



Data Science  
Academy

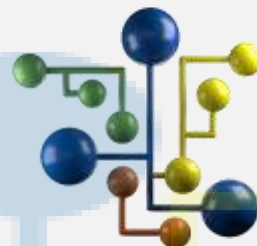
Data Science Academy [rodrigo.c.abreu@hotmail.com](mailto:rodrigo.c.abreu@hotmail.com) 5e207d48e32fc335fa60447d

# Machine Learning



Data Science  
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



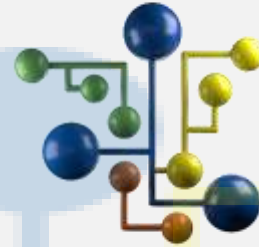
**Data Science  
Academy**

Seja muito bem-vindo(a)!



Data Science  
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

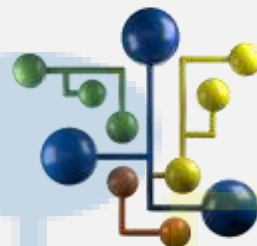


**Data Science  
Academy**

# Classificação com Naive Bayes



Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d

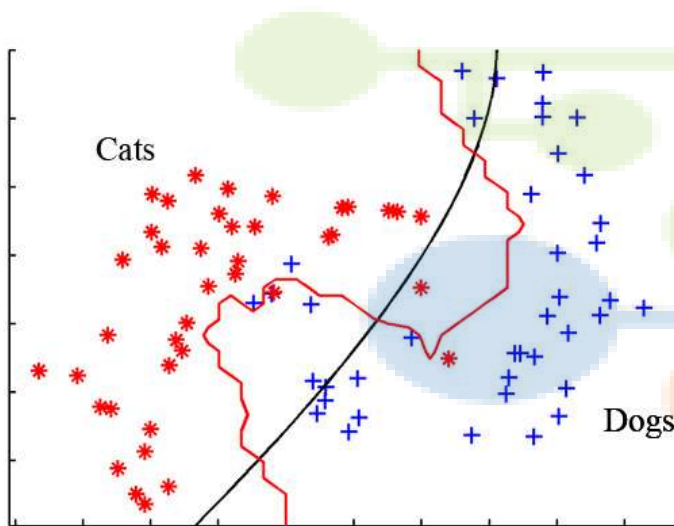


**Data Science  
Academy**

# Conhecendo o Naive Bayes



## Classificação

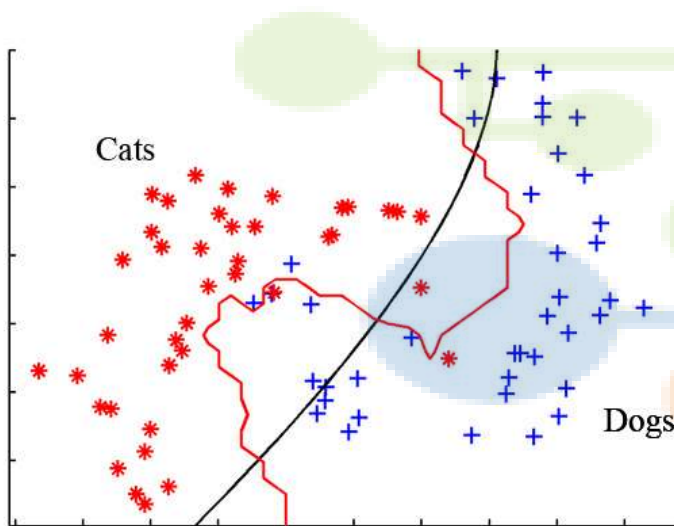


A classificação consiste no processo de encontrar, através de aprendizado de máquina, um modelo ou função que descreva diferentes classes de dados.

