



UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO

Departamento de Ciências de Computação

<http://www.icmc.usp.br>

SCC-630 - Capítulo 12

Aprendizado Probabilístico

João Luís Garcia Rosa¹

¹Departamento de Ciências de Computação
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo - São Carlos
joaoluis@icmc.usp.br

2011

Agradecimento

Agradeço à Profa. Maria Carolina Monard, que gentilmente permitiu que eu usasse seus slides [2] para preparação deste capítulo.

Sumário

1 Aprendizado Probabilístico

Material do Eamonn Keogh

Os próximos 25 slides contêm material do Prof. Eamonn Keogh [1], com adaptação da Profa. Maria Carolina Monard.

Fair Use Agreement

This agreement covers the use of all slides on this CD-Rom, please read carefully.

- You may freely use these slides for teaching, if
 - You send me an email telling me the class number/ university in advance.
 - My name and email address appears on the first slide (if you are using all or most of the slides), or on each slide (if you are just taking a few slides).
- You may freely use these slides for a conference presentation, if
 - You send me an email telling me the conference name in advance.
 - My name appears on each slide you use.
- You may not use these slides for tutorials, or in a published work (tech report/ conference paper/ thesis/ journal etc). If you wish to do this, email me first, it is highly likely I will grant you permission.

(c) Eamonn Keogh, eamonn@cs.ucr.edu

Classificador Naïve Bayes



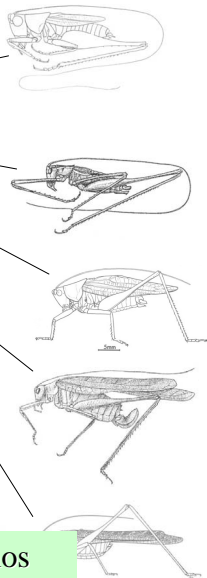
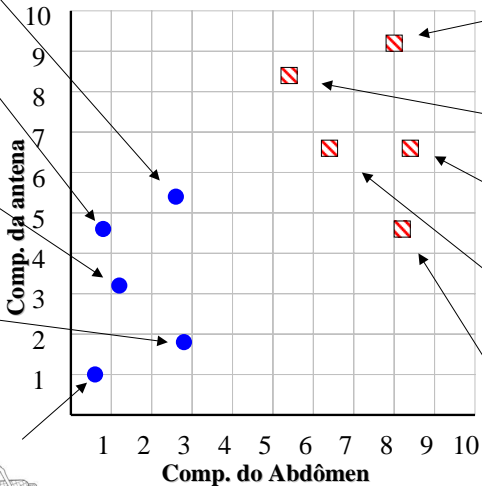
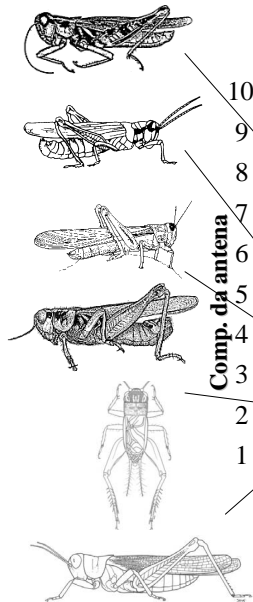
Thomas Bayes

1702 - 1761

Vamos iniciar com uma intuição visual, antes de ver a matemática...

