

# **Tópicos Especiais em Sistemas de Informação**

**Naive Bayes, probabilidade máximo a posterior e com número aleatório**

**Ely – [elydasilvamiranda@gmail.com](mailto:elydasilvamiranda@gmail.com)**

# Fonte

- Slides baseados em:
  - Curso da Alura:  
<https://cursos.alura.com.br/course/introducao-a-machine-learning-com-classificacao/>
  - <https://www.vooo.pro/insights/6-passos-faceis-para-aprender-o-algoritmo-naive-bayes-com-o-codigo-em-python/>

# Algoritmo Naive Bayes

- É uma técnica de classificação baseado no teorema de Bayes;
- Pressupõe a independência entre as variáveis: uma variável sim/não não condiciona a outra;
- Exemplo:
  - Altura e cor da pele;
  - Estado de origem e renda mensal;
- Útil com grande volumes de dados.

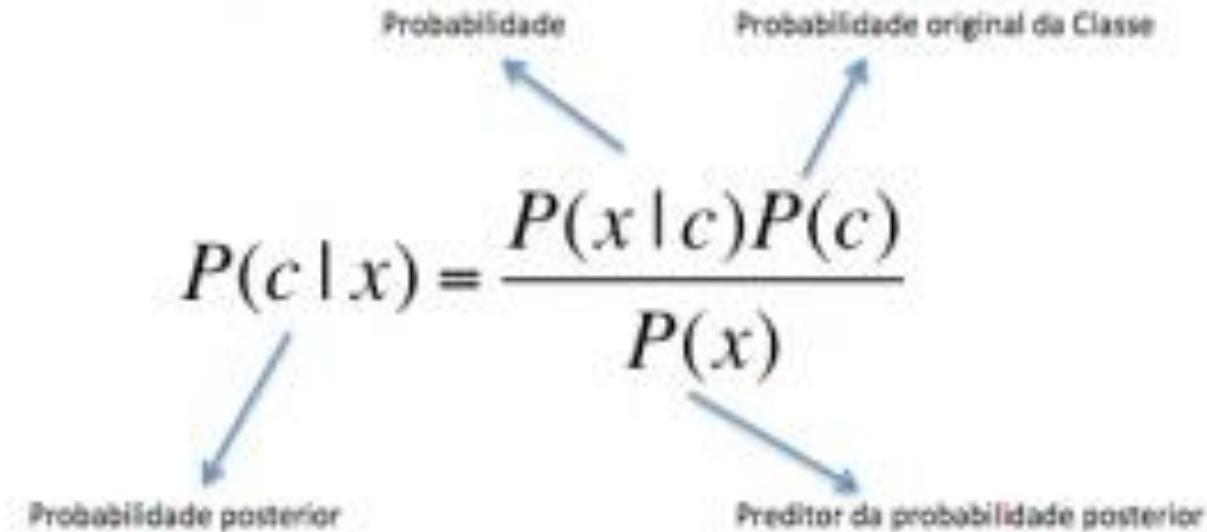
# Teorema de Bayes

- Dados  $c$  e  $x$ :
  - $c$  : um "alvo" ou marcação (classe);
  - $x$  : um dado ou variável (preditor);
- Exemplo:
  - Dado
  - $P(c | x) =$
  - $P(\text{Sim} | \text{Sol})$ : é a probabilidade de haver jogo caso o dia esteja ensolarado.

Tempo	Joga
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

- Fórmula de Bayes:



The diagram shows the formula  $P(c | x) = \frac{P(x | c)P(c)}{P(x)}$  with four blue arrows pointing to its components:  $P(c | x)$  is labeled 'Probabilidade posterior',  $P(x | c)$  is labeled 'Probabilidade',  $P(c)$  is labeled 'Probabilidade original da Classe', and  $P(x)$  is labeled 'Preditores da probabilidade posterior'.

