

# Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina

## Naive Bayes

Prof. Tiago A. Almeida

Tiago A. Almeida

## Classificador Naive-Bayes

- Método supervisionado (classificação)
  - Método baseado em estatística/probabilidades
- Cria modelos probabilísticos que permitem capturar incertezas sobre os dados
- Baseado no Teorema de Thomas Bayes (1702-1761): “Teorema de Bayes”

Tiago A. Almeida

## Classificador Naive-Bayes

- Trata-se de um algoritmo eficiente de aprendizado que gera hipóteses por meio do conhecimento estatístico extraído dos dados de treinamento
  - Hipóteses são criadas a partir de cálculos explícitos de probabilidades => **oferece robustez a ruídos nos dados de entrada**

Tiago A. Almeida

## Classificador Naive-Bayes

- **Classificação de texto**
  - considerado como um dos melhores algoritmos para classificação de documentos
- **Filtragem de spam**
  - aplicação mais conhecida dos classificadores Bayesianos na classificação de textos
    - DSPAM, SpamAssassin, SpamBayes, Bogofilter, ...

## Probabilidade em AM

- Sejam dois eventos A e B
  - A: atributo alvo (presença de uma doença)
    - Variável aleatória com dois valores: presença e ausência
  - B: atributo de entrada (resultado de um exame)
    - Variável aleatória com dois valores: positivo e negativo
  - $P(A)$ : probabilidade do evento A ocorrer (doença)
  - $P(B)$ : probabilidade do evento B ocorrer (exame +)

## Probabilidade em AM

### Exemplo:

Paciente	Teste	Doença
001	positivo	presente
002	negativo	presente
003	negativo	ausente
004	positivo	presente
005	positivo	ausente
006	positivo	presente
007	negativo	ausente
008	negativo	presente
009	positivo	ausente
010	positivo	presente

Probabilidade **a priori** pode ser estimada pela frequência

$$P(-) = 4/10 = 0,4$$

$$P(+) = 6/10 = 0,6$$

$$P(\text{presente}) = 6/10 = 0,6$$

$$P(\text{ausente}) = 4/10 = 0,4$$

O que se deseja em AM é a probabilidade **a posteriori**  
 $P(\text{doença}|\text{teste})$

## Probabilidade condicional

- Probabilidade **a priori** x **a posteriori** de um indivíduo estar doente
- Probabilidade de ocorrência de um evento pode depender da ocorrência de outro:  $P(A|B)$ 
  - Probabilidade de ocorrência de um evento A depende da ocorrência de um evento B
    - Ex.: Probabilidade de estar doente (A) dado que um exame (B) deu positivo
  - Se eventos são independentes  $\Rightarrow P(A|B) = P(A)$

## Probabilidade condicional

- É fácil estimar as probabilidades **a priori**: pela frequência
  - $P(B)$ : prob. do resultado do exame ser positivo
  - $P(A)$ : prob. do resultado do paciente estar doente
- É difícil estimar probabilidade **a posteriori**
  - $P(A|B)$ : probabilidade do paciente estar doente dado que um exame deu positivo
    - Uso do teorema de Bayes**: permite calcular probabilidade **a posteriori** de um evento

## Teorema de Bayes

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

## Teorema de Bayes

### Teorema de Bayes:

- $P(A|B) = P(B|A)P(A) / P(B)$ 
  - *Posteriori = (verossimilhança do dado x priori) / evidência*
  - $P(B)$ : lei da probabilidade total
    - Evento **A** pode ter dois possíveis resultados, **A<sub>1</sub> (A)** e **A<sub>2</sub> (~A)**, que formam uma partição

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2)$$

## Teorema de Bayes

### Lei da probabilidade total:

- $P(B)$ : lei da probabilidade total
  - Evento **A** pode ter dois possíveis resultados, **A<sub>1</sub> (A)** e **A<sub>2</sub> (~A)**, que formam uma partição
  - **Exemplo**
    - $P(\text{sexo}=M) = P(\text{sexo}=F) = 0.5$
    - $P(\text{olhos=escuro} | \text{sexo}=M) = 0.3$
    - $P(\text{olhos=escuro} | \text{sexo}=F) = 0.6$
  - $P(\text{olhos=escuro}) ???$