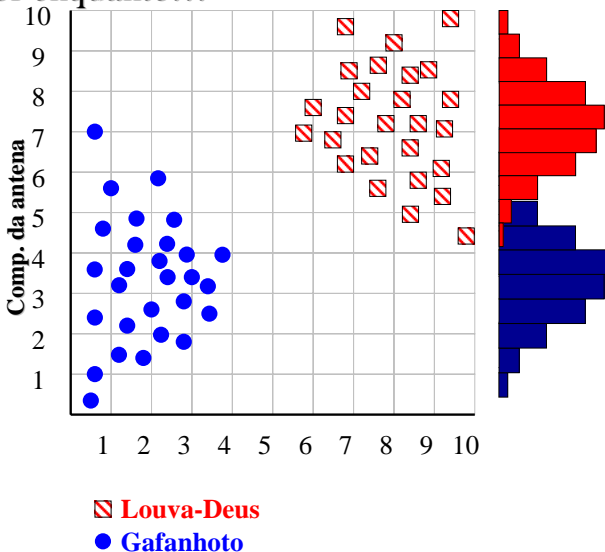
Gafanhoto

Esperança

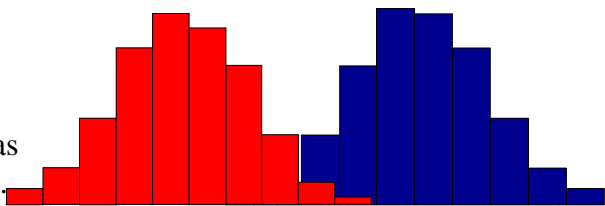


Lembra-se desse exemplo? Vamos pegar muitos outros dados...

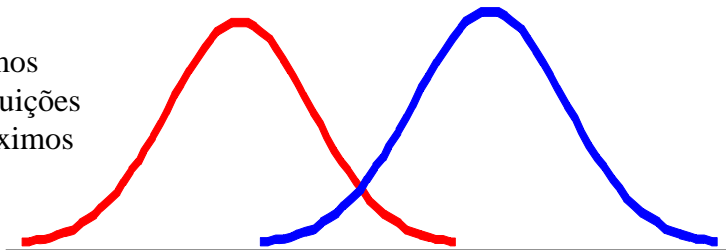
Com um monte de dados podemos construir um histograma. Vamos construir apenas um para “Comp. da antena”, por enquanto...



Podemos deixar os histogramas como estão ou podemos sumariá-los com duas distribuições normais.

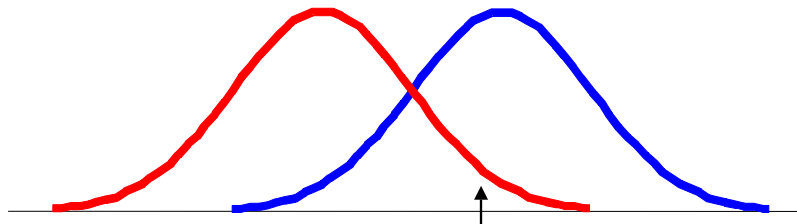


Para facilitar a visualização vamos usar duas distribuições normais nos próximos slides...

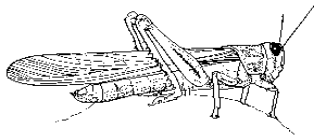


- Queremos classificar um inseto que encontramos. Sua antena tem 3 unidades de comprimento. Como classificá-lo?
- Podemos apenas nos perguntar, dada a distribuição de comprimento de antenas que vimos, é mais *provável* que nosso inseto seja um **Gafanhoto** ou um **Esperança**?
- Há uma maneira formal de discutir a classificação mais *provável*...

$p(c_j | d)$ = probabilidade da classe c_j , dado que observamos d



3

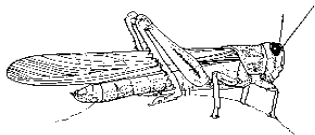
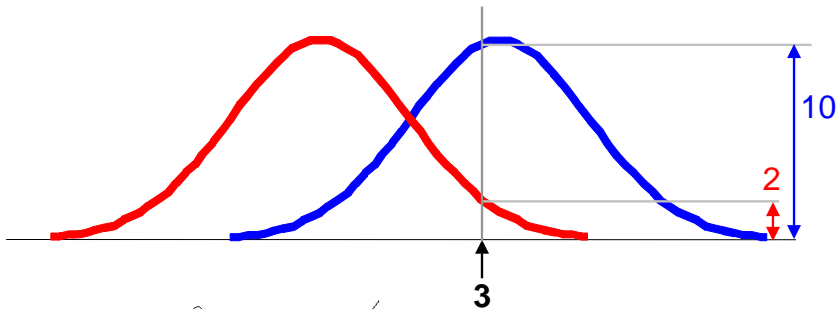