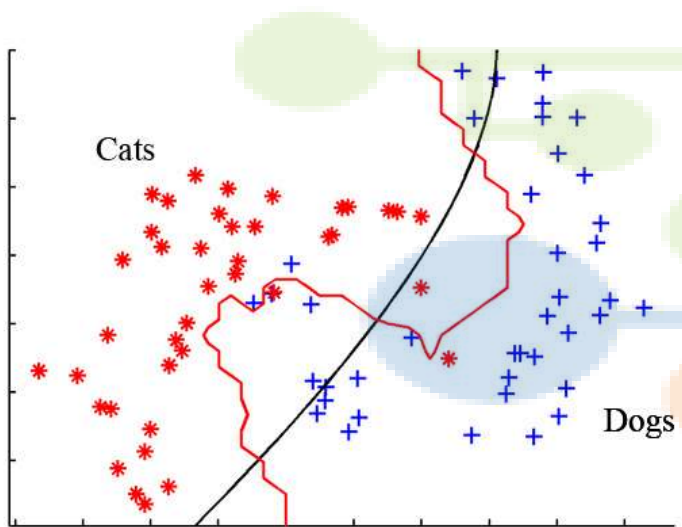




## Classificação



- Detecção de SPAM
- Organização automática de e-mails
- Identificação de páginas com conteúdo adulto
- Detecção de expressões e sentimentos



## Classificador Naïve Bayes



## Classificador Naïve Bayes







## Aplicações do Algoritmo Naive Bayes

- **Previsões multi-classes**
- **Classificação de textos/Filtragem de spam/Análise de sentimento**
- **Previsões em tempo real**
- **Sistema de Recomendação**



## Naive Bayes no Scikit-Learn

Gaussian

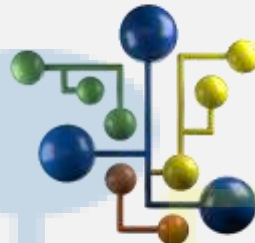
Multinomial

Bernoulli



Data Science  
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science  
Academy**

# A Teoria da Probabilidade



**Probabilidade** é o estudo sobre experimentos que, mesmo realizados em condições bastante parecidas, apresentam resultados que não são possíveis de prever.

Estudamos Probabilidade com a intenção de prever as possibilidades de ocorrência de uma determinada situação ou fato.





## Experimento Aleatório

Um experimento é considerado aleatório quando suas ocorrências podem apresentar resultados diferentes. Um exemplo disso acontece ao lançarmos uma moeda que possua faces distintas, sendo uma cara e outra coroa. O resultado desse lançamento é imprevisível, pois não há como saber qual a face que ficará para cima.





## Espaço Amostral

O espaço amostral ( $S$ ) determina as possibilidades possíveis de resultados. No caso do lançamento de uma moeda o conjunto do espaço amostral é dado por:  $S = \{\text{cara, coroa}\}$ , isso porque são as duas únicas respostas possíveis para esse experimento aleatório.







## Evento

Na probabilidade a ocorrência de um fato ou situação é chamado de evento. Sendo assim, ao lançarmos uma moeda estamos estabelecendo a ocorrência do evento. Temos então que, qualquer subconjunto do espaço amostral deve ser considerado um evento. Um exemplo pode acontecer ao lançarmos uma moeda três vezes, e obtermos como resultado do evento o seguinte conjunto:

$$E = \{\text{Cara, Coroa, Cara}\}$$





## Razão de Probabilidade

A razão de probabilidade é dada pelas possibilidades de um evento ocorrer levando em consideração o seu espaço amostral. Essa razão, que é uma fração, é igual ao número de elementos do evento (numerador) sobre o número de elementos do espaço amostral (denominador).





## Razão de Probabilidade

Considere os seguintes elementos:

$E$  é um evento.

$n(E)$  é o número de elementos do evento.

$S$  é espaço amostral.

$n(S)$  é a quantidade de elementos do espaço amostral.