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

Probabilidade posterior

Probabilidade

Probabilidade original da Classe

Preditores da probabilidade posterior

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

# Teorema de Bayes

- No slide anterior, temos:
  - $P(c | x)$  é a probabilidade posterior da classe ( $c$ , alvo) dada preditor ( $x$ , atributos);
  - $P(c)$  é a probabilidade original da classe;
  - $P(x | c)$  é a probabilidade que representa a probabilidade de preditor, dada a classe;
  - $P(x)$  é a probabilidade original do preditor.

# Calculando com Naive Bayes

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

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

# Calculando com Naive Bayes

- Passo 2: Calcular as probabilidades de cada clima e do "sim"/"não"

Clima	Não	Sim	
Nublado	0	4	$4/14 = 0,29$
Sol	2	3	$5/14 = 0,36$
Chuva	2	3	$5/14 = 0,36$
Total	5	9	
	$5/14 = 0,36$	$9/14 = 0,64$	



# Calculando com Naive Bayes

- Passo 3:
  - Usar a equação Bayesiana Naive para calcular a probabilidade posterior para cada classe:

$$P(\text{Sim} \mid \text{Sol})$$

$$P(\text{Sim} \mid \text{Chuva})$$

$$P(\text{Sim} \mid \text{Nublado})$$

$$P(\text{Não} \mid \text{Sol})$$

...

- A classe com maior probabilidade posterior é o resultado da previsão.

# Exemplo

- $P(\text{Sim} \mid \text{Sol}) = ?$ 
  - $P(\text{Sim} \mid \text{Sol}) = P(\text{Sol} \mid \text{Sim}) * P(\text{Sim}) / P(\text{Sol}) =$
  - $(3/9) * 0,64 / 0,36 = 0,60$
- Ou seja:
  - Há 60% de chances de haver jogo com Sol e 40% de chances de não haver jogo;
  - 60% prevalece.

# Prós

- É fácil e rápido para prever o conjunto de dados da classe de teste e de múltiplas;
- Tem melhor desempenho em comparação com outros modelos;
- Exige menos dados de treinamento;

# Contras

- Se a variável categórica tem uma categoria 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”.
- Nem sempre as variáveis são independentes;
  - Na vida real, é quase impossível que ter indicadores que sejam completamente isolados.

# Probabilidade máximo a posterior

- É a forma de escolha vista anteriormente;
- Nela, optamos pela maior probabilidade para prever resultados futuros;
- Exemplo:
  - Para uma variável simples "quantidade de votos"...
  - Perguntamos dentre dois candidatos quem ganhará uma eleição:

Quem vai ganhar	Total
Guilherme	70
Paulo	30

# Probabilidade máximo a posterior

- Ao treinarmos um algoritmo com esses dados, ele pode entender apenas essa probabilidade;
- Caso seja perguntado novamente o algoritmo pode responder que Guilherme seria eleito.

# Probabilidade com número aleatório

- Uma melhoria seria sortear números aleatórios entre 0 e 100;
- Se o número for menor ou igual a 70, significa que ele faz parte de 70% de chance (Guilherme);
- Caso contrário, entre 71 a 100, o voto seria para o Paulo que corresponde aos 30%;

Quem vai ganhar	Probabilidade	Aleatório
Guilherme	70%	1 até 70
Paulo	30%	71 a 100

# Probabilidade com número aleatório

- O mesmo exemplo com valores diferentes:

Quem vai ganhar	Total	Probabilidade	Aleatório
Guilherme	77	65	1 até 65
Paulo	42	35	66 a 100

- Essa regra funciona de acordo com as probabilidades;
- Não predizemos apenas com uma possibilidade;
- Existe a chance de escolher Guilherme ou Paulo, mesmo com maior probabilidade para Guilherme.



# Algoritmos de Machine Learning

- O comportamento básico de um algoritmo é:
  - Pegar uma quantidade de dados brutos;
  - Calcular as probabilidades;
  - Utilizar uma regra de decisão;
- Esses passos são bem similares ao que fazemos inconscientemente:
  - Vemos diversos animais e esses animais possuem pernas curtas, são rosinhas, não fazem *auau...*
  - Aprendemos que para as características, a maioria são porcos e a minoria são cachorrinhos;
  - Calculamos a probabilidade na nossa cabeça e concluimos se é um porco ou um cachorro.

