Inteligência Artificial

Naive Bayes

Tatiane Nogueira Rios
tatiane.nogueira@ufba.br

Recapitulando...

- Teorema de Bayes
  - Expressa a probabilidade a posteriori de uma hipótese baseado
    - na probabilidade a priori dessa hipótese
    - na probabilidade a priori da evidência
    - verossimilhança entre causa e efeito

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)}$$

Combinação de evidências

- A regra de Bayes pode ser útil para responder a consultas probabilísticas condicionadas sobre uma única peça de evidência.

- As informações probabilísticas estão frequentemente disponíveis sob a forma: $P(efeito|causa)$.

Combinação de evidências

- Mas o que acontece se tivermos duas ou mais evidências?
- Ex: $P(cárie|dordedente \wedge boticão)$ = ?

	dordedente		*~dordedente*	
	boticão	*~boticão*	*boticão*	*~boticão*
cárie	0,108	0,012	0,072	0,008
~cárie	0,016	0,064	0,144	0,576

Combinação de evidências

- Mas o que acontece se tivermos duas ou mais evidências?
- Ex: $P(cárie|dordedente \wedge boticão)$ = ?

	dordedente		*~dordedente*	
	boticão	*~boticão*	*boticão*	*~boticão*
cárie	0,108	0,012	0,072	0,008
~cárie	0,016	0,064	0,144	0,576

Distribuição conjunta total

Combinação de evidências

- A reformulação anterior só funciona **se forem conhecidas todas as probabilidades condicionais** da conjunção $dordedente \wedge boticão$ para cada valor de cárie.
  - Só é viável para apenas duas variáveis de evidência.
  - Se houver n variáveis de evidência possíveis (raio X, dieta, higiene oral etc..), então haveria 2^n combinações possíveis de valores observados para os quais precisaríamos conhecer probabilidades condicionais.

Combinação de evidências

É preciso encontrar asserções adicionais sobre o domínio para simplificar as expressões.

Essas variáveis são independentes, dada a presença ou a ausência de uma cárie.

Cada uma é diretamente causada pela cárie, mas nenhuma delas tem efeito direto sobre a outra: a dor de dente depende do estado dos nervos do dente, enquanto a precisão da ferramenta depende da habilidade do dentista, para o qual a dor de dente é irrelevante.

Logo, tem-se que:

$$P\ (dordedente \wedge boticão|cárie) = P(dordedente|carie)P(boticão|cárie)$$

Combinação de evidências

É preciso encontrar asserções adicionais sobre o domínio para simplificar as expressões.

Essas variáveis são independentes, dada a presença ou a ausência de uma cárie.

Cada uma é diretamente causada pela cárie, mas nenhuma delas tem efeito direto sobre a outra: a dor de dente depende do estado dos nervos do dente, enquanto a precisão da ferramenta depende da habilidade do dentista, para o qual a dor de dente é irrelevante.

Independência condicional

Logo, tem-se que:

$$P\ (dordedente \wedge boticão|cárie) = P(dordedente|carie)P(boticão|cárie)$$

Combinação de evidências

Quando há mais de uma evidência, os requisitos de informações são idênticos aos da inferência, utilizando cada item de evidência separadamente: a probabilidade a priori para a variável de consulta e a probabilidade condicional de cada efeito, dada sua causa (Regra de Bayes + Independência condicional):

$$P(cárie|dordedente \wedge boticão) = P(dordedente|cárie)\ P(boticão|cárie)\ P(cárie)$$

Combinação de evidências

- A decomposição de grandes domínios probabilísticos em subconjuntos conectados livremente por meio de independência condicional é um dos desenvolvimentos mais importantes na história recente da IA.

- Quando uma única causa influencia de maneira direta vários efeitos, todos **condicionalmente independentes**, dada a causa, tem-se a distribuição conjunta total:
  - $P(Causa, Efeito_1, \ldots, Efeito_n) = P(Causa)\prod_i P(Efeito_n|Causa)$

Combinação de evidências

- A decomposição de grandes domínios probabilísticos em subconjuntos conectados livremente por meio de independência condicional é um dos desenvolvimentos mais importantes na história recente da IA.