## Teorema de Bayes

### Lei da probabilidade total:

- $P(B)$ : lei da probabilidade total
  - Evento **A** pode ter dois possíveis resultados, **A<sub>1</sub> (A)** e **A<sub>2</sub> (~A)**, que formam uma partição
  - **Exemplo**
    - $P(\text{sexo}=M) = P(\text{sexo}=F) = 0.5$
    - $P(\text{olhos=escuro} | \text{sexo}=M) = 0.3$
    - $P(\text{olhos=escuro} | \text{sexo}=F) = 0.6$
  - **$P(\text{olhos=escuro}) =$** 
    - $P(\text{olhos=escuro} | \text{sexo}=F).P(\text{sexo}=F) + P(\text{olhos=escuro} | \text{sexo}=M).P(\text{sexo}=M) = 0.6*0.5 + 0.3*0.5 = \mathbf{0.45}$

## Teorema de Bayes

- A **probabilidade** de um evento A dado um evento B depende não apenas do relacionamento entre os eventos A e B [ $P(B|A)$ ], mas também da probabilidade marginal [ $P(A)$  e  $P(B)$ ] da ocorrência de cada evento

## Teorema de Bayes

### Exemplo

- Alguém diz que conversou com uma pessoa
  - Probabilidade de ter conversado com mulher é 50%

## Teorema de Bayes

### Exemplo

- Alguém diz que conversou com uma pessoa
  - Probabilidade de ter conversado com mulher é 50%
- Você recebe a informação de que a pessoa tinha cabelo comprido
  - Probabilidade de ter conversado com mulher aumenta
- Mas qual é a probabilidade?

## Teorema de Bayes

### Suponha que

- **M** - evento de ter conversado com mulher
- **C** - evento de ter conversado com alguém com cabelo comprido.
  - Para simplificar, suponha que
    - 50% da população é do sexo feminino  $\Rightarrow P(M) = 0.5$
    - 75% das mulheres têm cabelo comprido  $\Rightarrow P(C|M) = 0.75$
    - 30% dos homens têm cabelo comprido  $\Rightarrow P(C|H) = 0.3$
- Qual o valor de  $P(M|C)$ ?

## Teorema de Bayes

- Usando teorema de Bayes:
  - $P(M|C) = [P(C|M) P(M)] / P(C)$
- Usando a lei de probabilidade total:
  - $P(M|C) = [P(C|M) P(M)] / \{ [P(C|M) P(M)] + [P(C|H) P(H)] \}$

## Teorema de Bayes

- Usando teorema de Bayes:
  - $P(M|C) = [P(C|M) P(M)] / P(C)$
- Usando a lei de probabilidade total:
  - $P(M|C) = [P(C|M) P(M)] / \{ [P(C|M) P(M)] + [P(C|H) P(H)] \}$
  - $P(M|C) = (0.75 \times 0.5) / [(0.75 \times 0.5) + (0.3 \times 0.5)]$
  - $P(M|C) = 0.375 / (0.375 + 0.15)$
  - $P(M|C) = 0.375 / (0.525)$
  - $P(M|C) = 0.714 \Rightarrow 71\%$

## Teorema de Bayes

- **Exemplo 2**
  - Um entomologista acredita ter encontrado uma espécie rara de besouro devido às características das suas asas. Em espécies raras, 98% possuem tal característica. Em espécies comuns, apenas 5%. As espécies raras representam 0.1% da população. Qual a probabilidade do besouro ser raro?

