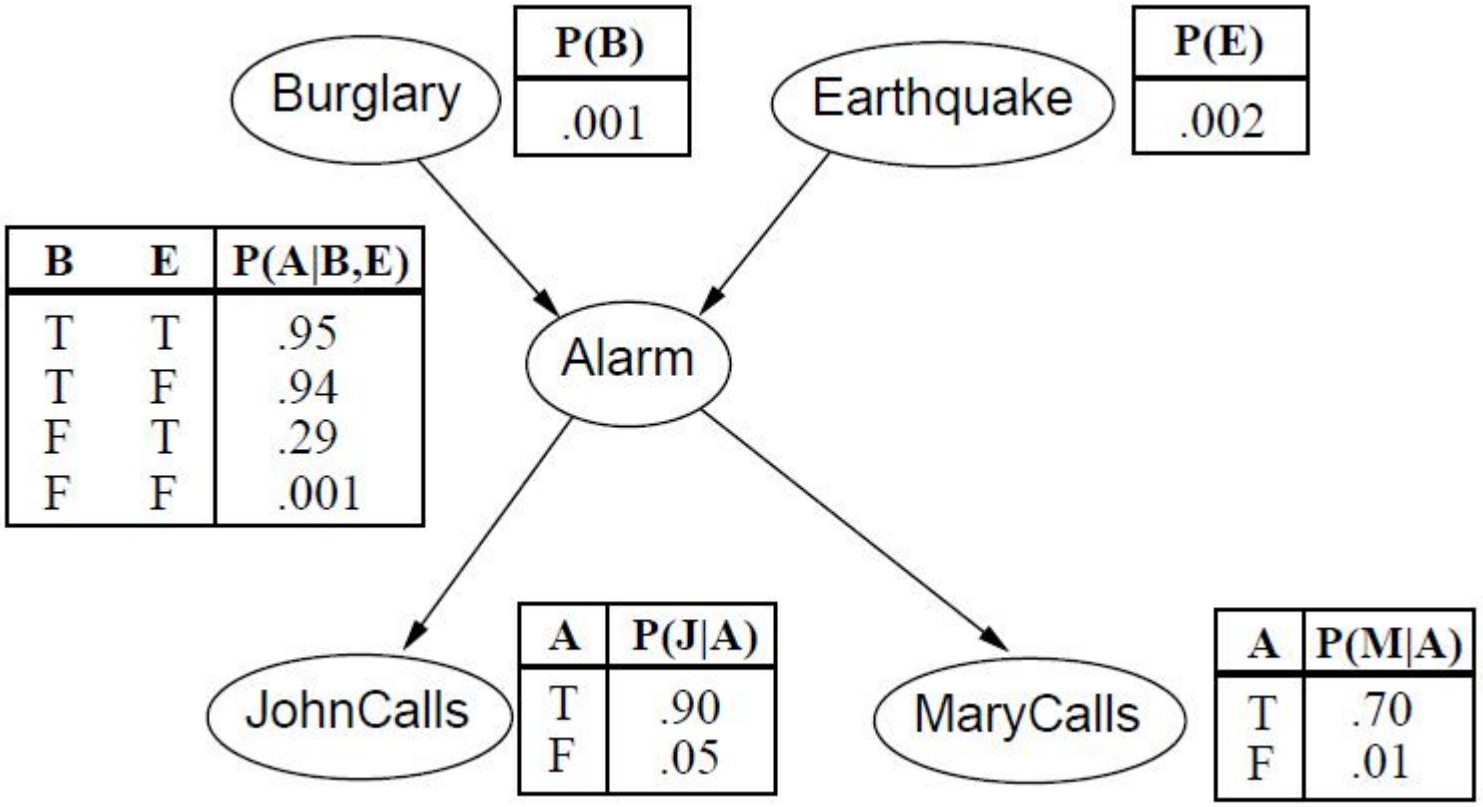
Example: Is it an Earthquake or burglar?

I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

Variables: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$
Network topology reflects "causal" knowledge:

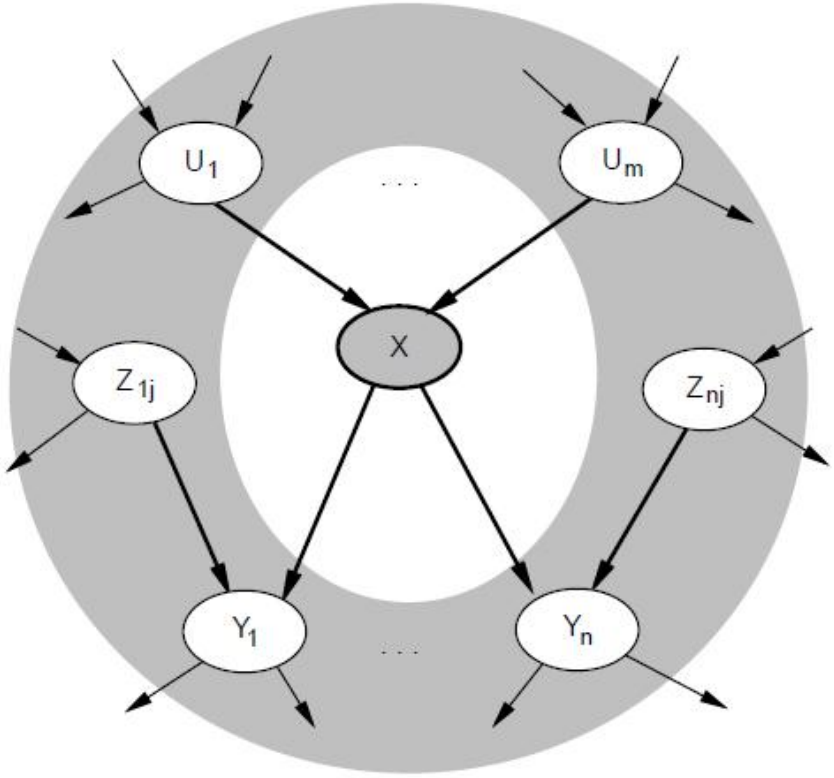
- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call

