Redes Bayesianas

Prof. Júlio Cesar Nievola
PPGIa – Escola Politécnica
PUCPR

Raciocínio com Incerteza

- Incerteza: qualidade ou estado de não ser conhecido com certeza

- Fontes de incerteza:
  - Ignorância: qual o lado da moeda está para cima (na mão)?
  - Complexidade: ex. meteorologia
  - Aleatoriedade física: qual lado da moeda vai estar para cima quando cair?

Rev. Thomas Bayes (1702-1761)

Introdução

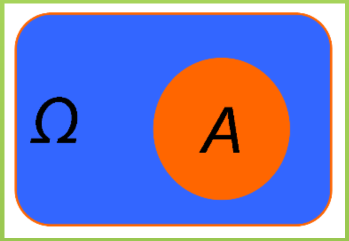
- As Redes Bayesianas são aplicadas em casos de incerteza quando se sabem certas probabilidades (condicionais) e se buscam probabilidades desconhecidas para situações específicas.
- Aplicações: bioinformática e medicina, engenharia, classificação de documentos, processamento de imagens, fusão de dados, sistemas de suporte à decisão, etc.
- Exemplos:
  - Inferência: P(Diagnóstico|Sintoma)
  - Detecção de anomalias: A observação é anômala?
  - Coleção ativa de dados: Qual é o próximo teste diagnóstico dado um conjunto de observações?

Variáveis Aleatórias Discretas

- Considere que A indica uma variável aleatória binária.
- Vamos considerar que A indica um evento, o qual não se tem certeza se acontece (1) ou não acontece (0).
- Exemplos:
  - O lançamento de uma moeda resultar em cara;
  - O paciente apresenta câncer.

Intuição sobre probabilidade

- Intuitivamente, a probabilidade do evento A é igual à proporção dos resultados em que é A verdadeiro.



- Ω é o conjunto de todos os resultados possíveis.
- Sua área é $P(\Omega) = 1$.
- A área em laranja corresponde aos resultados em que A é verdadeiro.
- $P(A)$ = área do círculo laranja. Obviamente, $0 \leq P(A) \leq 1$.

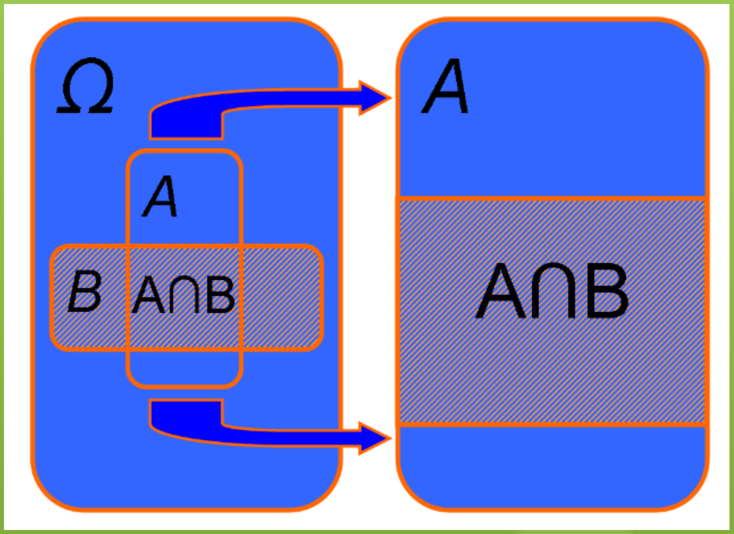
Axiomas de Kolmogorov e Teoremas

- Kolmogorov1: $P(A) \geq 0, \forall A \subseteq \Omega$
- Kolmogorov2: $P(\Omega) = 1$
- Kolmogorov3: $P(\bigcup_i A_i) = \sum_i P(A_i)$, em que A1, A2, … é qualquer sequência contável de eventos disjuntos

- $P(\bar{A}) = 1 - P(A)$, na qual $\bar{A} = \Omega \setminus A$
- $P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A) = P(A \cap B) + P(A \cap \bar{B})$