## Teorema de Bayes

- **Exemplo 2**
  - Um entomologista acredita ter encontrado uma espécie rara de besouro devido às características das suas asas. Em espécies raras, 98% possuem tal característica. Em espécies comuns, apenas 5%. As espécies raras representam 0.1% da população. Qual a probabilidade do besouro ser raro?
  - $P(R) = 0.001 \Rightarrow P(C) = 0.999$
  - $P(A|R) = 0.98$  e  $P(A|C) = 0.05$

## Teorema de Bayes

### Exemplo 2

- Um entomologista acredita ter encontrado uma espécie rara de besouro devido às características das suas asas. Em espécies raras, 98% possuem tal característica. Em espécies comuns, apenas 5%. As espécies raras representam 0.1% da população. Qual a probabilidade do besouro ser raro?
- $P(R) = 0.001 \Rightarrow P(C) = 0.999$
- $P(A|R) = 0.98$  e  $P(A|C) = 0.05$
- $P(R|A) = P(A|R) \times P(R) / \{[P(A|R) \times P(R) + P(A|C) \times P(C)]\}$

## Teorema de Bayes

### Exemplo 2

- Um entomologista acredita ter encontrado uma espécie rara de besouro devido às características das suas asas. Em espécies raras, 98% possuem tal característica. Em espécies comuns, apenas 5%. As espécies raras representam 0.1% da população. Qual a probabilidade do besouro ser raro?
- $P(R) = 0.001 \Rightarrow P(C) = 0.999$
- $P(A|R) = 0.98$  e  $P(A|C) = 0.05$
- $P(R|A) = P(A|R) \times P(R) / \{[P(A|R) \times P(R) + P(A|C) \times P(C)]\}$
- $P(R|A) = (0.98 \times 0.001) / [(0.98 \times 0.001) + (0.05 \times 0.999)] = 0.019$
- $P(R|A) = 1.9\%$**

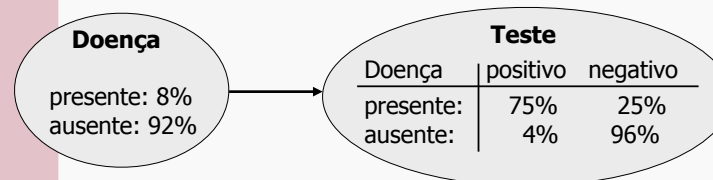
## Aprendizado Bayesiano

- Ocorre de acordo com experiências passadas
  - Ex. diagnóstico de uma doença X
    - Probabilidade de alguém possuir a doença é 8%
    - Teste para diagnóstico possui um grau de incerteza
      - 75% dos testes positivos e 96% dos testes negativos foram confirmados
  - Como representar essa informação?**

**Doença:** variável aleatória com dois possíveis valores (presente/ausente)  
**Resultado do teste:** dois valores (+ / -)

## Aprendizado Bayesiano

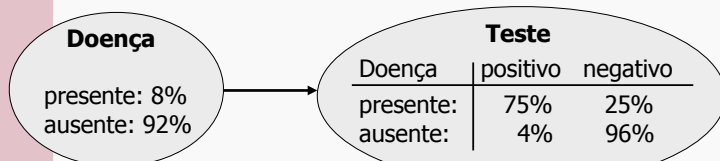
- Ex. diagnóstico de uma doença X
  - Valor da doença influencia valor do teste
  - $\Rightarrow$  representação em um grafo



- Nós = variáveis ou atributos
- Arestas = influência entre as variáveis

## Aprendizado Bayesiano

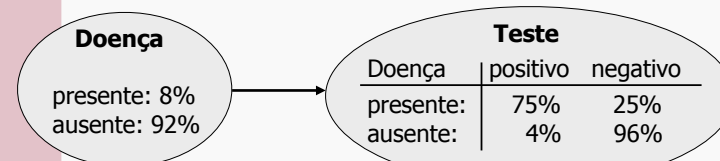
### Ex. diagnóstico de uma doença X



- Probabilidade de verdadeiros positivos (sensibilidade) = 75%
  - $P(\text{Teste} = \text{positivo} | \text{Doença} = \text{presente})$
- Probabilidade de verdadeiros negativos (especificidade) = 96%
  - $P(\text{Teste} = \text{negativo} | \text{Doença} = \text{ausente})$