Example - 2


Burglary
P(B)
.001
Earthquake
P(E)
.002
B E P(A|B,E)
T T .95
T F .94
F T .29
F F .001
Alarm
JohnCalls
A P(J|A)
T .90
F .05
MaryCalls
A P(M|A)
T .70
F .01

Markov Blanket (Cobertor de Markov)

Each node is conditionally independent of all others given its Markov blanket: parents + children + children's parents



Método para construção de uma rede

Need a method such that a series of locally testable assertions of conditional independence guarantees the required global semantics

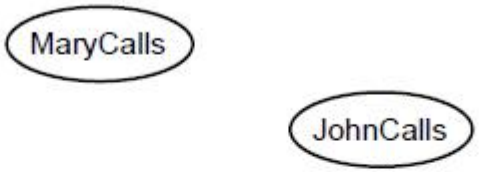
1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1$ to n
   add X_i to the network
   select parents from $X_1, \ldots, X_{i-1}$ such that
   $$\mathbf{P}(X_i|Parents(X_i)) = \mathbf{P}(X_i|X_1, \ldots, X_{i-1})$$

This choice of parents guarantees the global semantics:

$$\begin{aligned} \mathbf{P}(X_1, \ldots, X_n) &= \prod_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1}) \quad \text{(chain rule)} \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i|Parents(X_i)) \quad \text{(by construction)} \end{aligned}$$

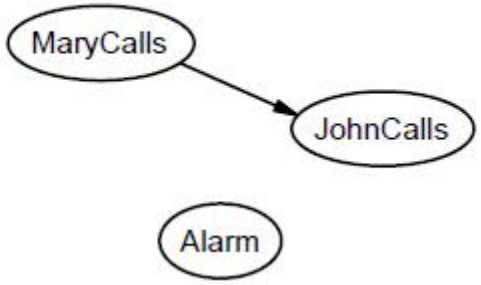
Exemplo

Suppose we choose the ordering M, J, A, B, E



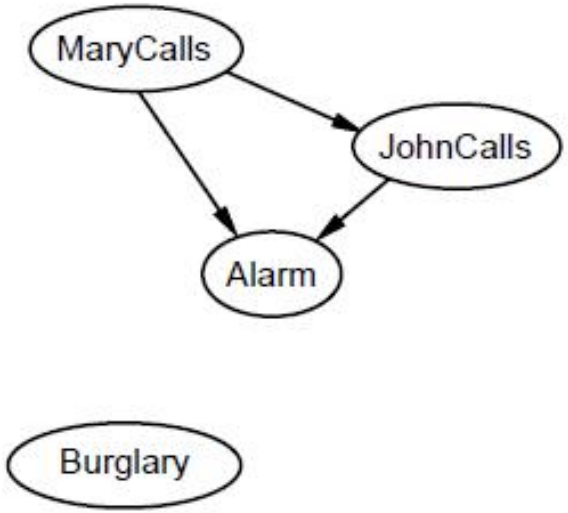
$P(J|M) = P(J)$?

Suppose we choose the ordering M, J, A, B, E



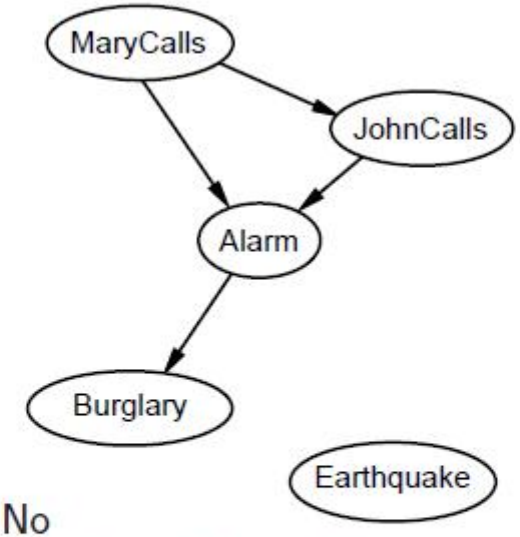
$P(J|M) = P(J)$? No
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$?

Suppose we choose the ordering M, J, A, B, E



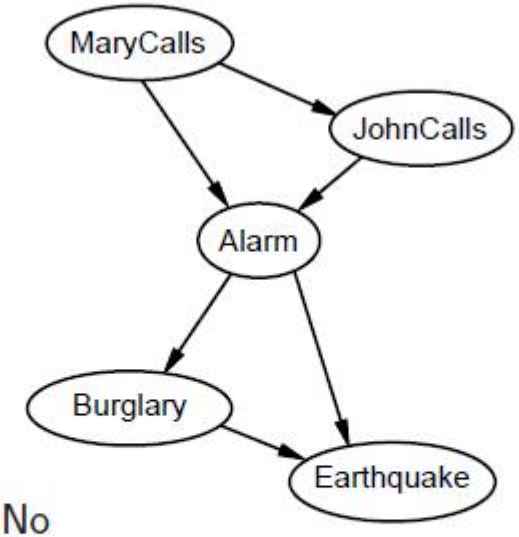
$P(J|M) = P(J)$? No
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? No
$P(B|A, J, M) = P(B|A)$?
$P(B|A, J, M) = P(B)$?

Suppose we choose the ordering M, J, A, B, E

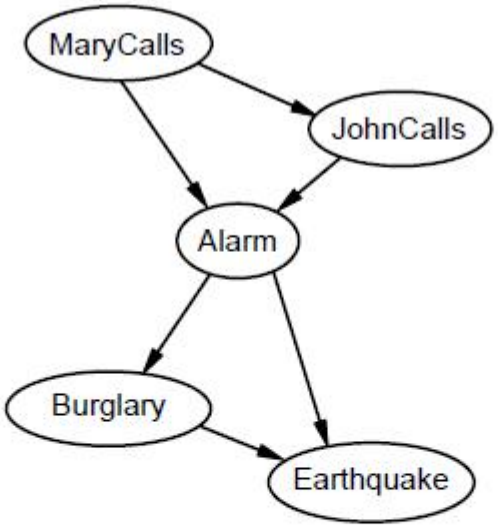


$P(J|M) = P(J)$? No
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? No
$P(B|A, J, M) = P(B|A)$? Yes
$P(B|A, J, M) = P(B)$? No
$P(E|B, A, J, M) = P(E|A)$?
$P(E|B, A, J, M) = P(E|A, B)$?