# Algoritmos e números aleatórios

- Na classificação de e-mails entre *spam* e não *spam* também fazemos algo bem similar;
- Usamos um dos 3 critérios que vimos até agora quando nos deparamos com um texto do *subject*:
  - Se o texto é comum em *spams*, imediatamente, sem outros critérios, concluimos que é *spam*;
  - Mesmo sabendo que o *subject* é de um *spam*:
    - Por escolha aleatória, decidimos abrir esse e-mail;
    - Nesse instante usamos a situação em que sorteamos um número;
    - Então, verificamos em qual probabilidade ele se encaixa.

# Algoritmos e números aleatórios

- Um outro exemplo são as ações da bolsa de valores;
- Avaliamos se a cotação de uma ação vai subir ou descer:
  - Acontece um evento comum que, na maioria das vezes, a ação subia:
    - Pelo método do maior valor, investiremos nessa ação, pois acreditamos que ela subirá;
  - Pode também acontecer um evento que em alguns momentos ela subiu ou desceu:
    - Utilizamos o critério da probabilidade,
    - Sorteamos um número na nossa cabeça e escolhemos se vamos ou não investir na ação.

# Probabilidades condicionais

- Em um site de venda de imóveis, foi observada o estado de origem dos clientes que compraram;
- Dado o estado temos a informação de quantos compraram ou não:

RJ	Total	Probabilidade
Comprou	37	68,5%
Não comprou	17	31,5%

SP	Total	Probabilidade
Comprou	67	27,5%
Não comprou	117	72,5%

# Probabilidades condicionais

- Temos agora duas variáveis:
  - **localidade**: estado que o cliente mora;
  - **compra**: se comprou ou não;
- Podemos analisar duas informações e prever se o cliente comprará ou não;

RJ	Total	Probabilidade
Comprou	37	68,5%
Não comprou	17	31,5%

SP	Total	Probabilidade
Comprou	67	27,5%
Não comprou	117	72,5%

# Probabilidades condicionais

- Utilizando o método *maximum a posteriori*:
  - Um cliente do Rio de Janeiro, compra ou não compra?
    - 68,5% compram e 31,5% não, em outras palavras, os clientes do Rio de Janeiro irão comprar.
  - E os clientes de São Paulo?
    - 72,5% não compram e 27,5% sim, logo, não irão comprar.
- Esse método parece apropriado?
- Será que o método com probabilidade com número aleatório não seria mais interessante?

# Mais de uma variável condicional

- Adicionando mais uma variável condicional:
  - O cliente ganha ou não mais de R\$ 5.000,00?
- Assim, teremos quatro tabelas:

RJ	Total	Probabilidade
Comprou	37	68,5%
Não comprou	17	31,5%

\$ > 5000	Total	Probabilidade
Comprou	32	80%
Não comprou	8	20%

SP	Total	Probabilidade
Comprou	67	27,5%
Não comprou	117	72,5%

\$ < 5000	Total	Probabilidade
Comprou	47	18%
Não comprou	211	82%

# Mais de uma variável condicional

- As variáveis condicionais são independentes:
  - a probabilidade de ambas acontecerem é a multiplicação das probabilidades individuais;
- Qual a probabilidade de  $P(\text{Comprar} \mid \text{SP}, \leq 5000)$ ?
  - $P(\text{Comprar} \mid \text{SP}) * P(\text{Comprar} \mid \leq 5000)$
  - $27,5/100 * 18/100 = 0,0495 = 5\%$
- Assim:
  - Pela regra máximo a posterior: não comprará;
  - Pela regra da probabilidade: número aleatórios e de 1 a 5, comprará... De 6 a 100, não comprará<sub>26</sub>



# **Tópicos Especiais em Sistemas de Informação**

**Naive Bayes, probabilidade máximo a posterior e com número aleatório**

**Ely – [elydasilvamiranda@gmail.com](mailto:elydasilvamiranda@gmail.com)**