



# Classificação: Naive Bayes

---

Fabrcio Olivetti de Franca

Universidade Federal do ABC

## 1. Naive Bayes

# Naive Bayes

---

Foi mencionado anteriormente que o uso da função logística remetia a probabilidade de certo exemplo pertencer a uma classe.

Por que então não estimar diretamente uma função de probabilidade?

A ideia é estimar  $P(y = y_i | x_i)$

O algoritmo Naive Bayes (que pode ser traduzido como Bayes “Ingênuo”) é um classificador probabilístico que utiliza o teorema de Bayes assumindo independência nos atributos do objeto.

Digamos que nossos objetos representam animais, e estamos analisando 4 características:

- Tamanho: pequeno ou grande,
- Pele: lisa ou peluda,
- Cor: marrom, verde ou vermelho,
- Carne: macia ou dura

Pele	Cor	Tamanho	Carne
peludo	marrom	grande	dura
peludo	verde	grande	dura
liso	vermelho	grande	macia
peludo	verde	grande	macia
peludo	vermelho	pequeno	dura
liso	vermelho	pequeno	dura
liso	marrom	pequeno	dura
peludo	verde	pequeno	macia
liso	verde	pequeno	dura
peludo	vermelho	grande	dura
liso	marrom	grande	macia
liso	verde	pequeno	macia
peludo	vermelho	pequeno	macia
liso	vermelho	grande	dura
liso	vermelho	pequeno	dura
peludo	verde	pequeno	dura

Digamos que queremos saber se é seguro ou perigoso comer certo animal, baseado em suas características.

O primeiro passo é pré-classificar uma pequena amostra dos dados.



Pele	Cor	Tamanho	Carne	Classe
peludo	marrom	grande	dura	seguro
peludo	verde	grande	dura	seguro
liso	vermelho	grande	macia	perigoso
peludo	verde	grande	macia	seguro
peludo	vermelho	pequeno	dura	seguro
liso	vermelho	pequeno	dura	seguro
liso	marrom	pequeno	dura	seguro
peludo	verde	pequeno	macia	perigoso
liso	verde	pequeno	dura	perigoso
peludo	vermelho	grande	dura	seguro
liso	marrom	grande	macia	seguro
liso	verde	pequeno	macia	perigoso
peludo	vermelho	pequeno	macia	seguro
liso	vermelho	grande	dura	perigoso
liso	vermelho	pequeno	dura	
peludo	verde	pequeno	dura	

Com uma base pré-classificada, devemos formular a seguinte questão:

*Dado um pequeno animal que tem pele lisa de cor vermelha e carne dura.  
É seguro come-lo?*

Nosso animal pode ser representado por:

$$x = [\text{pequeno}, \text{lisa}, \text{vermelho}, \text{dura}]$$

E a pergunta, em forma de probabilidade condicional:

$$P(y = \text{seguro} \mid x)$$

Essa probabilidade deve ser lida como:

Qual a probabilidade de um animal ser seguro para comer dado que ele é:  
*pequeno, lisa, vermelho, dura?*

Pelo Teorema de Bayes, temos que:

$$P(y = \text{seguro} \mid x) = \frac{P(x \mid y = \text{seguro}) \cdot P(y = \text{seguro})}{P(x)},$$

com  $P(x \mid y = \text{seguro})$  sendo denominado *likelihood*,  $P(y = \text{seguro})$  é o *prior* e  $P(x)$  o preditor.

Com essa fórmula, e a tabela de exemplos podemos calcular a probabilidade de cada animal não classificado pertencer a uma classe ou outra.

Expandindo temos:

$$P(y = \text{seguro} \mid x) = \frac{\prod_i P(x_i \mid y = \text{seguro}) \cdot P(y = \text{seguro})}{\prod_i P(x_i)}$$

Dado  $x = [\text{pequeno}, \text{lisa}, \text{vermelho}, \text{dura}]$ :

$$P(\text{pequeno}) =$$

$$P(\text{lisa}) =$$

$$P(\text{vermelho}) =$$

$$P(\text{dura}) =$$

Pele	Cor	Tamanho	Carne	Classe
peludo	marrom	grande	<b>dura</b>	seguro
peludo	verde	grande	<b>dura</b>	seguro
<b>liso</b>	<b>vermelho</b>	grande	macia	perigoso
peludo	verde	grande	macia	seguro
peludo	<b>vermelho</b>	<b>pequeno</b>	<b>dura</b>	seguro
<b>liso</b>	<b>vermelho</b>	<b>pequeno</b>	<b>dura</b>	seguro
<b>liso</b>	marrom	<b>pequeno</b>	<b>dura</b>	seguro
peludo	verde	<b>pequeno</b>	macia	perigoso
<b>liso</b>	verde	<b>pequeno</b>	<b>dura</b>	perigoso
peludo	<b>vermelho</b>	grande	<b>dura</b>	seguro
<b>liso</b>	marrom	grande	macia	seguro
<b>liso</b>	verde	<b>pequeno</b>	macia	perigoso
peludo	<b>vermelho</b>	<b>pequeno</b>	macia	seguro
<b>liso</b>	<b>vermelho</b>	grande	<b>dura</b>	perigoso
liso	vermelho	pequeno	dura	
peludo	verde	pequeno	dura	