- Quando uma única causa influencia de maneira direta vários efeitos, todos condicionalmente independentes, dada a causa, tem-se a distribuição conjunta total:
  - $P(Causa, Efeito_1, \ldots, Efeito_n) = P(Causa)\prod_i P(Efeito_n|Causa)$

Modelo bayesiano ingênuo = Naive Bayes

Classificador Naive Bayes

- Junto com árvores de decisão, redes neurais e método de agrupamento dos vizinhos mais próximos (KNN), Naive-Bayes é um dos métodos de aprendizagem mais práticos.
- Quando usar:
  - Quando o conjunto de dados disponível é de tamanho moderado à grande
  - Quando os atributos que descrevem as instâncias são **condicionalmente independentes das classes**
- Aplicações bem sucedidas:
  - Diagnóstico médico
  - Classificação de documentos textuais

Classificador Naive Bayes

- Tarefa de atribuir a uma nova instância a classe mais provável dados os valores de atributos que descrevem a instância.

$$v_{MAP} = \arg\max_{v_j \in V} P(v_j \mid a_1, a_2 \ldots a_n)$$

Teorema de Bayes

$$v_{MAP} = \arg\max_{v_j \in V} \frac{P(v_j \mid a_1, a_2 \ldots a_n) P(v_j)}{P(a_1, a_2 \ldots a_n)}$$

Por que P(a1,...,an) sumiu?

$$= \arg\max_{v_j \in V} P(v_j \mid a_1, a_2 \ldots a_n) P(v_j)$$

Classificador Naive Bayes

- Calcular P(a1, a2, ..., an | vj) é inviável, pois o número de termos como este é o número possível de instâncias (com todas as possibilidades de valores de atributos) multiplicado pelo número possível de classes vj.
  - Processo caro
  - Necessitaria de um conjunto de treinamento muito grande para evitar o problema dos dados serem esparsos

Classificador Naive Bayes

- O classificador naive bayes resolve o problema com a suposição ingênua de que os atributos são condicionalmente independentes dada a classe

- A conjunção de atributos <a1, a2, ..., an> corresponde à multiplicação das probabilidades de cada atributo

$$P(a_1, \ldots, a_n \mid v_j) = \prod_i P(a_i \mid v_j)$$

Classificador Naive Bayes

- Reescrevendo a equação anterior:

$$v_{map} = \arg\max_{v_j \in V} P(a1, a2, \ldots, an \mid v_j) P(v_j)$$

$$\Downarrow$$

$$v = \arg\max_{v_j \in V} P(v_j) \prod_i P(ai \mid v_j)$$

Classificador Naive Bayes

- Reescrevendo a equação anterior:

$$v_{map} = \arg\max_{v_j \in V} P(a1, a2, \ldots, an \mid v_j) P(v_j)$$

$$\Downarrow$$

$$v = \arg\max_{v_j \in V} P(v_j) \prod_i P(ai \mid v_j)$$

- Agora, o número de P(ai|vj) é o número de valores do atributo ai multiplicado pelo número de classes, o que é viável, pois é muito menos do que o cálculo anterior de P(a1, a2, ..., an|vj)

Classificador Naive Bayes

- Portanto,
  - O processo de aprendizado consiste em estimar P(vj) e P(ai|vj) a partir do conjunto de treinamento.
  - O conjunto de estimativas (probabilidades) consiste na hipótese aprendida, usada para classificar novos exemplos.

Exemplo 1

- Considere o conjunto de treinamento:

dia	aparência	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	**não**
D2	ensolarado	quente	alta	forte	**não**
D3	nublado	quente	alta	fraco	**sim**
D4	chuva	moderada	alta	fraco	**sim**
D5	chuva	fria	normal	fraco	**sim**
D6	chuva	fria	normal	forte	**não**
D7	nublado	fria	normal	forte	**sim**
D8	ensolarado	moderada	alta	fraco	**não**
D9	ensolarado	fria	normal	fraco	**sim**
D10	chuva	moderada	normal	fraco	**sim**
D11	ensolarado	moderada	normal	forte	**sim**
D12	nublado	moderada	alta	forte	**sim**
D13	nublado	quente	normal	fraco	**sim**
D14	chuva	moderada	alta	forte	**não**