Suppose we choose the ordering M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? No
$P(B|A, J, M) = P(B|A)$? Yes
$P(B|A, J, M) = P(B)$? No
$P(E|B, A, J, M) = P(E|A)$? No
$P(E|B, A, J, M) = P(E|A, B)$? Yes



Deciding conditional independence is hard in noncausal directions

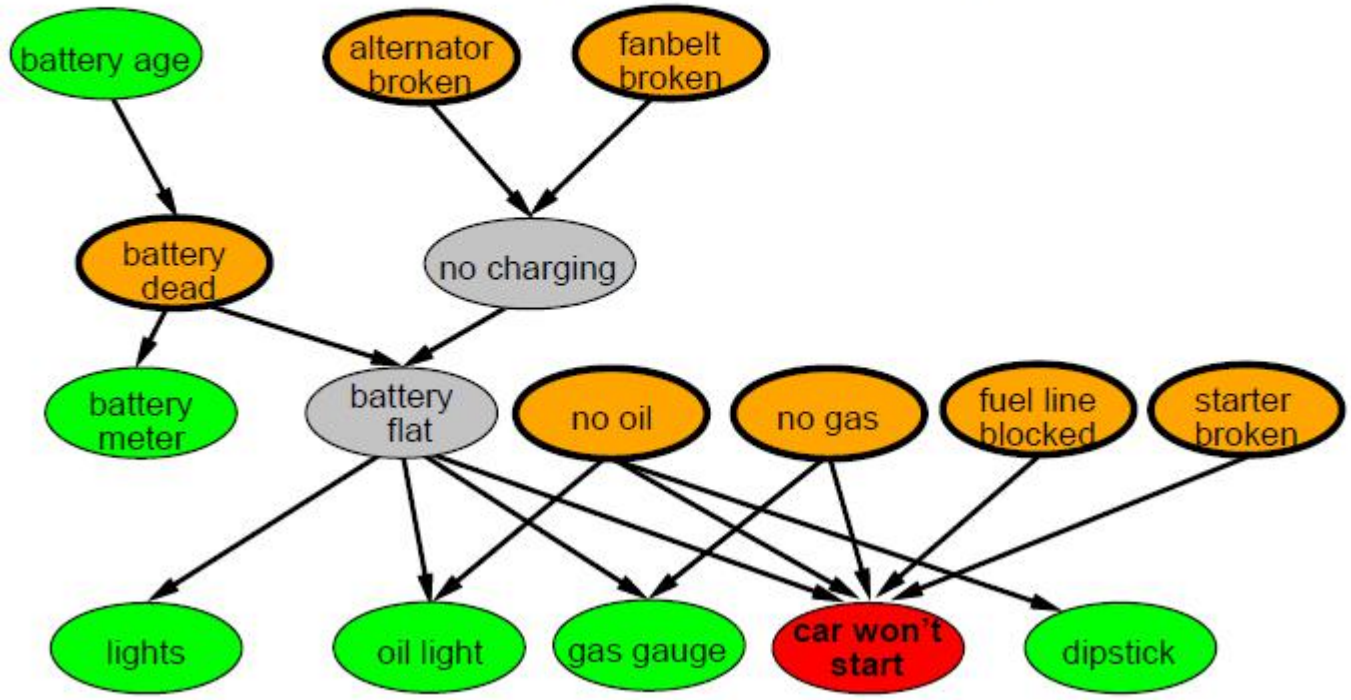
(Causal models and conditional independence seem hardwired for humans!)