## Aprendizado Bayesiano

### Ex. diagnóstico de uma doença X



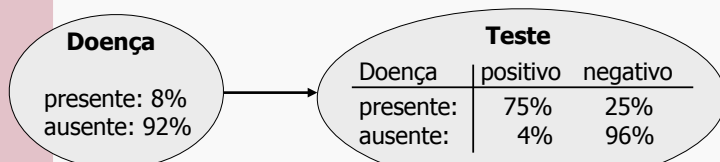
- Qual o poder preditivo do Teste com respeito à Doença?

$$P(\text{Teste} = \text{positivo}) = P(\text{Teste}) = P(\text{Teste} | \text{Doença})P(\text{Doença}) + P(\text{Teste} | \sim \text{Doença})P(\sim \text{Doença}) = 0,75 \times 0,08 + 0,04 \times 0,92 = 0,0968$$

$$P(\text{Teste} = \text{negativo}) = P(\sim \text{Teste}) = P(\sim \text{Teste} | \text{Doença})P(\text{Doença}) + P(\sim \text{Teste} | \sim \text{Doença})P(\sim \text{Doença}) = 0,25 \times 0,08 + 0,96 \times 0,92 = 0,9032$$

## Aprendizado Bayesiano

### Ex. diagnóstico de uma doença X



- Se sabemos que o resultado do Teste de um paciente foi +, podemos concluir que ele está doente?

## Aprendizado Bayesiano

- $P(y_i | \mathbf{x})$  = probabilidade de  $\mathbf{x}$  pertencer a classe  $y_i$

- Função de custo é minimizada se  $\mathbf{x}$  é associado a  $y_k$  para o qual  $P(y_k | \mathbf{x})$  é máxima

- Estimativa MAP** (*Maximum a Posteriori*)

- Predição:

$$y_{\text{MAP}} = \arg \max_i P(y_i | \mathbf{x})$$

## Aprendizado Bayesiano

- Funções que calculam probabilidades  $P(y_i|\mathbf{x})$  são chamadas **discriminantes**
  - Separam exemplos de classes diferentes
  - Teorema de Bayes provê método para calcular  $P(y_i|\mathbf{x})$

$$P(y_i | \mathbf{x}) = \frac{P(y_i) P(\mathbf{x} | y_i)}{P(\mathbf{x})}$$

## Aprendizado Bayesiano

- Teorema de Bayes provê método para calcular  $P(y_i|\mathbf{x})$

$$P(y_i | \mathbf{x}) = \frac{P(y_i) P(\mathbf{x} | y_i)}{P(\mathbf{x})}$$

$P(\mathbf{x})$  pode ser ignorado, pois é o mesmo para todas as classes, não afetando os valores relativos de suas probabilidades

## Naive Bayes

$$P(y_i | \mathbf{x}) = \frac{P(y_i) P(\mathbf{x} | y_i)}{P(\mathbf{x})}$$

- Classificador Bayesiano mais simples
  - Assume que os valores dos atributos de um exemplo são **independentes** entre si dada a classe

## Naive Bayes

- Temos:

$$P(y_i|\vec{x}) \propto P(y_i) \prod_{j=1}^n P(x^j|y_i)$$

Todas as probabilidades necessárias são computadas a partir dos dados de treinamento



## Naive Bayes

- Naive Bayes:
  - $P(y_i)$ : manter contador para cada classe
  - $P(x_i|y_i)$ : Probabilidade de observar um valor de atributo dado que o exemplo pertence a uma classe
    - Depende do tipo do atributo
      - **Qualitativo**: manter contador para cada valor de atributo por classe
      - **Quantitativo**: assumir distribuição para os valores do atributo (ex: Normal) ou discretizar o atributo
        - Discretização é mais comum: = min(10, número de valores diferentes) intervalos de mesmo tamanho