Exemplo 1

- Classifique a instância abaixo como sim (jogar tênis) ou não.
  - <aparência=ensolarado, temperatura=fria, umidade=alta, vento=forte>

dia	aparência	temperatura	umidade	vento	jogar_tênis
D1	ensolarado	quente	alta	fraco	**não**
D2	ensolarado	quente	alta	forte	**não**
D3	nublado	quente	alta	fraco	**sim**
D4	chuva	moderada	alta	fraco	**sim**
D5	chuva	fria	normal	fraco	**sim**
D6	chuva	fria	normal	forte	**não**
D7	nublado	fria	normal	forte	**sim**
D8	ensolarado	moderada	alta	fraco	**não**
D9	ensolarado	fria	normal	fraco	**sim**
D10	chuva	moderada	normal	fraco	**sim**
D11	ensolarado	moderada	normal	forte	**sim**
D12	nublado	moderada	alta	forte	**sim**
D13	nublado	quente	normal	fraco	**sim**
D14	chuva	moderada	alta	forte	**não**

Exercício 1

(Aprendizado probabilístico) Tom é estudante da UFBA. Recentemente, seu humor tem sido altamente influenciado por dois fatores: as condições climáticas (Tempo) e seu desempenho nos estudos (Resultado). Naturalmente, ele gosta de um bom tempo e odeia resultados ruins. Mais importante, Tom se preocupa com seu desempenho nos estudos. Ele se sente feliz quando passa em uma prova e triste quando perde. Agora Tom quer predizer sua felicidade de acordo com estes dois fatores utilizando suas experiências prévias. Tabelas A e B mostram estes dados.

Tempo (T)	Resultado (R)	Felicidade (F)
Ruim	Perdeu	0
Bom	Perdeu	0
Bom	Perdeu	0
Bom	Perdeu	0
Ruim	Passou	0
Ruim	Passou	1
Ruim	Passou	1
Bom	Passou	1

Tabela A: 2 Fatores

Exercício 1a

Usando a Tabela A: Se a situação hoje fosse: T=Bom, R=Passou, e Tom utilizasse um classificador de Naive Bayes, como ele iria predizer sua felicidade? Por favor, apresente todos os cálculos e o resultado da classificação.

Tempo (T)	Resultado (R)	Felicidade (F)
Ruim	Perdeu	0
Bom	Perdeu	0
Bom	Perdeu	0
Bom	Perdeu	0
Ruim	Passou	0
Ruim	Passou	1
Ruim	Passou	1
Bom	Passou	1

Exercício 1a

Usando a Tabela A: Se a situação hoje fosse: T=Bom, R=Passou, e Tom utilizasse um classificador de Naive Bayes, como ele iria predizer sua felicidade? Por favor, apresente todos os cálculos e o resultado da classificação.

Tempo (T)	Resultado (R)	Felicidade (F)
Ruim	Perdeu	0
Bom	Perdeu	0
Bom	Perdeu	0
Bom	Perdeu	0
Ruim	Passou	0
Ruim	Passou	1
Ruim	Passou	1
Bom	Passou	1

P(T=Bom|F=0)P(R=Passou|F=0)P(F=0)
3/5 * 1/5 * 5/8 = 0.075 = 0.075/(0.3333+0.075) = 19%

P(T=Bom|F=1)P(R=Passou|F=1)P(F=1) =
1/3 * 3/3 = 0.3333 = 0.3333/(0.3333+0.075) = 81%

Resultado da classificação = Felicidade (1) = Feliz

Exercício 1b

Usando a Tabela A: Se a situação hoje fosse T=Ruim e R=Perdeu, e Tom utilizasse um classificador de Naive Bayes, como ele iria predizer sua felicidade? Por favor, apresente todos os cálculos e o resultado da classificação.

Exercício 1 - Resposta b

Usando a Tabela A: Se a situação hoje fosse T=Ruim e R=Perdeu, e Tom utilizasse um classificador de Naive Bayes, como ele iria predizer sua felicidade? Por favor, apresente todos os cálculos e o resultado da classificação.

Tempo (T)	Resultado (R)	Felicidade (F)
Ruim	Perdeu	0
Bom	Perdeu	0
Bom	Perdeu	0
Bom	Perdeu	0
Ruim	Passou	0
Ruim	Passou	1
Ruim	Passou	1
Bom	Passou	1

P(T=Ruim|F=0)P(R=Perdeu|F=0)P(F=0) =
2/5 * 4/5 * 5/8 = 0.2

P(T=Ruim|F=1)P(R=Perdeu|F=1)P(F=1) =
2/3 * 0/3 * 3/8 = 0

Resultado da classificação = Felicidade (0) = Triste

Exercício 2

(Aprendizado probabilístico) Tom percebeu que seu vizinho sempre faz caminhada quando o tempo está bom e fica em casa quando o tempo está ruim. Tom acha que não faria mal ter mais informações, por isso ele acrescenta mais um fator, Vizinho (V), na tabela. Na nova tabela, Tabela B, você pode ver que sempre que o tempo está bom (T=Bom), o vizinho faz caminhada (V=Sai), e sempre que o tempo está ruim (T=Ruim), o vizinho fica em casa (V=Fica).