Assessing conditional probabilities is hard in noncausal directions

Network is less compact: $1+2+4+2+4=13$ numbers needed

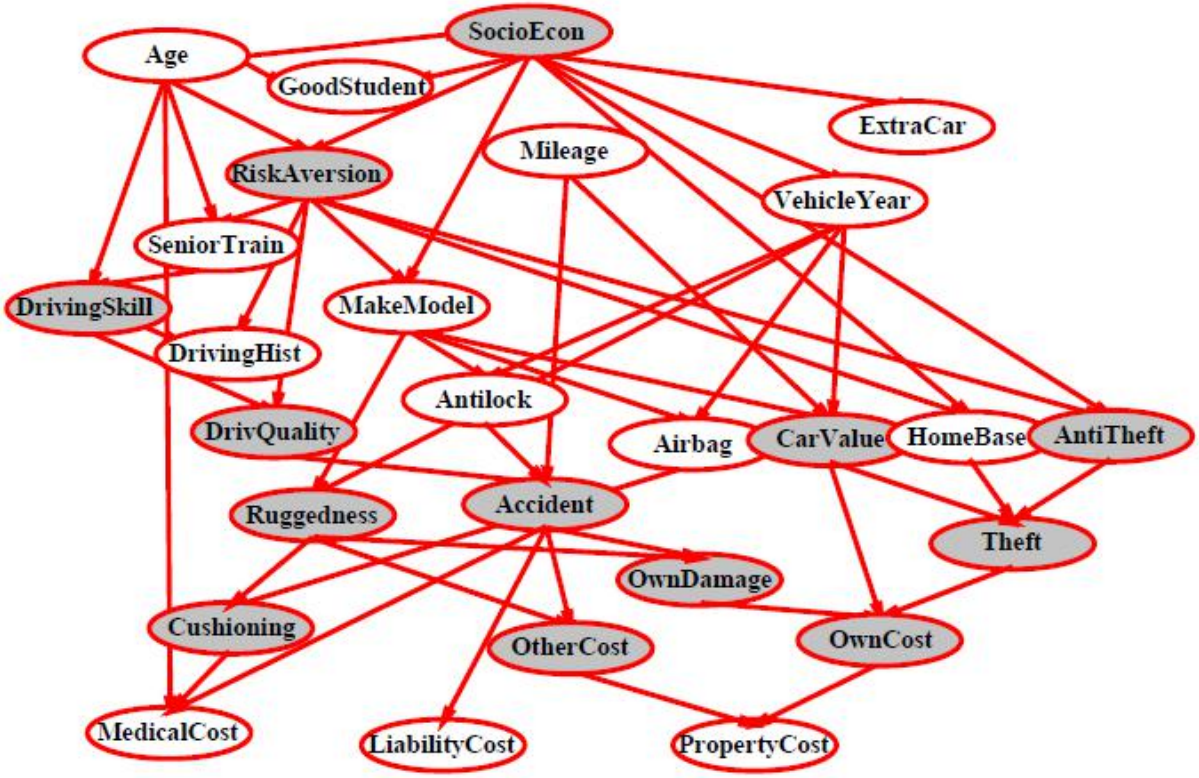
Outro Exemplo: Conserto de Carro

Initial evidence: car won't start
Testable variables (green), "broken, so fix it" variables (orange)
Hidden variables (gray) ensure sparse structure, reduce parameters



Exemplo: Seguro de Carro

- Problema: Estimar custos (Medical, Liability, Property) dados as informações do segurado e outras disponíveis por outras fontes (em Cinza)



I-map and D-map and Perfect Map

- I-map: All direct dependencies in the system being modeled are explicitly shown via arcs. (Independence Map or I-map for short).

- D-map: If every arc in a BN happens to correspond to a direct dependence in the system, then the BN is said to be a Dependence-map (or, D-map for short).

- A BN which is both an I-map and a D-map is said to be a perfect map.

Sumário

- Redes Bayesianas ou Redes de crença
- Inferência probabilística
- Aprendizado em método probabilísticos
- Métodos simplificados: Bayes ingênuo e Noisy-OR

Inferência em Redes Bayesianas

- Dada uma rede, devemos ser capaz de inferir a partir dela isto é :

- Busca responder questões simples, P(X| E=e)
  - Ex. : $P(NoGas|Gauge=empty, Lights=on, Starts=false)$
- Ou questões conjuntivas: P(X_i , X_j | E=e)
  - Usando o fato:

$$\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})$$

- A inferência pode ser feita a partir da distribuição conjunta total ou por **enumeração**

Inferência com Distribuição Conjunta Total: Exemplo

Por exemplo para saber

P(A|b) temos

P(A|b)= P(A, b)/P(b)=

<P(a, b)/P(b);P(¬a , b)/P(b) > =

A	B	C	P(A,B,C)
F	F	F	P(A=F,B=F,C=F)
F	F	T	P(A=F,B=F,C=T)
F	T	F	..
F	T	T	.
T	F	F	..
T	F	T	...
T	T	F	....
T	T	T	....
			.....
			P(A=T,B=T,C=T)

$= \alpha < P(a, b); P(\neg a , b) >$

$= \alpha [<P(a,b,c)+P(a,b,\neg c); P(\neg a,b,c)+P(\neg a,b, \neg c)>]$

Observe que α pode ser visto como um fator de normalização para o vetor resultante da distribuição de probabilidade, pedida P(A|b). Assim pode-se evitar seu cálculo, Simplesmente normalizando <P(a,b); P(¬a , b) >

Inferência em Redes Bayesianas

Simple queries: compute posterior marginal $\mathbf{P}(X_i|\mathbf{E}=\mathbf{e})$
e.g., $P(NoGas|Gauge=empty, Lights=on, Starts=false)$

Conjunctive queries: $\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})$

Optimal decisions: decision networks include utility information;
probabilistic inference required for $P(outcome|action, evidence)$

Value of information: which evidence to seek next?

Sensitivity analysis: which probability values are most critical?

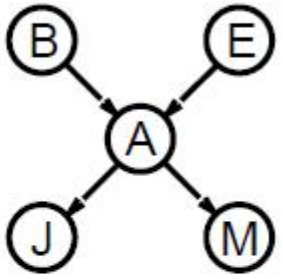
Explanation: why do I need a new starter motor?

Inferência por Enumeração

Slightly intelligent way to sum out variables from the joint without actually constructing its explicit representation

Simple query on the burglary network:

$$\begin{aligned}&\mathbf{P}(B|j,m)\\&=\mathbf{P}(B,j,m)/P(j,m)\\&=\alpha\mathbf{P}(B,j,m)\\&=\alpha\textstyle\sum_e\sum_a\mathbf{P}(B,e,a,j,m)\end{aligned}$$