Comp. da antena é 3

$p(c_j | d)$ = probabilidade da classe c_j , dado que observamos d

$$P(\text{Gafanhoto} | 3) = 10 / (10 + 2) = 0.833$$

$$P(\text{Esperança} | 3) = 2 / (10 + 2) = 0.166$$

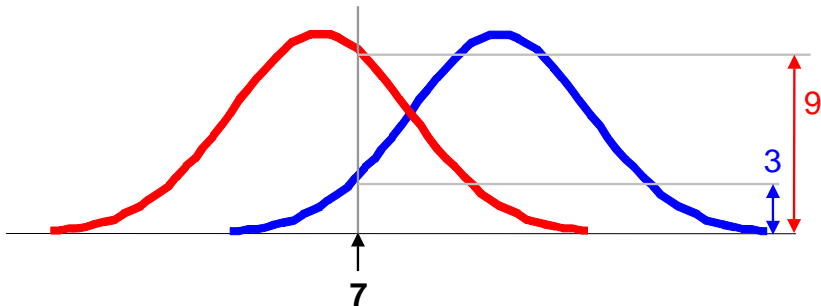


Comp. da antena é 3

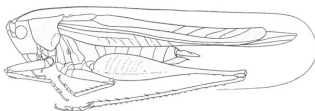
$p(c_j | d) = \text{probabilidade da classe } c_j, \text{ dado que observamos } d$

$$P(\text{Gafanhoto} | 7) = 3 / (3 + 9) = 0.250$$

$$P(\text{Esperança} | 7) = 9 / (3 + 9) = 0.750$$



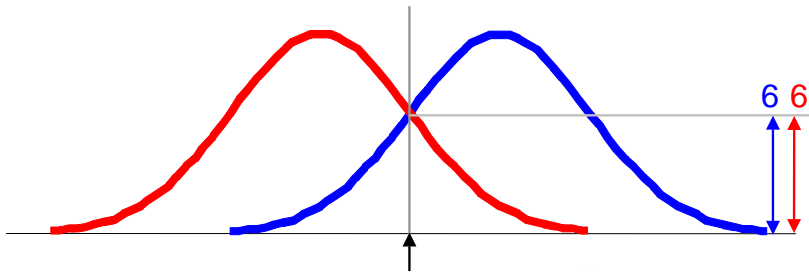
Comp. da antena é 7



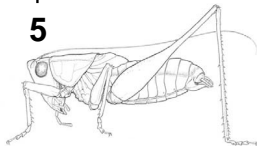
$p(c_j | d) = \text{probabilidade da classe } c_j, \text{ dado que observamos } d$

$$P(\text{Gafanhoto} | 5) = 6 / (6 + 6) = 0.500$$

$$P(\text{Esperança} | 5) = 6 / (6 + 6) = 0.500$$



Comp. da antena é 5



Classificadores Bayesianos

Essa foi uma intuição visual para um caso simples do Classificador Bayesiano, também chamado de:

- Naive Bayes
- Bayes Ingênuo
- Bayes Simples

Vamos ver agora alguns formalismos matemáticos, e mais exemplos, mas vamos manter a idéia básica na cabeça.

*Descubra a probabilidade de um **exemplo nunca visto antes** pertencer a cada classe, então simplesmente selecione a classe mais provável.*

Classificadores Bayesianos

- Classificador Bayesiano usa o **teorema de Bayes**, que diz

$$p(c_j | d) = \frac{p(d | c_j) p(c_j)}{p(d)}$$

- $p(c_j | d)$ = probabilidade do exemplo d ser da classe c_j ,
É isto que estamos tentando calcular
- $p(d | c_j)$ = probabilidade de gerar exemplo d dada a classe c_j ,
Nós podemos imaginar que ser da classe c_j , leva a ter a característica d com alguma probabilidade
- $p(c_j)$ = probabilidade de ocorrência da classe c_j ,
Representa apenas quão freqüente a classe c_j , é na nossa base
- $p(d)$ = probabilidade de exemplo d ocorrer

Isto pode ser ignorado, por ser igual para todas as classes

Assuma que temos duas classes

$C_1 = \text{homem}$, e $C_2 = \text{mulher}$.