$$P(E) = \frac{n(E)}{n(S)}$$

Com  $n(S) \neq 0$

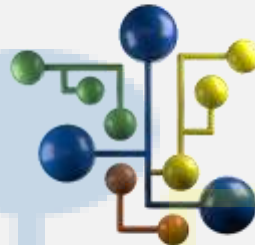
$$0 \leq P(E) \leq 1$$





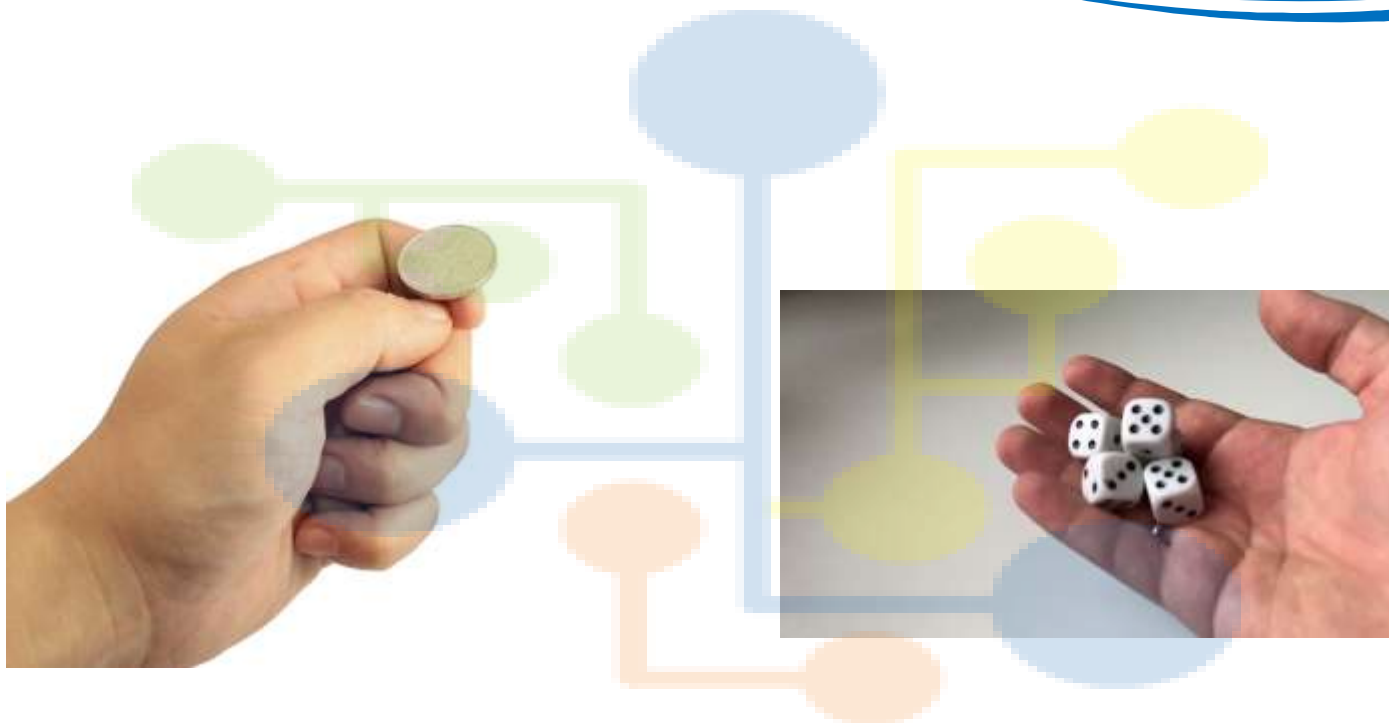
Data Science  
Academy

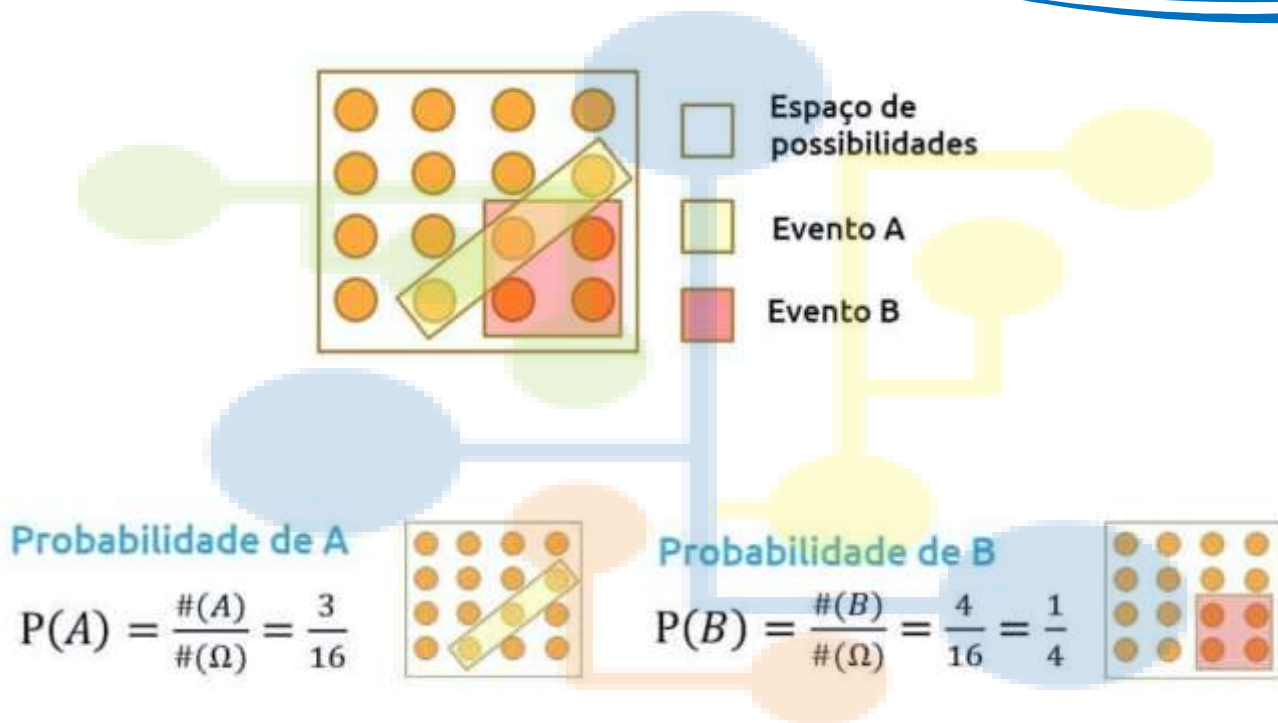
Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