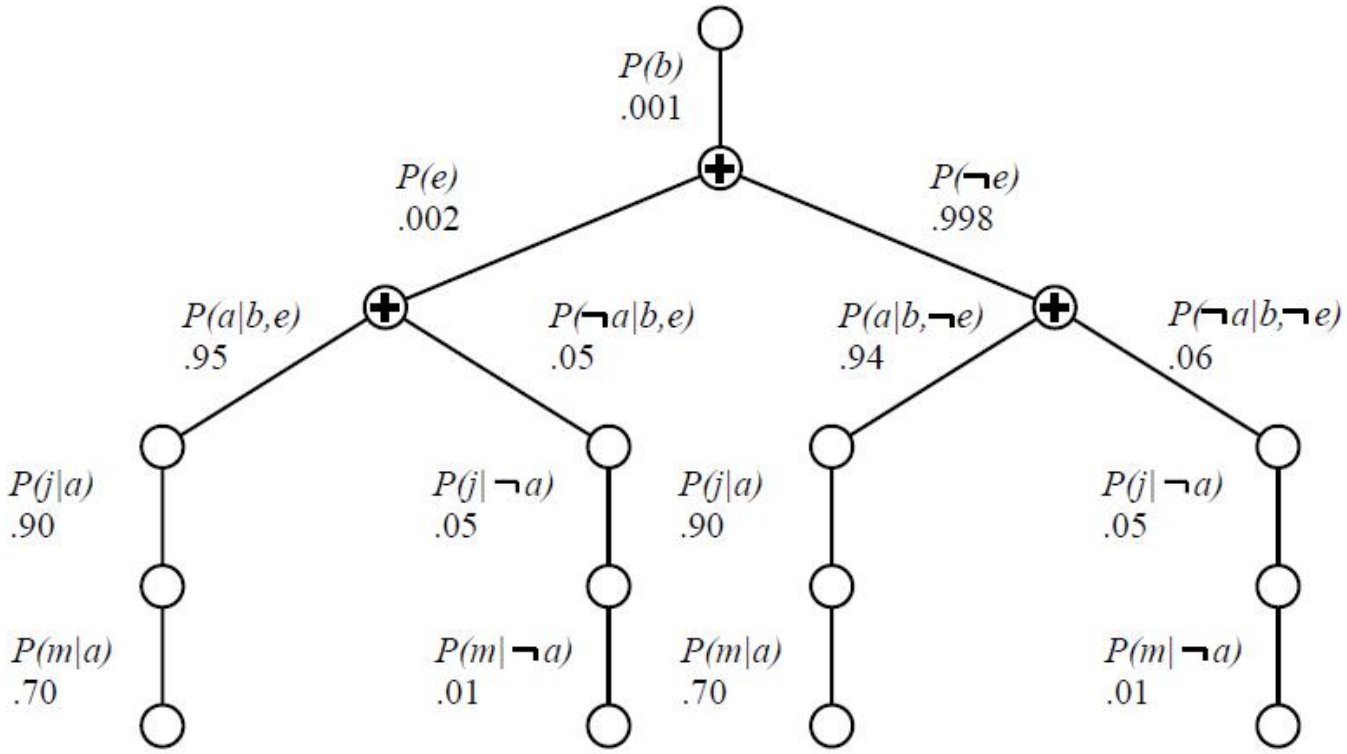
Rewrite full joint entries using product of CPT entries:

$$\begin{aligned}&\mathbf{P}(B|j,m)\\&=\alpha\textstyle\sum_e\sum_a\mathbf{P}(B)P(e)\mathbf{P}(a|B,e)P(j|a)P(m|a)\\&=\alpha\mathbf{P}(B)\textstyle\sum_e P(e)\sum_a\mathbf{P}(a|B,e)P(j|a)P(m|a)\end{aligned}$$

Recursive depth-first enumeration: $O(n)$ space, $O(d^n)$ time

Inferência por Enumeração - 2

Enumeration is inefficient: repeated computation
e.g., computes $P(j|a)P(m|a)$ for each value of e



- Pode ser melhorada através do armazenamento dos valores já calculados (Programação Dinâmica)

Calculando P(b) não normalizado

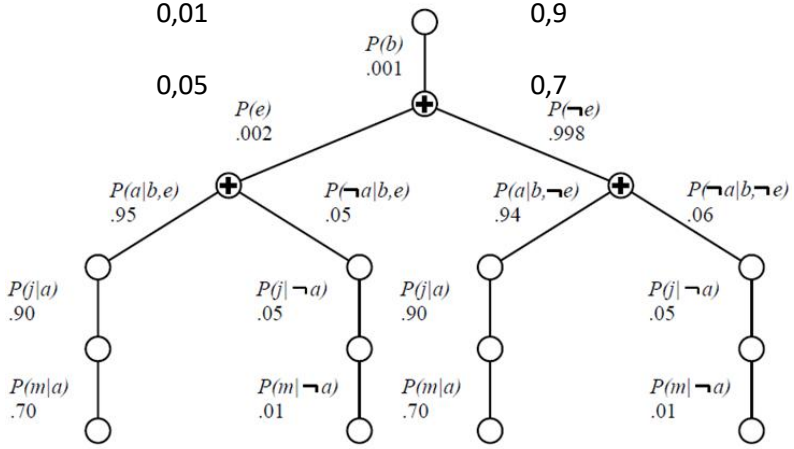
"P(b) nao normalizado"

0,0005922

0,001

\+ 0,5922426

		0,001197				0,591046	
*		0,002	*			0,998	
+		0,598525	+			0,59223	
	0,5985	0,000025		0,5922		0,00003	Produtorio
	0,95	0,05		0,94		0,06	
	0,9	0,01		0,9		0,01	
	0,7	0,05		0,7		0,05	



Calculando P(não b) não normalizado

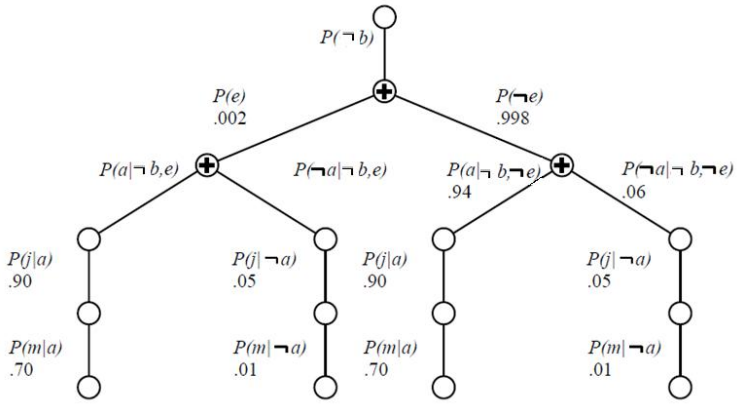
"P(nao b) nao normalizado"

0,001492

0,999

\+ 0,001493

		0,000366				0,001127	
*		0,002	*			0,998	
+		0,183055	+			0,00113	
	0,1827	0,000355		0,00063		0,0005	
	0,29	0,71		0,001		0,999	
	0,9	0,01		0,9		0,01	
	0,7	0,05		0,7		0,05	