Probabilidade Condicional

- $P(A \cap B)$ = proporção do espaço em que A é verdadeiro e que também é B verdadeiro
- Definição formal: $P(B|A) = \frac{P(A \cap B)}{P(A)}$



Exemplo de Probabilidade Condicional

- Escolhe-se uma carta de um baralho com 52 cartas.
- A = a carta é uma figura. P(A) = 12/52 = 3/13.
- B = a carta é dama. P(B) = 4/52 = 1/13.
  - $P(B \cap A) = P(B) = 1/13$
- Aplicando a definição tem-se o resultado intuitivo
  - $P(B|A) = \frac{{}^{1}/_{13}}{{}^{3}/_{13}} = {}^{1}/_{3}$
  - $P(A|B) = \frac{{}^{1}/_{13}}{{}^{1}/_{13}} = 1$
- C = é carta de espadas. P(C) = ¼
- Deve-se observar que $P(C|A) = P(C) = 1/4$, ou seja, o evento C é independente de A.

Eventos Independentes

- Definição: Dois eventos A e B são independentes se e somente se $P(A \cap B) = P(A)P(B)$. A independência de A e B implica:
  - $P(A|B) = P(A),\ se\ P(B) \neq 0$
  - $P(B|A) = P(B),\ se\ P(A) \neq 0$

- Independência Condicional: Dois eventos são condicionalmente independentes dado C se e somente se $P(A \cap B|C) = P(A|C)P(B|C)$. Notação: $I(A, B|C)$.
  - $P(A|B, C) = P(A|C), se\ P(B|C) \neq 0$
  - $P(B|A, C) = P(B|C), se\ P(A|C) \neq 0$

Terminologia do Teorema de Bayes

- Seja **A** uma amostra de dados ("*evidência*"): classe desconhecida.
- Seja H a *hipótese* de que X pertence à classe C.
- Classificação é determinar P(H|**A**), (*probabilidade a posteriori),* a probabilidade de que a hipótese seja verdadeira para os dados observados **A.**
- P(H) (*probabilidade a priori*), a probabilidade inicial:
  - E.g., **A** comprará um computador, independente da idade, salário, …
- P(**A**): probabilidade que os dados sejam observados.
- P(**A**|H) (*verossimilhança*), a probabilidade de observar os dados **X**, dado que a hipótese é verdadeira:
  - E.g., dado que **A** comprará um computador, a probabilidade de que A esteja na faixa 31..40 de renda.

Regra de Bayes

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

- De outra forma:

$$Posterior = \frac{Verossimilhança\ x\ APriori}{ProbEvidência}$$

- Condições:
  - Existência das probabilidades a priori A; e
  - Existência das probabilidades a priori B.

Classificação baseada no Teorema de Bayes

- Considere **Y** a classe. Por exemplo, deseja-se prever se indivíduos vão tentar enganar durante a declaração de renda.
- Seja **X** = (**X**$_1$, **X**$_2$, …, **X**$_k$) o conjunto de atributos (proprietário de imóvel, estado civil, receita anual etc.).
- Pode-se tratar **X** e **Y** como variáveis aleatórias e determinar $P(X|Y)$ (probabilidade a posteriori).
- Conhecendo a probabilidade $P(X|Y)$ pode-se relacionar o registro **X** à classe que maximiza a probabilidade a posterior.
- Como se pode estimar $P(X|Y)$ a partir dos dados de treinamento?