Temos uma pessoa cujo sexo não conhecemos, digamos “*drew*” ou *d*.

Classificar *drew* como homem ou mulher é equivalente a se perguntar se é mais provável que *drew* é **homem** ou **mulher**, ou seja, qual é maior $p(\text{homem} | \text{drew})$ ou $p(\text{mulher} | \text{drew})$

(nota: “Drew pode ser o nome de um homem ou de uma mulher.”)



Drew Barrymore



Drew Carey

Qual a probabilidade de ser chamado “*drew*” dado que você é um **homem**?

$$p(\text{homem} | \text{drew}) = \frac{p(\text{drew} | \text{homem}) p(\text{homem})}{p(\text{drew})}$$

Qual a probabilidade de ser um **homem**?

Qual a probabilidade de ser chamado “*drew*”? (na verdade isso é irrelevante por ser a mesma para ambas as classes).



Oficial Drew

Este é o oficial Drew (que me prendeu em 1997). O oficial Drew é um **homem** ou **mulher**?

Por sorte, temos uma pequena base de dados com nomes e sexo.

E podemos usá-la para aplicar a regra de Bayes...

Nome	Sexo
Drew	homem
Claudia	mulher
Drew	mulher
Drew	mulher
Alberto	homem
Karin	mulher
Nina	mulher
Sergio	homem

$$p(c_j | d) = \frac{p(d | c_j) p(c_j)}{p(d)}$$



Oficial Drew

$$p(c_j | d) = \frac{p(d | c_j) p(c_j)}{p(d)}$$

Nome	Sexo
Drew	homem
Claudia	mulher
Drew	mulher
Drew	mulher
Alberto	homem
Karin	mulher
Nina	mulher
Sergio	homem

$$p(\text{homem} | \text{drew}) = \frac{1/3 * 3/8}{3/8} = \frac{0.125}{3/8}$$

$$p(\text{mulher} | \text{drew}) = \frac{2/5 * 5/8}{3/8} = \frac{0.250}{3/8}$$

É mais provável
que o oficial
Drew seja
mulher.



Oficial Drew é uma mulher!

Oficial Drew

$$p(\text{homem} | \text{drew}) = \frac{1/3 * 3/8}{3/8} = \frac{0.125}{3/8}$$

$$p(\text{mulher} | \text{drew}) = \frac{2/5 * 5/8}{3/8} = \frac{0.250}{3/8}$$

Até agora só consideramos a classificação Bayesiana quando temos um atributo (o “*Comp. da antena*”, ou o “*Nome*”). Mas podemos ter muitas características. Como usamos todas as características?

$$p(c_j | d) = \frac{p(d | c_j) p(c_j)}{p(d)}$$

Nome	Mais de 170CM	Olhos	Comp. do cabelo	Sexo
Drew	não	Azuis	Curto	homem
Claudia	sim	Castanhos	Longo	mulher
Drew	não	Azuis	Longo	mulher
Drew	não	Azuis	Longo	mulher
Alberto	sim	Castanhos	Curto	homem
Karin	não	Azuis	Longo	mulher
Nina	sim	Castanhos	Curto	mulher
Sergio	sim	Azuis	Longo	homem

- Para simplificar a tarefa, **Classificadores Naïve Bayes** assumem que os atributos tem distribuições independentes, e portanto, estimam que:

$$p(d|c_j) = p(d_1|c_j) * p(d_2|c_j) * \dots * p(d_n|c_j)$$

A probabilidade da classe c_j gerar exemplo d , igual a...

A probabilidade da classe c_j gerar o valor observado para o atributo 1, multiplicado por...

A probabilidade da classe c_j gerar o valor observado para o atributo 1, multiplicado por...