Tempo (T)	Resultado (R)	Vizinho (V)	Felicidade (F)
Ruim	Perdeu	Fica	0
Bom	Perdeu	Sai	0
Bom	Perdeu	Sai	0
Bom	Perdeu	Sai	0
Ruim	Passou	Fica	0
Ruim	Passou	Fica	1
Ruim	Passou	Fica	1
Bom	Passou	Sai	1

Tabela B: 3 Fatores

Exercício 2a

Usando a Tabela B: Se a situação hoje fosse: T=Bom, R=Passou, V=Saiu e Tom utilizasse um classificador de Naive Bayes, como ele iria predizer sua felicidade? Por favor, apresente todos os cálculos e o resultado da classificação.

Tempo (T)	Resultado (R)	Vizinho (V)	Felicidade (F)
Ruim	Perdeu	Fica	0
Bom	Perdeu	Sai	0
Bom	Perdeu	Sai	0
Bom	Perdeu	Sai	0
Ruim	Passou	Fica	0
Ruim	Passou	Fica	1
Ruim	Passou	Fica	1
Bom	Passou	Sai	1

Exercício 2a

Usando a Tabela B: Se a situação hoje fosse: T=Bom, R=Passou, V=Saiu e Tom utilizasse um classificador de Naive Bayes, como ele iria predizer sua felicidade? Por favor, apresente todos os cálculos e o resultado da classificação.

Tempo (T)	Resultado (R)	Vizinho (V)	Felicidade (F)
Ruim	Perdeu	Fica	0
Bom	Perdeu	Sai	0
Bom	Perdeu	Sai	0
Bom	Perdeu	Sai	0
Ruim	Passou	Fica	0
Ruim	Passou	Fica	1
Ruim	Passou	Fica	1
Bom	Passou	Sai	1

P(T=Bom|F=0)P(R=Passou|F=0)P(F=0)
3/5 * 1/5 * 5/8 = 0.075 = 0.075/(0.3333+0.075) = 19%

P(T=Bom|F=1)P(R=Passou|F=1)P(F=1) =
1/3 * 3/3 = 0.3333 = 0.3333/(0.3333+0.075) = 81%

Resultado da classificação = Felicidade (1) = Feliz

Exercício 2 - Resposta b

Porque o novo fator não melhora a performance do classificador de Naive Bayes? Por que?

Porque o tempo e o vizinho são não condicionalmente independentes.

Exercício 2 - Resposta b

Porque o novo fator não melhora a performance do classificador de Naive Bayes? Por que?

Exercício 2c

Usando a Tabela B: Se Tom usasse um **classificador bayesiano (Teorema de Bayes)** ao invés de um **classificador naive bayes**, e a situação hoje fosse T=Bom e R=Passou, e Vizinho=Saiu, como ele iria predizer sua felicidade? Por favor, apresente todos os cálculos e o resultado da classificação.

Exercício 2c

Usando a Tabela B: Se Tom usasse um **classificador bayesiano** ao invés de um **classificador de naive bayes**, e a situação hoje fosse T=Bom e R=Passou, e Vizinho=Saiu, como ele iria predizer sua felicidade? Por favor, apresente todos os cálculos e o resultado da classificação.

Tempo (T)	Resultado (R)	Vizinho (V)	Felicidade (F)
Ruim	Perdeu	Fica	0
Bom	Perdeu	Sai	0
Bom	Perdeu	Sai	0
Bom	Perdeu	Sai	0
Ruim	Passou	Fica	0
Ruim	Passou	Fica	1
Ruim	Passou	Fica	1
Bom	Passou	Sai	1

P(T=Bom, R=Passou, Vizinho=Saiu|F=0) P(F=0) =
0 * 5/7 = 0
P(T=Bom, R=Passou, Vizinho=Saiu|F=1) P(F=1) =
1/3 * 3/8 = 0.125

Resultado da classificação = Felicidade (1) = Feliz

Exemplo 3

- Classifique o comentário "Great phone buy it" em positivo ou negativo.