Algoritmo de Enumeração

function ENUMERATION-ASK(X, **e**, bn) **returns** a distribution over X
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayes net with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$ /* **Y** = *hidden variables* */

  $\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
  for each value x_i of X **do**
    $\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(bn.VARS, $\mathbf{e}_{x_i}$)
      where $\mathbf{e}_{x_i}$ is **e** extended with $X = x_i$
  return NORMALIZE($\mathbf{Q}(X)$)

---

function ENUMERATE-ALL($vars$, **e**) **returns** a real number
  if EMPTY?($vars$) **then return** 1.0
  $Y \leftarrow$ FIRST($vars$)
  if Y has value y in **e**
    then return $P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), **e**)
    else return $\sum_y P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_y$)
      where $\mathbf{e}_y$ is **e** extended with $Y = y$

Inferência por Enumeração

- Algoritmo de Enumeração permite determinar uma distribuição de probabilidade condicional
- P(variável de saída| evidências conhecidas)

- Também é possível responder perguntas conjuntivas usando o fato:

$$\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})$$

- Demonstração?….

Demonstração

$$\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})$$

como:

$$\mathrm{P}(A, B) = \mathrm{P}(A \mid B)\,\mathrm{P}(B)$$

$$\mathbf{P}(X_i, X_j|\mathbf{E}=\mathbf{e}) = \frac{\mathbf{P}(X_i, X_j, \mathbf{E}=\mathbf{e})}{\mathbf{P}(\mathbf{E}=\mathbf{e})} =$$

$$\frac{\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})\,\mathbf{P}(X_i, \mathbf{E}=\mathbf{e})}{\mathbf{P}(\mathbf{E}=\mathbf{e})} =$$

$$\mathbf{P}(X_j|X_i, \mathbf{E}=\mathbf{e})\ \mathbf{P}(X_i|\mathbf{E}=\mathbf{e})$$

Inferência por Enumeração

- Como observado, a enumeração tende a recalcular várias vezes alguns valores

- Pode-se eliminar parte do retrabalho através da técnica de programação dinâmica (eliminação de variável)… Basicamente, os valores já calculados são armazenados em uma tabela e selecionados quando novamente necessários…(mais informações Russel, cap. 14)

Inferência por Eliminação de Variável

Enumeration is inefficient: repeated computation

e.g., computes $P(J=true|a)P(M=true|a)$ for each value of e

Variable elimination: carry out summations right-to-left, storing intermediate results (factors) to avoid recomputation

$$
\begin{aligned}
&\mathbf{P}(B|J=true, M=true) \\
&= \alpha \underbrace{\mathbf{P}(B)}_{B} \Sigma_e \underbrace{P(e)}_{E} \Sigma_a \underbrace{\mathbf{P}(a|B,e)}_{A} \underbrace{P(J=true|a)}_{J} \underbrace{P(M=true|a)}_{M} \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a \mathbf{P}(a|B,e) P(J=true|a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a \mathbf{P}(a|B,e) f_J(a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) \Sigma_a f_A(a,b,e) f_J(a) f_M(a) \\
&= \alpha \mathbf{P}(B) \Sigma_e P(e) f_{\bar{A}JM}(b,e) \text{ (sum out } A\text{)} \\
&= \alpha \mathbf{P}(B) f_{\bar{E}\bar{A}JM}(b) \text{ (sum out } E\text{)} \\
&= \alpha f_B(b) \times f_{\bar{E}\bar{A}JM}(b)
\end{aligned}
$$

Operações Básicas da Eliminação de Variáveis

Pointwise product of factors f_1 and f_2:

$$f_1(x_1, \ldots, x_j, y_1, \ldots, y_k) \times f_2(y_1, \ldots, y_k, z_1, \ldots, z_l) = f(x_1, \ldots, x_j, y_1, \ldots, y_k, z_1, \ldots, z_l)$$

E.g., $f_1(a, b) \times f_2(b, c) = f(a, b, c)$

Summing out a variable from a product of factors: move any constant factors outside the summation:

$$\sum_x f_1 \times \cdots \times f_k = f_1 \times \cdots \times f_i \sum_x f_{i+1} \times \cdots \times f_k = f_1 \times \cdots \times f_i \times f_{\bar{X}}$$

assuming $f_1, \ldots, f_i$ do not depend on X

Sumário

- Redes Bayesianas ou Redes de crença
- Inferência probabilística
- Aprendizado em método probabilísticos
- Métodos simplificados: Bayes ingênuo e Noisy-OR

Aprendizado em modelos probabilísticos

- Aprender em redes bayesianas é o processo de determinar a topologia da rede (isto é, seu grafo direcionado) e as tabelas de probabilidade condicional

- Problemas?
  - Como determinar a topologia?
  - Como estimar as probabilidades ?
  - Quão complexas são essas tarefas?
    - Isto é quantas topologias e quantas probabilidades precisariam ser determinadas….

Tamanho das Tabelas de Probabilidade Condicional e Distribuição Conjunta Total

- Vamos supor que cada variável é influenciada por no máximo k outras variáveis (Naturalmente, k<n=total de variáveis).