## Análise do algoritmo

- Vantagens do NB:
  - Rápido e exige poucos recursos para armazenamento
  - Robusto às características irrelevantes
    - Características irrelevantes cancelam outras, sem afetar os resultados
  - Ótimo se as suposições de independência forem verdadeiras
    - Se a independência assumida de fato existir, então o classificador será ótimo para o problema
  - Bom *baseline*, principalmente para a classificação de texto

## Análise do algoritmo

- Desvantagens do NB:
  - Desconsidera relações entre os atributos
  - Dificuldade para atributos contínuos

## Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina

### Classificação de textos

Prof. Tiago A. Almeida

## É spam?

De: WebMail UFSCar <paolina.swider@ufscar.br>  
 Assunto: Caro Membro ufscar.br  
 Data: 15 de novembro de 2012 12:26:01 BRST  
 Para: undisclosed-recipients: ;  
 Responder A: upxuv@07168.cn

Ocultar

Spam

Caro Membro ufscar.br,

Devido ao congestionamento em toda ufscar.br! Contas de usuários, ufscar.br webmail seria encerrar todas as contas não utilizadas.

A fim de evitar a desativação de sua conta, você terá que confirmar seu e-mail, preencha suas informações de login abaixo, clicando no botão de resposta.

As informações pessoais solicitadas são para a segurança da sua Conta ufscar.br. Por favor, deixe todas as informações solicitadas.

Nome completo: .....  
 Nome de usuário: .....  
 Senha: .....  
 Confirmar senha: .....  
 Data de Nascimento: .....

Depois você deve seguir as instruções na folha, a sua conta ufscar.br não será interrompida e continuará sendo normal.

Obrigado por seu habitual co-operação. Pedimos desculpas por qualquer inconveniente.

UFSCar Atendimento ao Cliente WebMail  
 Número do processo: 8941624  
 Propriedade: Segurança da Conta

## Quem é o autor?

- 1787/1788: cartas anônimas tentaram convencer Nova York a alterar a Constituição dos EUA:
  - Jay, Madison, Hamilton
- Autoria de 12 cartas em disputa
- 1963: problema resolvido por Mosteller & Wallace usando métodos Bayesianos



James Madison



Alexander Hamilton

## Autor ou autora?

- “By 1925 present-day Vietnam was divided into three parts under French colonial rule. The southern region embracing Saigon and the Mekong delta was the colony of Cochin-China; the central area with its imperial capital at Hue was the protectorate of Annam...”
- “Clara never failed to be astonished by the extraordinary felicity of her own name. She found it hard to trust herself to the mercy of fate, which had managed over the years to convert her greatest shame into one of her greatest assets...”

S. Argamon, M. Koppel, J. Fine, A. R. Shimoni, 2003. “Gender, Genre, and Writing Style in Formal Written Texts,” Text, volume 23, number 3, pp. 321–346

## Comentário positivo ou negativo?

- “Incrivelmente decepcionante.”
- “Sátira cheia de personagens malucos e ricamente aplicado, e algumas reviravoltas grandes.”
- “Esta é a maior comédia já filmada.”
- “Foi patético. A pior parte foram as cenas de luta.”

## Qual é o assunto do artigo?

### MEDLINE Article



### Assunto

- Epidemiologia
- Embriologia
- Química
- Terapia

## Classificação de texto

- Atribuir categorias de assuntos, temas ou gêneros
- Detecção de spam
- Identificação de autoria
- Identificação de idade / sexo
- Identificação de linguagem
- ...