DOC	WORDS	CLASS
1	DONT BUY	NEGATIVE
2	PHONE GOT HANGED	NEGATIVE
3	BATTERY DRAINS FAST	NEGATIVE
4	DURABLE PHONE	POSITIVE
5	GREAT CAMERA	POSITIVE
6	GREAT PHONE BUY IT	?

Treinamento

WORDS	CLASS
DONT BUY	NEGATIVE
PHONE GOT HANGED	NEGATIVE
BATTERY DRAINS FAST	NEGATIVE
DURABLE PHONE	POSITIVE
GREAT CAMERA	POSITIVE

Teste

WORDS	CLASS
GREAT PHONE BUY IT	?

Exemplo 3

- Classifique o comentário "Great phone buy it" em positivo ou negativo.

DOC	WORDS	CLASS
1	DONT BUY	NEGATIVE
2	PHONE GOT HANGED	NEGATIVE
3	BATTERY DRAINS FAST	NEGATIVE
4	DURABLE PHONE	POSITIVE
5	GREAT CAMERA	POSITIVE
6	GREAT PHONE BUY IT	?

Dados estruturados

WORDS	POSITIVE	NEGATIVE
DONT	0	1
BUY	0	1
PHONE	1	1
GOT	0	1
HANGED	0	1
BATTERY	0	1
DRAINS	0	1
FAST	0	1
DURABLE	1	0
GREAT	1	0
CAMERA	1	0

Exemplo 3

- Classifique o comentário "Great phone buy it" em positivo ou negativo.

DOC	WORDS	CLASS
1	DONT BUY	NEGATIVE
2	PHONE GOT HANGED	NEGATIVE
3	BATTERY DRAINS FAST	NEGATIVE
4	DURABLE PHONE	POSITIVE
5	GREAT CAMERA	POSITIVE
6	GREAT PHONE BUY IT	?

P(positive) = ⅖ = 0.4

P(negative) = ⅗ = 0.6

Exemplo 3

- Classifique o comentário "Great phone buy it" em positivo ou negativo.

WORDS	POSITIVE	NEGATIVE
DONT	0	1
BUY	0	1
PHONE	1	1
GOT	0	1
HANGED	0	1
BATTERY	0	1
DRAINS	0	1
FAST	0	1
DURABLE	1	0
GREAT	1	0
CAMERA	1	0

P(great | positive) = 1/4
P(phone | positive) = 1/4
P(buy | positive) = 0/4
P(it | positive) = 0/4

P(great | negative) = 0/8
P(phone | negative) = 1/8
P(buy | negative) = 1/8
P(it | negative) = 0/8

Exemplo 3

- Classifique o comentário "Great phone buy it" em positivo ou negativo.

WORDS	POSITIVE	NEGATIVE
DONT	0	1
BUY	0	1
PHONE	1	1
GOT	0	1
HANGED	0	1
BATTERY	0	1
DRAINS	0	1
FAST	0	1
DURABLE	1	0
GREAT	1	0
CAMERA	1	0

P(great | positive) = 1/4
P(phone | positive) = 1/4
P(buy | positive) = 0/4
P(it | positive) = 0/4

P(great | negative) = 0/8
P(phone | negative) = 1/8
P(buy | negative) = 1/8
P(it | negative) = 0/8

Como evitar este overfitting?

Exemplo 3

- Classifique o comentário "Great phone buy it" em positivo ou negativo.

WORDS	POSITIVE	NEGATIVE
DONT	0	1
BUY	0	1
PHONE	1	1
GOT	0	1
HANGED	0	1
BATTERY	0	1
DRAINS	0	1
FAST	0	1
DURABLE	1	0
GREAT	1	0
CAMERA	1	0

P(great | positive) = 1/4
P(phone | positive) = 1/4
P(buy | positive) = 0/4
P(it | positive) = 0/4

P(great | negative) = 0/8
P(phone | negative) = 1/8
P(buy | negative) = 1/8
P(it | negative) = 0/8

Laplace Smoothing !

Exemplo 3

- Classifique o comentário "Great phone buy it" em positivo ou negativo.