- Supondo variáveis booleanas, cada tabela de probabilidade condicional (CPT) terá no máximo 2^k entradas (ou probabilidades). Logo ao total haverá no máximo n* 2^k entradas

- Enquanto, na distribuição conjunta Total haverá 2^n entradas. Por exemplo, para n=30 com no máximo cinco pais (k=5) isto significa 960 ao invés de mais um bilhão (2^{30})

A	B	C	P(A,B,C)
F	F	F	P(A=F,B=F,C=F)
F	F	T	P(A=F,B=F,C=T)
F	T	F	..
F	T	T	.
T	F	F	..
T	F	T	...
T	T	F	....
T	T	T	....
			.....
			P(A=T,B=T,C=T)

Número de "entradas" da Distribuição Conjunta e na Rede Bayesiana - 2

- Em domínios onde **cada variável** pode ser diretemante **influenciada** por **todas** as outras, tem-se a rede totalmente conectada e assim exige-se a **quantidade de entradas da mesma ordem** da distribuição conjunta total

- Porém se essa dependência for tênue, **pode não valer** a pena a complexidade adicional na rede em relação ao pequeno ganho em exatidão

- Via de regra, se nos fixarmos em um **modelo causal** acabaremos tendo de especificar uma quantidade menor de números, e os números frequentemente serão mais fáceis de calcular. (Russel,Norvig, 2013, pg. 453)

- **Modelos causais** são aqueles onde se especifica no sentido causa efeito, isto é P(efeito|causa) ao invés de P(causa|efeito), oque geralmente é necessário para diagnóstico

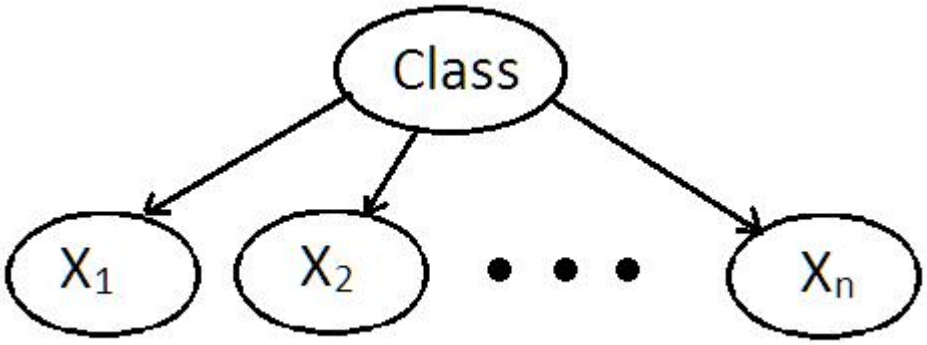
Simplificando a representação tabelas de probabilidade condicional (CPT)

- Vimos que que o número de entradas de uma CPT cresce exponencialmente
  - Para o caso binário e K pais, a CPT de um nó terá 2^k probabilidades a serem calculadas

- Vejamos duas abordagens para simplificar a rede através da adoção de hipóteses simplificadoras
  - Bayes Ingênuo e
  - OU-ruidoso

Naïve Bayes (Bayes Ingênuo)

- Uma classe particular e simples de redes bayesianas é chamada de Bayes Ingênuo (Naïve Bayes)
- Ela é simples por supor independência condicional entre todas as variáveis X dada a variável Class
- As vezes, chamado também de classificador Bayes, por ser frequentemente usado como abordagem inicial para classificação



Naïve Bayes (Bayes Ingênuo) - 2

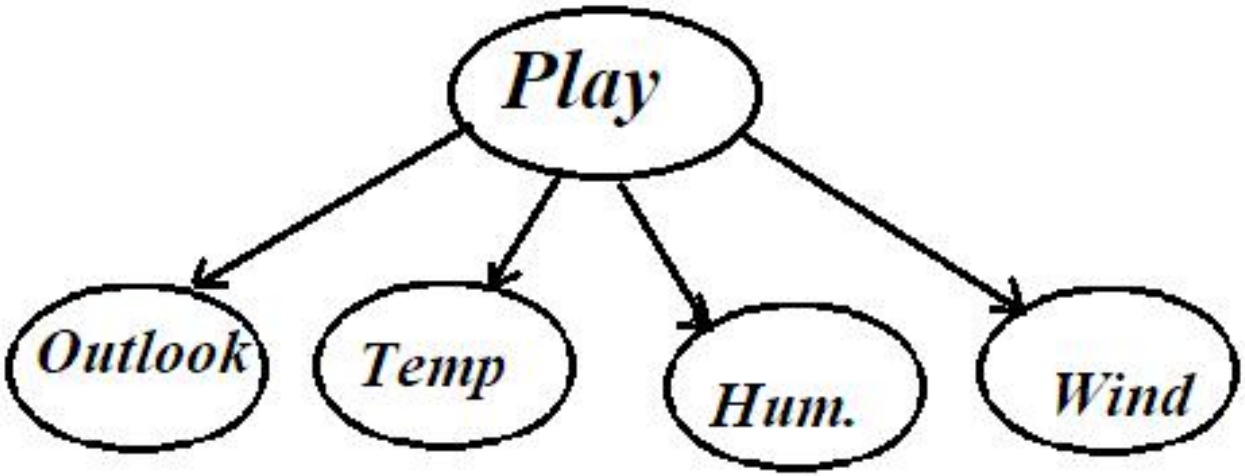
- A topologia simples traz a vantagem da representação concisa da Distribuição Conjunta Total.
- Como todo os nós tem no máximo um pai, cada CPT de no X tem apenas duas entradas e uma entrada no nó classe. Logo, (2n-1) entradas para toda a rede. Naïve Bayes é **linear** em relação ao número de nós (n) !!!!
- "Na prática, sistemas de Bayes ingênuos podem funcionar surpreendentemente bem…." . pg. 438

Exemplo de Naïve Bayes

- Vamos retomar o exemplo do jogo de tênis

Ex	Céu	Temperatura	Umidade	Vento	JogarTênis
X1	Ensolarado	Quente	Alta	Fraco	NÃO
X2	Ensolarado	Quente	Alta	Forte	NÃO
X3	Nublado	Quente	Alta	Fraco	SIM
X4	Chuvoso	Boa	Alta	Fraco	SIM
X5	Chuvoso	Fria	Normal	Fraco	SIM
X6	Chuvoso	Fria	Normal	Forte	NÃO
X7	Nublado	Fria	Normal	Forte	SIM
X8	Ensolarado	Boa	Alta	Fraco	NÃO
X9	Ensolarado	Fria	Normal	Fraco	SIM
X10	Chuvoso	Boa	Normal	Fraco	SIM
X11	Ensolarado	Boa	Normal	Forte	SIM
X12	Nublado	Boa	Alta	Forte	SIM
X13	Nublado	Quente	Normal	Fraco	SIM
X14	Chuvoso	Boa	Alta	Forte	NÃO

Usando a abordagem Bayes ingênuo



- Problema a resolver:

Outlook	Temperature	Humidity	Windy	Play
sunny	cool	high	true	?