## Classificação de texto: definição

- **Entradas**
  - Documento  $d$
  - Conjunto de classes  $C = \{c_1, c_2, \dots, c_j\}$
- **Saída**
  - Classe resultante  $c \in C$

## Classificação de texto: manual

- Regras baseadas em combinações de palavras ou outras características
  - spam: *blacklist* OU (“\$” E “viagra”)
- Acurácia pode ser alta
  - Se as regras forem cuidadosamente refinadas por especialista
  - Construção e manutenção de regras tem custo alto

## Classificação de texto: supervisionada

### Entradas

- Documento  $d$
- Conjunto de classes  $C = \{c_1, c_2, \dots, c_j\}$
- Conjunto de treinamento com  $m$  documentos rotulados  $(d_1, c_1), \dots, (d_m, c_m)$

### Saída

- Classificador  $f(d) \rightarrow c$

## Naive Bayes - Intuição

- Classificação é baseada no teorema de Bayes
- A representação do documento normalmente é muito simples
  - Bag of words*

## Representação

### Bag of words

$Y(\text{I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet.}) = C$

## Representação

### Bag of words

$Y(\text{I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet.}) = C$

## Representação

### ▪ Bag of words

Y(  
 x love sweet  
 satirical great  
 fun  
 whimsical  
 romantic laughing  
 recommend  
 several happy again  
 ) = C

## Representação

### ▪ Bag of words

Y(  

great	2
love	2
recommend	1
laugh	1
happy	1
...	...

 ) = C

## Naive Bayes - Classificação

### ▪ Para um documento **d** e uma classe **c**

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)}$$

## Naive Bayes - Classificação

$$c_{MAP} = \operatorname{argmax}_{c \in C} P(c|d)$$

MAP é "maximum a posteriori" = classe com maior probabilidade

$$= \operatorname{argmax}_{c \in C} \frac{P(d|c)P(c)}{P(d)}$$

Equação de Bayes

$$= \operatorname{argmax}_{c \in C} P(d|c)P(c)$$

Remoção do denominador

## Naive Bayes - Classificação

$$c_{MAP} = \operatorname{argmax}_{c \in C} P(d | c) P(c)$$

$$= \operatorname{argmax}_{c \in C} P(x_1, x_2, \dots, x_n | c) P(c)$$

Documento d  
representado  
como atributos  
x1..xn

## Naive Bayes - Independência

$$P(x_1, x_2, \dots, x_n | c)$$

- Suposições “ingênuas”:
  - Posicionamento:** a posição de cada palavra no texto não influencia na escolha da classe
  - Independência condicional:** a probabilidade de ocorrência de cada atributo  $P(x_i | c_j)$  é independente, dada a classe  $c$

$$P(x_1, \dots, x_n | c) = P(x_1 | c) \cdot P(x_2 | c) \cdot P(x_3 | c) \cdot \dots \cdot P(x_n | c)$$

## Naive Bayes - Classificador

$$c_{MAP} = \operatorname{argmax}_{c \in C} P(x_1, x_2, \dots, x_n | c) P(c)$$

$$c_{NB} = \operatorname{argmax}_{c \in C} P(c_j) \prod_{x \in X} P(x | c)$$

## Naive Bayes - Classificador

posições ← posições das palavras no documento teste

$$c_{NB} = \operatorname{argmax}_{c_j \in C} P(c_j) \prod_{i \in \text{posições}} P(x_i | c_j)$$

## Naive Bayes - Aprendizado

- Objetivo principal: calcular estimativas de máxima verossimilhança
  - simplesmente usar as frequências nos dados

$$\hat{P}(c_j) = \frac{N(C = c_j)}{N}$$

$$\hat{P}(x_i | c_j) = \frac{N(X_i = x_i, C = c_j)}{N(C = c_j)}$$

## Naive Bayes - Aprendizado

$$\hat{P}(X_i = w | c_j) = \text{quantidade de vezes que a palavra } w \text{ aparece entre todos os documentos da classe } c_j.$$

- Criar um mega-documento (**M**) composto pela concatenação de todos os documentos pertencentes à classe  $j$ 
  - Usar a frequência de  $w$  em **M**

## Naive Bayes - Aprendizado

- **Problema com a máxima verossimilhança**
  - O que ocorre ao processar uma palavra  $w$  que não está presente nos documentos de treinamento?