Exemplo de dados

Registro	Proprietário de imóvel (binário)	Estado Civil (categórico)	Saláro anual (contínuo)	Declara corretamente? (classe)
1	Sim	Solteiro	125.000,00	Não
2	Não	Casado	100.000,00	Não
3	Não	Solteiro	70.000,00	Não
4	Sim	Casado	120.000,00	Não
5	Não	Divorciado	95.000,00	Sim
6	Não	Casado	60.000,00	Não
7	Sim	Divorciado	220.000,00	Não
8	Não	Solteiro	85.000,00	Sim
9	Não	Casado	75.000,0o	Não
10	Não	Solteiro	90.000,00	Sim

Abordagem Bayesiana

- A estimativa precisa da probabilidade a posteriori para toda possível combinação de atributos e classes requer um conjunto de treinamento muito grande, mesmo para uma quantidade moderada de atributos.
  - Pode-se, em lugar disto, utilizar o Teorema de Bayes:

$$P(Y|X) = \frac{P(X|Y)\ x\ P(Y)}{P(X)}$$

- P(X) é uma constante e pode ser calculada como um multiplicador de normalização.
- P(Y) pode ser facilmente estimado a partir do conjunto de treinamento (fraçao dos exemplos que pertence a cada classe).
- $P(X|Y)$ é uma tarefa mais desafiadora. Métodos:
  - Classificador Naïve Bayes;
  - Rede Bayesiana.

Classificador Naïve Bayes

- Assume-se que os atributos são condicionalmente independentes, dado o rótulo da classe **y**:

$$P(X|Y=y)=\prod_{i=1}^{k}P(X_i|Y=y)$$

então:

$$P(Y|X)=\frac{P(Y)\prod_{i=1}^{k}P(X_i|Y=y)}{P(X)}$$

- Precisa-se, então, estimar

$P(X_i|Y)$ para i = 1, ..., k.

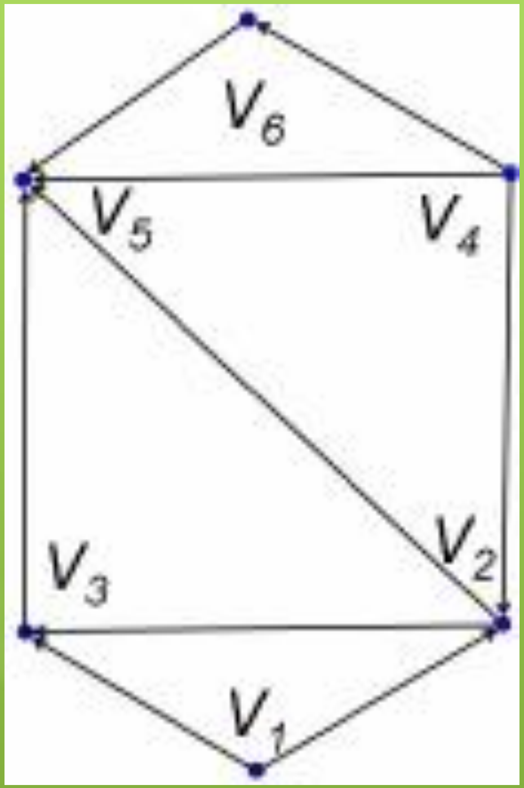
Classificador Naïve Bayes: Discussão

- Robusto em relação a pontos de ruído isolados.
- Pode trabalhar com valores faltantes.
- Robusto a valores irrelevantes.
- Atributos correlacionados podem degradar o desempenho pois a independência condicional não é respeitada.
  - Redes Bayesianas levam em conta a dependência entre atributos.

Grafos Direcionados

- Um grafo direcionado ou dígrafo G é um par ordenado G:=(V,A) em que:
  - V é um conjunto, cujos elementos são chamados de vértices ou nós;
  - $A \subseteq V \times V$ é um conjunto de pares de vértices ordenados, chamados de arcos ou ramos direcionados.
- Uma DAG é um grafo direcionado sem ciclos.

Noções de Teoria dos Grafos



- V1 e V4 são pais de V2;
- V5, V3 e V2 são descendentes de V2;
- V4 e V2 são ancestrais de V3;
- V6 e V4 são não descendentes de V1.