Solução:

- P(Play|Outlook,Temp,Hum,Wind)=
- P(Outlook,Temp,Hum,Wind|Play)P(Play)/P(Outlook, Temp,Hum,Wind)=
- Regra da cadeia e indepêndencia:
- P(Outlook|Play)P(Temp|Play)P(Hum|Play)P(Wind|Play)P(Play)/ P(Outlook,Temp,Hum,Wind)

- O método de inferência por enumeração já visto é aplicável!!!
- Estima-se as probabilidades pelo conjunto de treinamento

Contagens e probabilides estimadas pelo conjunto de treinamento

Outlook			Temperature			Humidity			Windy			Play	
	yes	*no*		*yes*	*no*		*yes*	*no*		*yes*	*no*	*yes*	*no*
sunny	2	3	hot	2	2	high	3	4	false	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3		
rainy	3	2	cool	3	1								
~P(Outlook|Play)													
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	false	6/9	2/5	9/14	5/14
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	true	3/9	3/5		
rainy	3/9	2/5	cool	3/9	1/5								

- P(Play=s|Outlook=sunny,Temp=cool,Hum=high,Wind=true)=

- P(sunny|play)P(cool|play)P(high|play)P(true|play)P(Play) /P(evidencia) = 2/9*3/9*3/9*3/9*9/14 / P(e) =0.0053/P(e)

Solução 3 – continuação

- Da mesma forma,
- P(sunny|play)P(cool|play)P(high|play)P(true|play)P(Play)/P(e) = 3/5*1/5*4/5*3/5*5/14/P(e) =0.0206/P(e)
- Mas P(H,e) e P(not H,e) tem que somar 1, assim:

$$\text{Probability of } yes = \frac{0.0053}{0.0053+0.0206} = 20.5\%,$$

$$\text{Probability of } no = \frac{0.0206}{0.0053+0.0206} = 79.5\%.$$

Estimativas de Probabilides

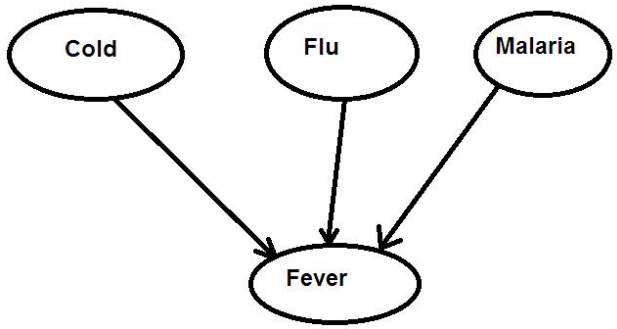
- Qual a estimativa da probabilidade P(Outlook=overcast|Play=no)?

Outlook			Temperature			Humidity			Windy		Play	
	yes	*no*		*yes*	*no*		*yes*	*no*		*yes*	*no*	*yes*	*no*
sunny	2	3	hot	2	2	high	3	4	false	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3		
rainy	3	2	cool	3	1								
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	false	6/9	2/5	9/14	5/14
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	true	3/9	3/5		
rainy	3/9	2/5	cool	3/9	1/5								

- Zero! Isto é razoável? Como resolver?
- Uma Solução: estimador de Laplace (Laplace smoothing). Seja V o número de valores possíveis para A, estima-se P(A|B) :
- P(A=a|B=b) = [N(A=a,B=b)+1]/[N(B=b)+V]

Criando Distribuições Condicionais Conjuntas Compactadas....

- Alguns problemas podem ser modelados com uma abordagem do tipo Noisy-OR (ou ruidoso). A técnica parte de duas hipóteses:
  - Todas as causas de uma variável ser acionada estão listadas (pode-se adicionar uma causa geral "outros")
    - Isto é, P (Fever | F,F,F) = 0
  - Há independência condicionais entre oque causa a "falha" da variável pai acionar a variável filho (efeito). Exemplo: o que impede a gripe de causar febre em alguém é independente do que impede o resfriado de causar febre.
    - Isto é, P (not Fever| Cold,Flu,Malaria) = P(not Fever|Cold)P(not Fever| Flu)P(not Fever | Malaria)
- Exemplo:
- P(Not fever |malaria) =0.1
- P(Not fever| flu) =0.2
- P(Not fever| cold)=0.6



Noisy -OR

Noisy-OR distributions model multiple noninteracting causes

1) Parents $U_1 \ldots U_k$ include all causes (can add leak node)
2) Independent failure probability q_i for each cause alone

Cold	*Flu*	*Malaria*	$P(Fever)$	$P(\neg Fever)$
F	F	F	**0.0**	
F	F	T		**0.1**
F	T	F		**0.2**
F	T	T		
T	F	F		**0.6**
T	F	T		
T	T	F		
T	T	T		

Number of parameters **linear** in number of parents

- $P(X \mid u_1, \cdots u_j, \neg u_{j+1}, \ldots \neg u_k) = \langle 1 - \prod^j_{i=1} q_i ; \prod^j_{i=1} q_i \rangle$
- q_i is the probability of cause i fails !!

Noisy -OR

Noisy-OR distributions model multiple noninteracting causes

1) Parents $U_1 \ldots U_k$ include all causes (can add leak node)
2) Independent failure probability q_i for each cause alone

Cold	*Flu*	*Malaria*	$P(Fever)$	$P(\neg Fever)$
F	F	F	**0.0**	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Number of parameters **linear** in number of parents

- $P(X \mid u_1, \cdots u_j, \neg u_{j+1}, \ldots \neg u_k) = \langle 1 - \prod^j_{i=1} q_i ; \prod^j_{i=1} q_i \rangle$
- q_i is the probability of cause i fails !!