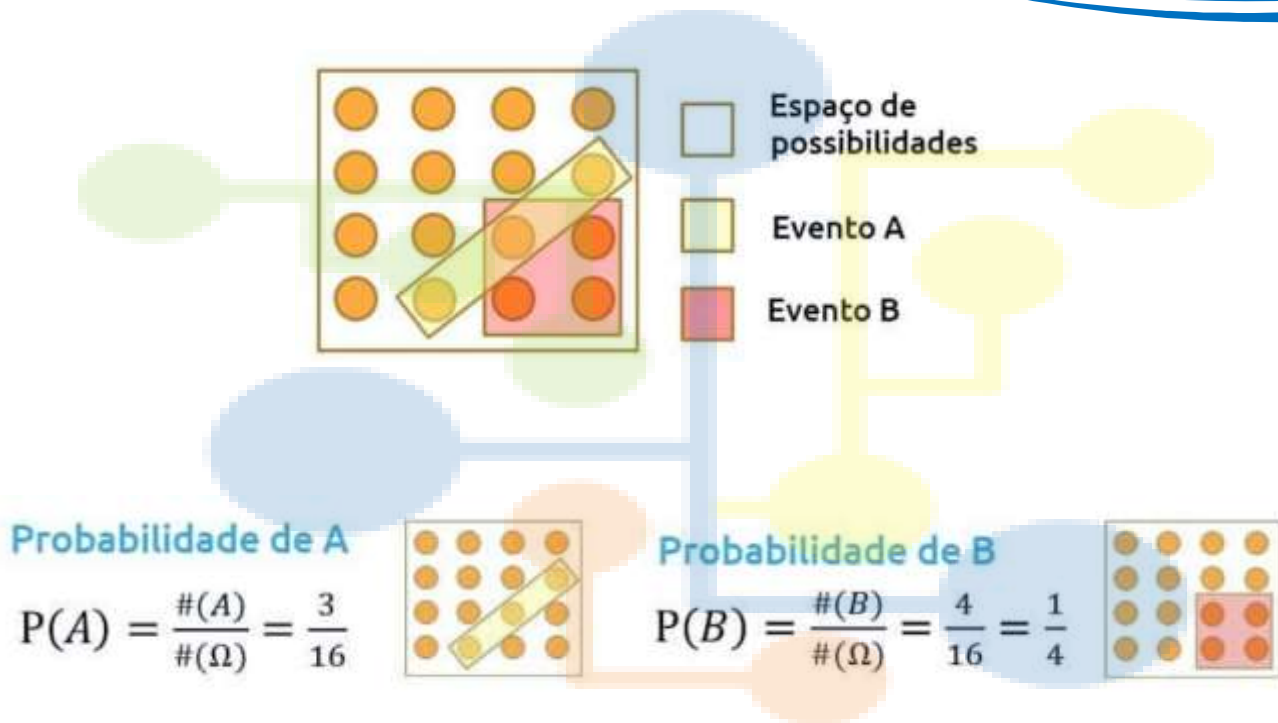
**Data Science  
Academy**

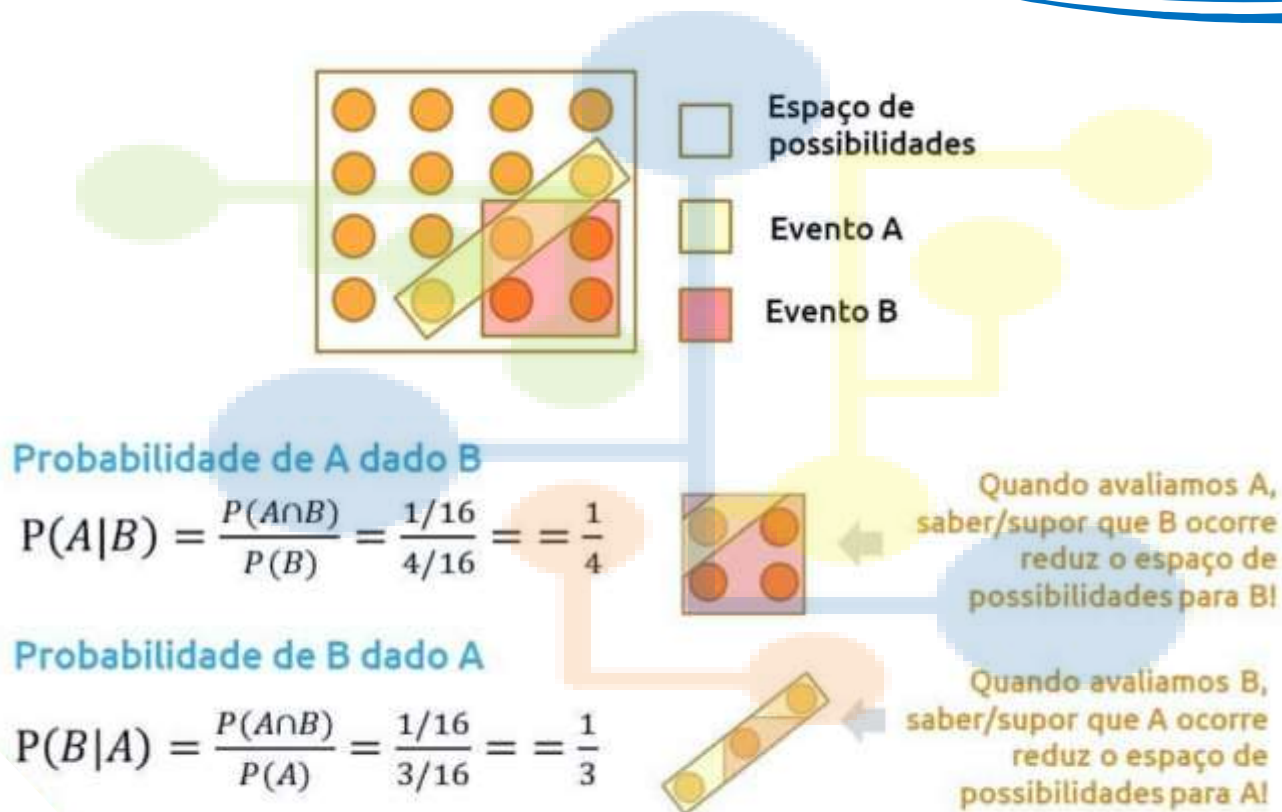
# Probabilidade Condicional







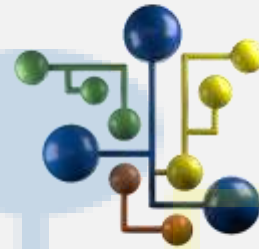






Data Science  
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



**Data Science  
Academy**

Teorema de Bayes



A regra de Bayes mostra como alterar as probabilidades a priori tendo em conta novas evidências de forma a obter probabilidades a posteriori.

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

$P(x)$  e  $P(c)$  são as probabilidades a priori de  $x$  e  $c$ .

$P(c|x)$  e  $P(x|c)$  são as probabilidades a posteriori de  $c$  condicional a  $x$  e de  $x$  condicional a  $c$  respectivamente.



## Teorema de Bayes

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

Probabilidade posterior

Probabilidade original da Classe

Probabilidade Original do Predictor

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

- $P(c | x)$  é a probabilidade posterior da classe (c, alvo) dada o predictor (x, atributos).
- $P(c)$  é a probabilidade original da classe.
- $P(x | c)$  é a probabilidade do predictor dada a classe.
- $P(x)$  é a probabilidade original do predictor.