- Para simplificar a tarefa, **Classificadores Naïve Bayes** assumem que os atributos tem distribuições independentes, e portanto, estimam que:

$$p(d|c_j) = p(d_1|c_j) * p(d_2|c_j) * \dots * p(d_n|c_j)$$

$$p(\text{oficial drew}|c_j) = p(\text{Mais de}_{170\text{cm}} = \text{sim}|c_j) * p(\text{Olhos} = \text{Azuis}|c_j) * \dots$$



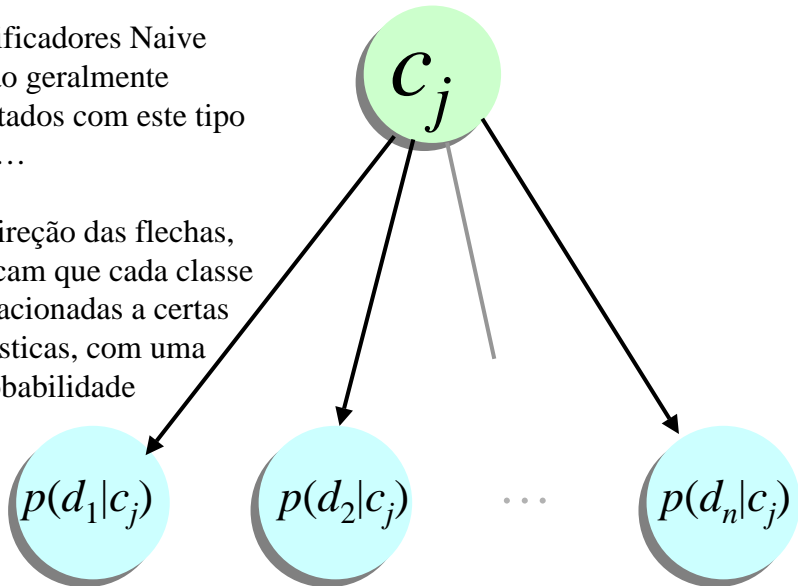
Oficial Drew
tem olhos
Azuis Mais
de 170_{cm} de
altura, e
cabelo longo

$$p(\text{oficial drew} | \text{mulher}) = 2/5 * 3/5 * \dots$$

$$p(\text{oficial drew} | \text{homem}) = 2/3 * 2/3 * \dots$$

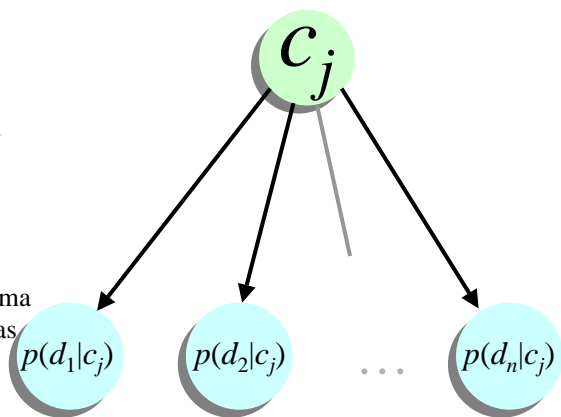
Os classificadores Naive Bayes são geralmente representados com este tipo de grafo...

Note a direção das flechas, que indicam que cada classe estão relacionadas a certas características, com uma certa probabilidade



O Naïve Bayes é rápido e eficiente em termos de espaço

Podemos descobrir todas as probabilidades varrendo apenas uma vez a base de dados e armazená-las em uma tabela (pequena) ...



Sexo	Mais de 190 _{cm}	
homem	sim	0.15
	não	0.85
mulher	sim	0.01
	não	0.99

Sexo	Cabelo Longo	
homem	sim	0.05
	não	0.95
mulher	sim	0.70
	não	0.30

Sexo	
homem	
mulher	

O Naïve Bayes NÃO é sensível a atributos irrelevantes...

Suponha que estamos tentando classificar o sexo das pessoas baseado em várias características, incluindo cor dos olhos (É claro que a cor dos olhos é completamente irrelevante para o sexo de uma pessoa)

$$p(\text{Jessica} | c_j) = p(\text{olhos} = \text{castanhos} | c_j) * p(\text{usa_vestido} = \text{sim} | c_j) * \dots$$

$$p(\text{Jessica} | \text{mulher}) = 9,000/10,000 * 9,975/10,000 * \dots$$

$$p(\text{Jessica} | \text{homem}) = 9,001/10,000 * 2/10,000 * \dots$$

Quase a mesma!