$$\hat{P}(X_i = w | c_j) = 0$$

$$c_{MAP} = \operatorname{argmax}_c \hat{P}(c) \prod_i \hat{P}(x_i | c)$$

- **$C_{MAP} = 0$**

## Naive Bayes - Aprendizado

- **Suavização das estimativas**

- Laplace (Adicionar 1):

$$\hat{P}(x_i | c_j) = \frac{N(X_i = x_i, C = c_j) + 1}{N(C = c_j) + k}$$

# de valores de  $X_i$

- Aplicação em texto:

$$\hat{P}(w | c) = \frac{\text{count}(w, c) + 1}{\text{count}(c) + |V|}$$

## Naive Bayes - Aprendizado

- Extrair *dicionário* da base de treinamento

### Calcular os termos $P(c_j)$ :

Para cada  $c_j$  em  $C$  faça

$doc_j \leftarrow$  todos os docs da classe =  $c_j$

$$P(c_j) \leftarrow \frac{|doc_j|}{|\text{total \# documentos}|}$$

### Calcular os termos $P(w_k|c_j)$ :

$Texto_j \leftarrow$  único doc contendo todos os  $doc_j$

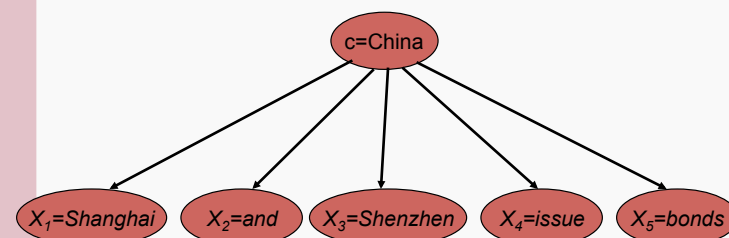
Para cada palavra  $w_k$  do Dicionário

$n_k \leftarrow$  # de ocorrências de  $w_k$  em  $Texto_j$

$$P(w_k | c_j) \leftarrow \frac{n_k + 1}{n + |\text{Dicionário}|}$$

## Naive Bayes - Modelagem

- NB modela a classe como conjunto de palavras



## Naive Bayes - Modelagem

- Associar cada palavra ( $w$ ):  $P(w | c)$
- Associar cada sentença ( $s$ ):  $P(s | c) = \prod P(w | c)$

Class *pos*

0.1	I	0.1	love	0.05	this	0.01	fun	0.1	film
0.1	love	0.1	0.1	0.05	0.01	0.1			
0.01	this								
0.05	fun								
0.1	film								
...									

$$P(s | \text{pos}) = 0.0000005$$

## Naive Bayes - Modelagem

- Em qual classe  $s$  tem maior probabilidade de estar associada?

Model pos		Model neg						
0.1	I	0.2	I					
0.1	love	0.001	love					
0.01	this	0.01	this					
0.05	fun	0.005	fun					
0.1	film	0.1	film					

	I	love	this	fun	film
0.1	0.1	0.01	0.05	0.1	
0.2	0.001	0.01	0.005	0.1	

$$P(s | \text{pos}) > P(s | \text{neg})$$



## Naive Bayes - Exemplo

	Doc	Palavras	Classe
Treino	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Teste	5	Chinese Chinese Chinese Tokyo Japan	?

## Naive Bayes - Exemplo

	Doc	Palavras	Classe
Treino	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Teste	5	Chinese Chinese Chinese Tokyo Japan	?