Definição de Rede Bayesiana

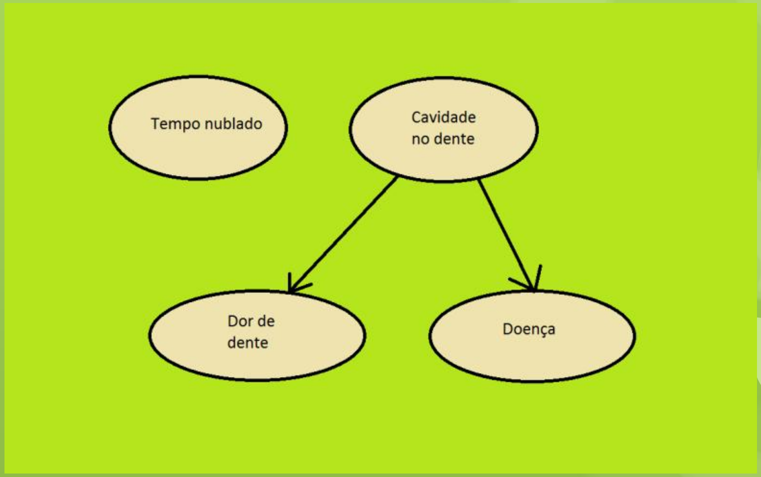
- Elementos de uma Rede Bayesiana:
  - DAG: codifica as relações de dependência entre um conjunto de variáveis;
  - Uma tabela de probabilidades associando cada nó com seus nós pais imediatos.

- Cada nó do grafo representa uma variável.

- Cada arco estabelece uma relação de dependência entre os pares de variáveis.

Exemplo 01

- A topologia da rede codifica as relações de dependência condicional:

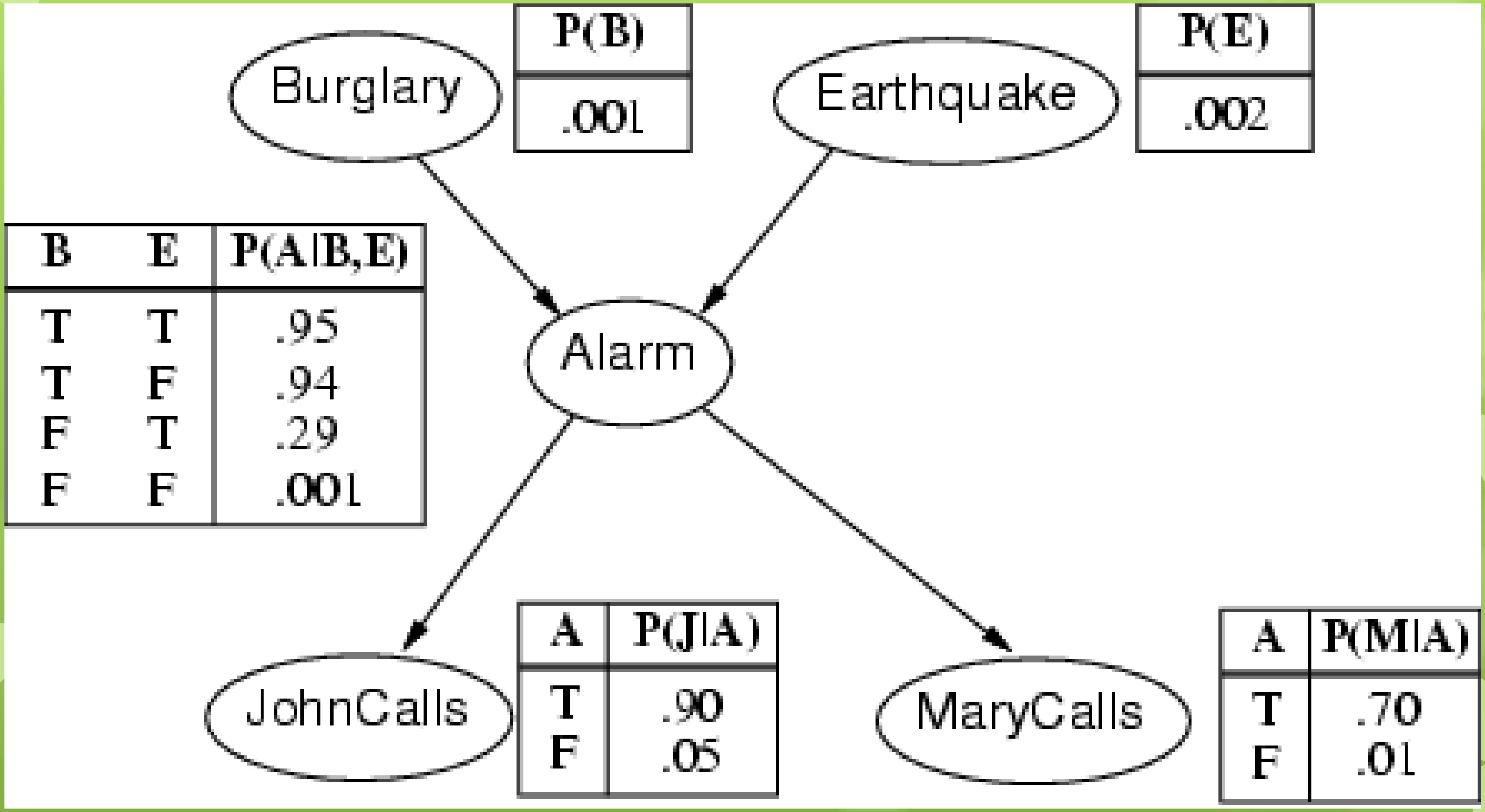


- *Tempo nublado* é independente das outras variáveis.
- *Dor de dente* e *Doença* são condicionalmente dependentes dada *Cavidade no dente.*