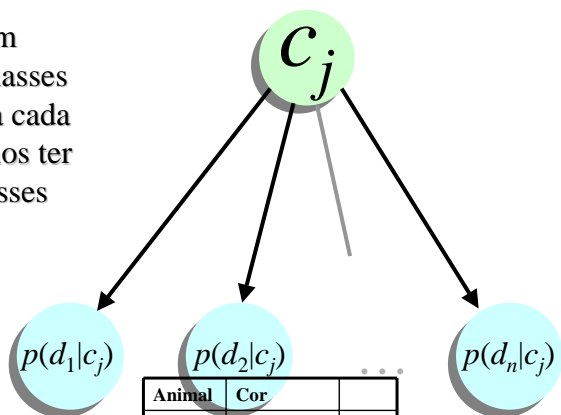
Entretanto, ele assume que temos estimativas das probabilidades boas o suficiente, portanto, quanto mais dados melhor

Um ponto óbvio. Usamos um problema simples de duas classes e dois valores possíveis para cada exemplo. Entretanto, podemos ter um número arbitrário de classes ou de valores de atributos

Animal	Peso > 10 _{kg}	
Gato	sim	0.15
	não	0.85
Cão	sim	0.91
	não	0.09
Porco	sim	0.99
	não	0.01

Animal	Cor	
Gato	Preto	0.33
	Branco	0.23
	Marrom	0.44
Cão	Preto	0.97
	Branco	0.03
	Marrom	0.90
Porco	Preto	0.04
	Branco	0.01
	Marrom	0.95

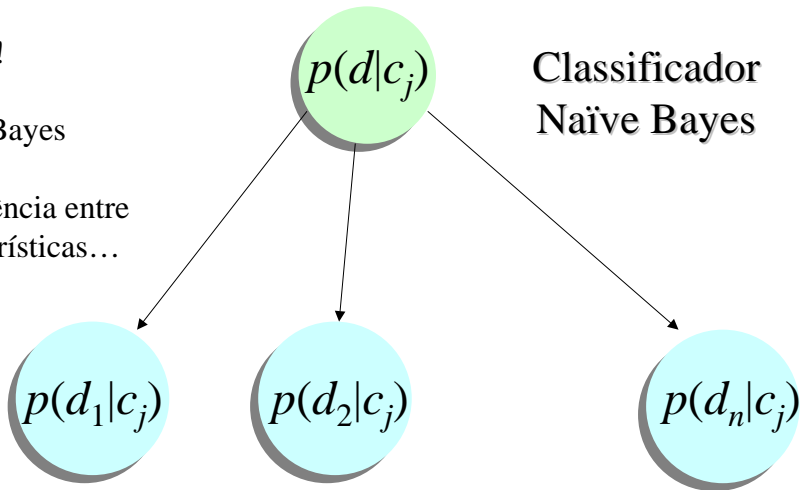
Animal
Gato
Cão
Porco



Problema!

O Naïve Bayes
assume
independência entre
as características...

Classificador
Naïve Bayes



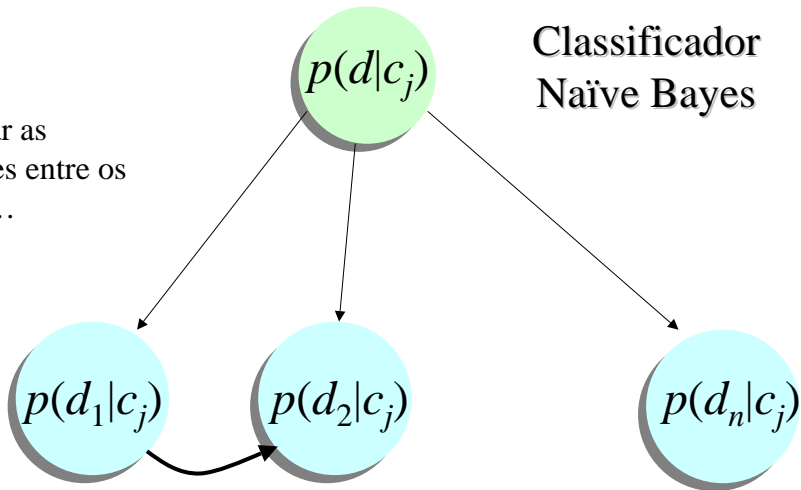
Sexo	Mais de 170 cm	
homem	sim	0.15
	não	0.85
mulher	sim	0.01
	não	0.99