WORDS	POSITIVE	NEGATIVE
DONT	0	1
BUY	0	1
PHONE	1	1
GOT	0	1
HANGED	0	1
BATTERY	0	1
DRAINS	0	1
FAST	0	1
DURABLE	1	0
GREAT	1	0
CAMERA	1	0

Laplace Smoothing

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

P(GREAT | POSITIVE) =	1+1 / 4+11	=	2/15	= **0.13**
P(PHONE | POSITIVE) =	1+1 / 4+11	=	2/25	= **0.13**
P(BUT | POSITIVE) =	0+1 / 4+11	=	1/15	= **0.07**
P(IT| POSITIVE) =	0+1 / 4+11	=	1/15	= **0.07**
P(GREAT | NEGATIVE) =	0+1 / 8+11	=	1/19	= **0.05**
P(PHONE | NEGATIVE) =	1+1 / 8+11	=	2/19	= **0.11**
P(BUT | NEGATIVE) =	1+1 / 8+11	=	2/19	= **0.11**
P(IT| NEGATIVE) =	0+1 / 8+11	=	1/19	= **0.05**

Exemplo 3

- Classifique o comentário "Great phone buy it" em positivo ou negativo.

WORDS	POSITIVE	NEGATIVE
DONT	0	1
BUY	0	1
PHONE	1	1
GOT	0	1
HANGED	0	1
BATTERY	0	1
DRAINS	0	1
FAST	0	1
DURABLE	1	0
GREAT	1	0
CAMERA	1	0

Laplace Smoothing

Hiperparâmetro

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

P(GREAT | POSITIVE) =	1+1 / 4+11	=	2/15	= **0.13**
P(PHONE | POSITIVE) =	1+1 / 4+11	=	2/25	= **0.13**
P(BUT | POSITIVE) =	0+1 / 4+11	=	1/15	= **0.07**
P(IT| POSITIVE) =	0+1 / 4+11	=	1/15	= **0.07**
P(GREAT | NEGATIVE) =	0+1 / 8+11	=	1/19	= **0.05**
P(PHONE | NEGATIVE) =	1+1 / 8+11	=	2/19	= **0.11**
P(BUT | NEGATIVE) =	1+1 / 8+11	=	2/19	= **0.11**
P(IT| NEGATIVE) =	0+1 / 8+11	=	1/19	= **0.05**

Exemplo 3

- Classifique o comentário "Great phone buy it" em positivo ou negativo.

P(GREAT | POSITIVE) =	1+1 / 4+11	=	2/15	= **0.13**
P(PHONE | POSITIVE) =	1+1 / 4+11	=	2/25	= **0.13**
P(BUT | POSITIVE) =	0+1 / 4+11	=	1/15	= **0.07**
P(IT| POSITIVE) =	0+1 / 4+11	=	1/15	= **0.07**
P(GREAT | NEGATIVE) =	0+1 / 8+11	=	1/19	= **0.05**
P(PHONE | NEGATIVE) =	1+1 / 8+11	=	2/19	= **0.11**
P(BUT | NEGATIVE) =	1+1 / 8+11	=	2/19	= **0.11**
P(IT| NEGATIVE) =	0+1 / 8+11	=	1/19	= **0.05**

P(positive) = ⅖ = 0.4

P(negative) = ⅗ = 0.6

P(POSITIVE) = P(GREAT | POSITIVE) P(PHONE | POSITIVE) P(BUT | POSITIVE) P(IT| POSITIVE) P(POSITIVE)
= (0.13) x (0.13) x (0.07) x (0.07) x (0.4)
P(POSITIVE) = 0.00003

P(NEGATIVE) = P(GREAT | NEGATIVE) P(PHONE | NEGATIVE) P(BUT | NEGATIVE) P(IT| NEGATIVE) P(NEGATIVE)
= (0.05) x (0.11) x (0.11) x (0.05) x (0.6)
P(NEGATIVE) = 0.00002

Referências

- Faceli et al., Inteligência Artificial – Uma Abordagem de Aprendizado de Máquina, LTC, 2015.
- Smola, A. and Vishwanathan, S.V.N., Introduction to Machine Learning, Cambridge University Press, 2008
- RUSSELL, S.; NORVIG, P. Inteligência Artificial. Editora Campus, 2013, Capítulo 14.
- https://inst.eecs.berkeley.edu/~cs188/sp20/assets/notes/n9.pdf
- https://inst.eecs.berkeley.edu/~cs188/sp20/assets/lecture/lec27_naivebayes.pdf
- https://www.youtube.com/watch?v=0kTUXhsdhtY