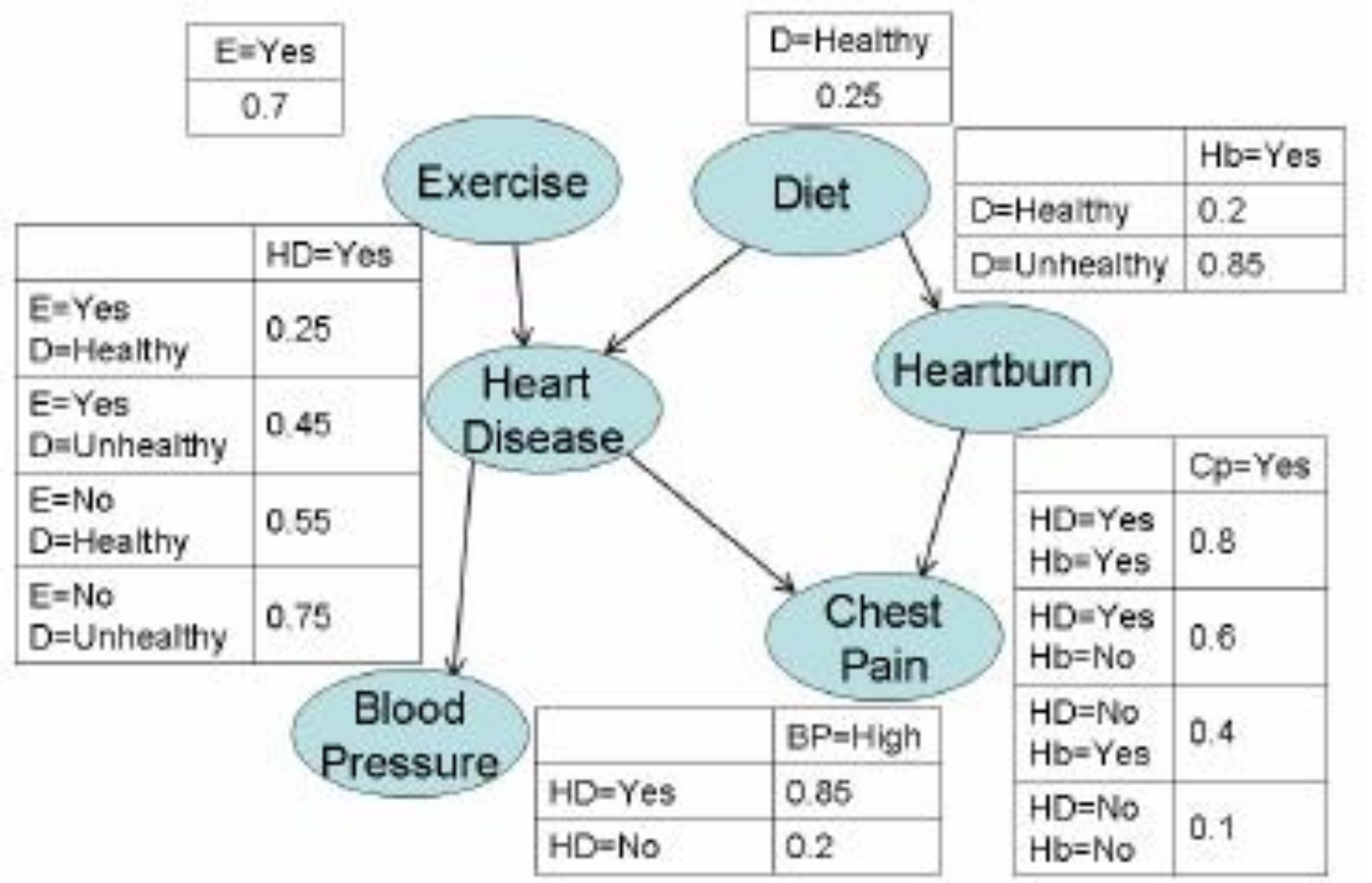
Exemplo 02

- Alguém está no trabalho e seu vizinho John telefona para dizer que o alarme da casa dele (em LA) disparou, sua vizinha Mary não telefona. Às vezes o alarme dispara devido a pequenos tremores de terra. Será que há um ladrão em casa?
- Variáveis: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*
- A topologia da rede reflete o conhecimento "causal":
  - Um ladrão pode disparar o alarme
  - Um tremor de terra pode disparar o alarme
  - O alarme pode fazer Mary me telefonar
  - O alarme pode fazer John me telefonar

Exemplo 02 (cont.)



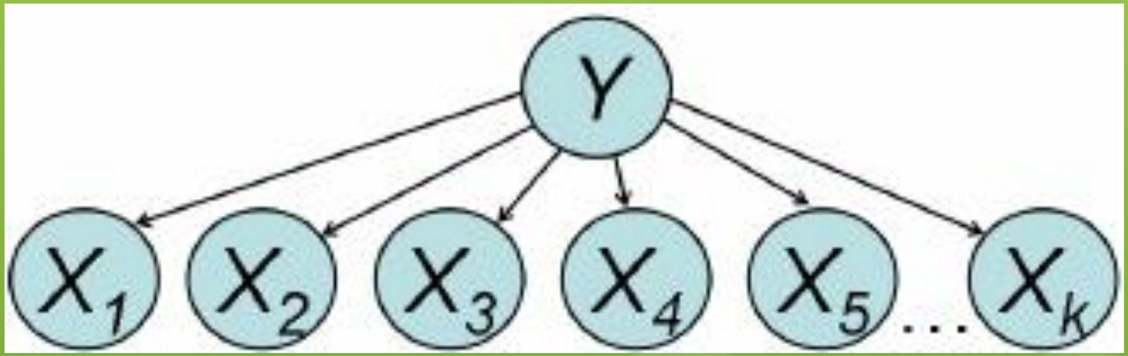
Rede Bayesiana – DAG + Tabelas



Representação de um Classificador Naïve Bayes

- Um classificador Naïve Bayes assume independência condicional dos atributos X_1, X_2, …, X_k em relação à classe Y.

- Isto pode ser representado pela seguinte Rede Bayesiana:



Rede Bayesiana - Discussão

- É uma forma elegante de codificar dependências probabilísticas causais.
- O modelo de dependência pode ser representado graficamente.
- Construir uma rede demanda esforço, mas acrescentar uma variável é quase imediato.
- Adequado mesmo com dados faltantes.
- Devido à natureza probabilística do modelo, o método é robusto ao sobreajuste ("overfitting").