Sexo	Mais de 100 kg	
homem	sim	0.11
	não	0.80
mulher	sim	0.05
	não	0.95

Solução

Considerar as correlações entre os atributos...

Classificador Naïve Bayes



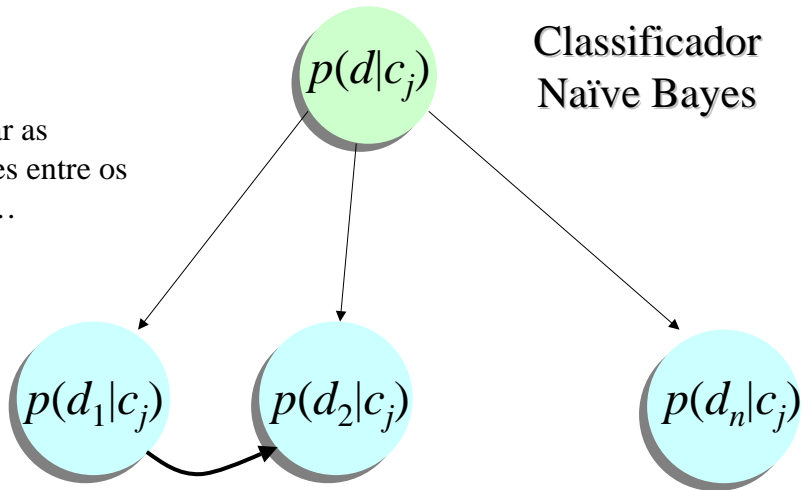
Sexo	Mais de 170 cm	
homem	sim	0.15
	não	0.85
mulher	sim	0.01
	não	0.99

Sexo	Mais de 100 kg	
homem	sim e Mais de 170 cm	0.11
	não e Mais de 170 cm	0.59
	sim e NÃO Mais de 170 cm	0.05
	não e NÃO Mais de 170 cm	0.35
mulher	sim e Mais de 170 cm	0.01

Solução

Considerar as correlações entre os atributos...

Classificador Naïve Bayes



Mas como encontramos o conjunto de arcos conectores??

Vantagens e Desvantagens do Naïve Bayes

- Vantagens:
 - Treinamento rápido (varredura única).
 - Rápido para classificar.
 - Não sensível a características irrelevantes.
 - Lida com dados reais e discretos.
 - Lida bem com dados contínuos.
- Desvantagem:
 - Assume independência das características.

Sumário da Classificação

- Vimos 4 técnicas de classificação principais:
 - Classificador linear simples,
 - vizinho mais próximo,
 - árvore de decisão,
 - Bayes.
- Há outras técnicas:
 - Redes neurais,
 - máquinas de suporte vetorial,
 - algoritmos genéticos,
 - sistemas classificadores fuzzy...
- Em geral, não há um classificador melhor para todos os problemas.
- Temos que considerar o que esperamos conseguir e os dados em si.

Referências I



[1] Eamonn Keogh,

Professor, Computer Science & Engineering Department,
University of California - Riverside.

[http:](http://www.cs.ucr.edu/~eamonn/tutorials.html)

[//www.cs.ucr.edu/~eamonn/tutorials.html](http://www.cs.ucr.edu/~eamonn/tutorials.html)



[2] Monard, M. C.

Slides da disciplina SCC630 - Inteligência Artificial. ICMC -
USP, 2010.