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w|c) = \frac{\text{count}(w,c)+1}{\text{count}(c)+|V|}$$

## Naive Bayes - Exemplo

	Doc	Palavras	Classe
Treino	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Teste	5	Chinese Chinese Chinese Tokyo Japan	?

$$\hat{P}(c) = \frac{N_c}{N}$$

$$P(c) = 3/4$$

$$P(j) = 1/4$$

$$\hat{P}(w|c) = \frac{\text{count}(w,c)+1}{\text{count}(c)+|V|}$$

$$P(\text{Chinese}|c) = (5+1) / (8+6) = 6/14 = 3/7$$

$$P(\text{Tokyo}|c) = (0+1) / (8+6) = 1/14$$

$$P(\text{Japan}|c) = (0+1) / (8+6) = 1/14$$

$$P(\text{Chinese}|j) = (1+1) / (3+6) = 2/9$$

$$P(\text{Tokyo}|j) = (1+1) / (3+6) = 2/9$$

$$P(\text{Japan}|j) = (1+1) / (3+6) = 2/9$$

## Naive Bayes - Exemplo

	Doc	Palavras	Classe
Treino	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Teste	5	Chinese Chinese Chinese Tokyo Japan	?

$$\hat{P}(c) = \frac{N_c}{N}$$

$$P(c) = 3/4$$

$$P(j) = 1/4$$

### Escolhendo a classe

$$P(c|d5) \propto 3/4 * (3/7)^3 * 1/14 * 1/14$$

$$\approx 0.0003$$

$$P(j|d5) \propto 1/4 * (2/9)^3 * 2/9 * 2/9$$

$$\approx 0.0001$$

$$\hat{P}(w|c) = \frac{\text{count}(w,c)+1}{\text{count}(c)+|V|}$$

$$P(\text{Chinese}|c) = (5+1) / (8+6) = 6/14 = 3/7$$

$$P(\text{Tokyo}|c) = (0+1) / (8+6) = 1/14$$

$$P(\text{Japan}|c) = (0+1) / (8+6) = 1/14$$

$$P(\text{Chinese}|j) = (1+1) / (3+6) = 2/9$$

$$P(\text{Tokyo}|j) = (1+1) / (3+6) = 2/9$$

$$P(\text{Japan}|j) = (1+1) / (3+6) = 2/9$$

# Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina

## Medidas de avaliação

Prof. Tiago A. Almeida

## Classificador - Avaliação

- Tabela de confusão (ou contingência)

	Pred=pos	Pred=neg
Classe=pos	<i>tp</i>	<i>fn</i>
Classe=neg	<i>fp</i>	<i>tn</i>

## Avaliação - Acurácia

	Pred=pos	Pred=neg
Classe=pos	<i>tp</i>	<i>fn</i>
Classe=neg	<i>fp</i>	<i>tn</i>

$$Ac = \frac{tp + tn}{tp + tn + fp + fn}$$

## Avaliação - Precisão e Revocação

	Pred=pos	Pred=neg
Classe=pos	<i>tp</i>	<i>fn</i>
Classe=neg	<i>fp</i>	<i>tn</i>

$$Prec_{pos} = \frac{tp}{tp + fp}$$

$$Rec_{pos} = \frac{tp}{tp + fn}$$

$$Prec_{neg} = \frac{tn}{tn + fp}$$

$$Rec_{neg} = \frac{tn}{fp + tn}$$

## Avaliação - F-medida

	Pred=pos	Pred=neg
Classe=pos	<i>tp</i>	<i>fn</i>
Classe=neg	<i>fp</i>	<i>tn</i>

$$F = 2 \frac{Pre \cdot Rec}{Pre + Rec}$$

## Referências

### Ilustrações utilizadas:

- [http://www.webciencia.com/11\\_05coracao.htm](http://www.webciencia.com/11_05coracao.htm)
- <http://desenhospaintcolor.blogspot.com/2010/05/desenho-de-dado-para-colorir-desenho-de.html>
- <http://sermineirouai.blogspot.com/2011/02/minas-gerais-quem-te-conhece.html>
- [http://www.organizacaocultural.ufba.br/index.php?instituicao/exibir\\_pub/40](http://www.organizacaocultural.ufba.br/index.php?instituicao/exibir_pub/40)

### Alguns slides:

- Prof Dr André C. P. L. F. Carvalho

### Material probabilidade:

<http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-050j-information-and-entropy-spring-2008/>