# Teorema de Bayes

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

Probabilidade posterior

Probabilidade original da Classe

Probabilidade Original do Predictor

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

Como funciona o Teorema de Bayes?



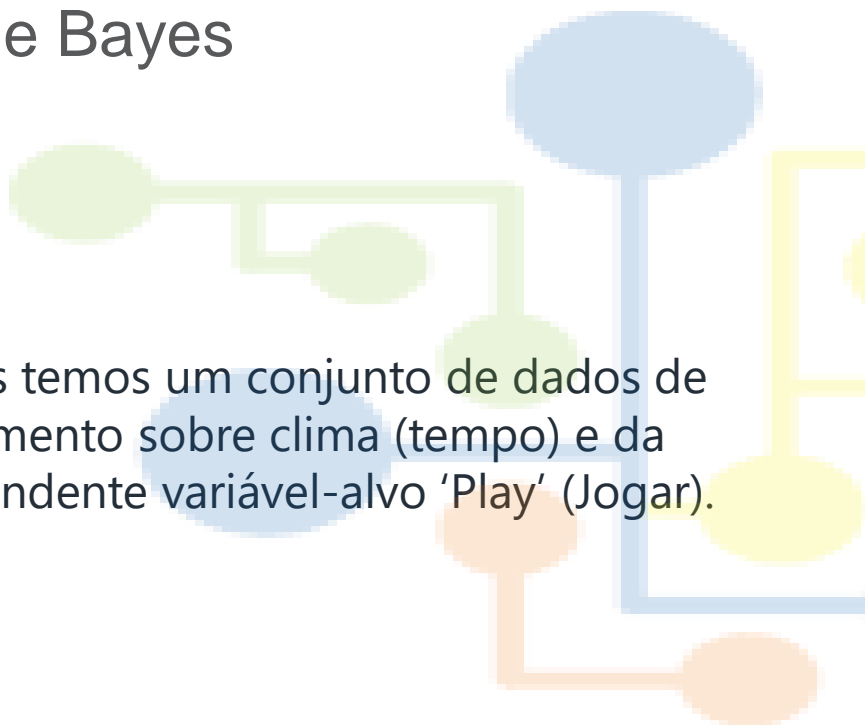


Em teoria da probabilidade o Teorema de Bayes mostra a relação entre uma probabilidade condicional e a sua inversa.



## Teorema de Bayes

Aqui nós temos um conjunto de dados de treinamento sobre clima (tempo) e da correspondente variável-alvo 'Play' (Jogar).



TEMPO	"PLAY"
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Sol	Sim
Chuva	Sim
Sol	Não
Nublado	Sim
Nublado	Sim
Chuva	Não



## Teorema de Bayes

Passo 1: Converter o conjunto de dados em uma tabela de frequência

TEMPO	"PLAY"
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Sol	Sim
Chuva	Sim
Sol	Não
Nublado	Sim
Nublado	Sim
Chuva	Não

Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	5	9



# Teorema de Bayes

Passo 2: Criar tabela de Probabilidade para encontrar as probabilidades de cada ocorrência e de cada combinação.

TEMPO	"PLAY"
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Sol	Sim
Chuva	Sim
Sol	Não
Nublado	Sim
Nublado	Sim
Chuva	Não

Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	5	9

=5/14	=9/14
0.36	0.64

=4/14	0.29
=5/14	0.36
=5/14	0.36



# Teorema de Bayes

Passo 3: Usamos a equação do Teorema de Bayes para calcular a probabilidade posterior para cada classe. A classe com maior probabilidade posterior é o resultado da previsão.

TEMPO	"PLAY"
Sol	Não
Nublado	Sim
Chuva	Sim
Sol	Sim
Sol	Sim
Nublado	Sim
Chuva	Não
Chuva	Não
Sol	Sim
Chuva	Sim
Sol	Não
Nublado	Sim
Nublado	Sim
Chuva	Não

Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	5	9
	=5/14	=9/14
	0.36	0.64

=4/14	0.29
=5/14	0.36
=5/14	0.36

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



## Teorema de Bayes

Os jogadores irão praticar esporte se o tempo estiver ensolarado.  
Esta afirmação está correta?

Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
<b>Total</b>	<b>5</b>	<b>9</b>

=5/14	=9/14
0.36	0.64

=4/14	0.29
=5/14	0.36
=5/14	0.36

$$P(\text{Sim} \mid \text{Sol}) = \frac{P(\text{Sol} \mid \text{Sim}) * P(\text{Sim})}{P(\text{Sol})}$$





## Teorema de Bayes

Os jogadores irão praticar esporte se o tempo estiver ensolarado.  
Esta afirmação está correta?



Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	5	9

Aqui temos:

$$P(\text{Sol} \mid \text{Sim}) = 3/9 = 0.33$$

$$P(\text{Sol}) = 5/14 = 0.36$$



# Teorema de Bayes

Os jogadores irão praticar esporte se o tempo estiver ensolarado.  
Esta afirmação está correta?



Tabela de Frequência		
Clima	Não	Sim
Nublado	0	4
Sol	2	3
Chuva	3	2
Total	5	9

Aqui temos:

$$P(\text{Sol} \mid \text{Sim}) = 3/9 = 0.33$$

$$P(\text{Sol}) = 5/14 = 0.36$$

$$P(\text{Sim}) = 9/14 = 0.64$$



## Teorema de Bayes

Os jogadores irão praticar esporte se o tempo estiver ensolarado.  
Esta afirmação está correta? **Sim, a afirmação está correta!**

Agora, é só colocar na fórmula:

$$P(\text{Sim} \mid \text{Sol}) = P(\text{Sol} \mid \text{Sim}) * P(\text{Sim}) / P(\text{Sol})$$

$$P(\text{Sim} \mid \text{Sol}) = 0.33 * 0.64 / 0.36 = 0.60$$



## Problema da Frequência Zero

- Probabilidade correspondente será zero!
- Probabilidade a posteriori será também zero!



A ideia principal é que a probabilidade de um evento **A** dado um evento **B** (i.e. a probabilidade de alguém ter câncer de mama sabendo, ou dado, que a mamografia deu positivo para o teste) depende não apenas do relacionamento entre os eventos **A** e **B** (i.e., a precisão, ou exatidão, da mamografia), mas também da probabilidade marginal (ou "probabilidade simples") da ocorrência de cada evento.



Data Science  
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



Data Science Academy

## Melhorando o Poder de Classificação do Modelo Naive Bayes





Se os atributos contínuos não têm distribuição normal, devemos usar a transformação ou métodos diferentes para convertê-los em distribuição normal.





Remova variáveis correlacionadas. Os atributos altamente correlacionados podem levar a um excesso de importância de uma característica, reduzindo a capacidade de generalização do modelo.







Classificadores Naive Bayes têm opções limitadas para ajuste de parâmetros, tais como como  $\alpha = 1$  para suavização, `fit_prior = [Verdade | Falso]` para aprendizagem a partir de probabilidades anteriores. Nós recomendamos focar no pré-processamento de dados e seleção de atributos.





Você pode querer aplicar alguma técnica ensemble como “bagging” e “boosting”, mas na prática esses métodos não ajudariam, pois a finalidade destes métodos é reduzir a variância.

Naive Bayes não tem variância para minimizar.





Data Science  
Academy

Data Science Academy rodrigo.c.abreu@hotmail.com 5e207d48e32fc335fa60447d



Data Science Academy

## Vantagens e Desvantagens de Modelos Naive Bayes





## Vantagens

- É fácil e rápido para prever o conjunto de dados da classe de teste. Também tem um bom desempenho na previsão de classes múltiplas.
- Quando a suposição de independência prevalece, um classificador Naive Bayes tem melhor desempenho em comparação com outros modelos como regressão logística, e você precisa de menos dados de treinamento.
- O desempenho é bom em caso de variáveis categóricas de entrada em comparação a variáveis numéricas. Para variáveis numéricas, assume-se a distribuição normal (curva de sino, que é uma suposição forte).





## Desvantagens

- Se a variável categórica tem uma categoria (no conjunto de dados de teste) que não foi observada no conjunto de dados de treinamento, então o modelo irá atribuir uma probabilidade de 0 (zero) e não será capaz de fazer uma previsão. Isso é muitas vezes conhecido como “Zero Frequency”. Para resolver esse problema, podemos usar a técnica de “suavização” (smoothing). Uma das técnicas mais simples de “suavização” (smoothing) é a chamada estimativa de Laplace.
- Uma limitação do Naive Bayes é a suposição de preditores independentes. Na vida real, é quase impossível ter um conjunto de indicadores que sejam completamente independentes.





Continue Trilhando uma Excelente Jornada de Aprendizagem!

**Muito Obrigado!**