Dado  $x = [\text{pequeno}, \text{lisa}, \text{vermelho}, \text{dura}]$ :

$$P(\text{pequeno}) = 7/14$$

$$P(\text{lisa}) = 7/14$$

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

$$P(\text{dura}) = 8/14$$

$$P(x) = 1960/38416 = 0.05$$

E, finalmente:

$$P(\text{pequeno} \mid y = \text{seguro}) =$$

$$P(\text{lisa} \mid y = \text{seguro}) =$$

$$P(\text{vermelho} \mid y = \text{seguro}) =$$

$$P(\text{dura} \mid y = \text{seguro}) =$$

Pele	Cor	Tamanho	Carne	Classe
peludo	marrom	grande	<b>dura</b>	seguro
peludo	verde	grande	<b>dura</b>	seguro
liso	vermelho	grande	macia	perigoso
peludo	verde	grande	macia	seguro
peludo	<b>vermelho</b>	<b>pequeno</b>	<b>dura</b>	seguro
<b>liso</b>	<b>vermelho</b>	<b>pequeno</b>	<b>dura</b>	seguro
<b>liso</b>	marrom	<b>pequeno</b>	<b>dura</b>	seguro
peludo	verde	pequeno	macia	perigoso
liso	verde	pequeno	dura	perigoso
peludo	<b>vermelho</b>	grande	<b>dura</b>	seguro
<b>liso</b>	marrom	grande	macia	seguro
liso	verde	pequeno	macia	perigoso
peludo	<b>vermelho</b>	<b>pequeno</b>	macia	seguro
liso	vermelho	grande	dura	perigoso
liso	vermelho	pequeno	dura	
peludo	verde	pequeno	dura	

E, finalmente:

$$P(\text{pequeno} \mid y = \text{seguro}) = 4/9$$

$$P(\text{lisa} \mid y = \text{seguro}) = 3/9$$

$$P(\text{vermelho} \mid y = \text{seguro}) = 4/9$$

$$P(\text{dura} \mid y = \text{seguro}) = 6/9$$

$$P(x \mid y = \text{seguro}) = 288/6561 = 0.04$$

Pele	Cor	Tamanho	Carne	Classe
peludo	marrom	grande	dura	seguro
peludo	verde	grande	dura	seguro
<b>liso</b>	<b>vermelho</b>	grande	macia	perigoso
peludo	verde	grande	macia	seguro
peludo	vermelho	pequeno	dura	seguro
liso	vermelho	pequeno	dura	seguro
liso	marrom	pequeno	dura	seguro
peludo	verde	<b>pequeno</b>	macia	perigoso
<b>liso</b>	verde	<b>pequeno</b>	<b>dura</b>	perigoso
peludo	vermelho	grande	dura	seguro
liso	marrom	grande	macia	seguro
<b>liso</b>	verde	<b>pequeno</b>	macia	perigoso
peludo	vermelho	pequeno	macia	seguro
<b>liso</b>	<b>vermelho</b>	grande	<b>dura</b>	perigoso
liso	vermelho	pequeno	dura	
peludo	verde	pequeno	dura	

E, finalmente:

$$P(\text{pequeno} \mid y = \text{perigoso}) = 3/5$$

$$P(\text{lisa} \mid y = \text{perigoso}) = 4/5$$

$$P(\text{vermelho} \mid y = \text{perigoso}) = 2/5$$

$$P(\text{dura} \mid y = \text{perigoso}) = 2/5$$

$$P(x \mid y = \text{perigoso}) = 48/625 = 0.08$$

Com isso podemos calcular:

$$P(y = \text{seguro} \mid x) = 0.04 \cdot 0.64 / 0.05 = 0.51$$

$$P(y = \text{perigoso} \mid x) = 0.08 \cdot 0.36 / 0.05 = 0.58$$

Note que  $P(y = \text{seguro} \mid x) \neq 1 - P(y = \text{perigoso} \mid x)$

Existem alguns pontos nesse algoritmo a serem observados:

- $P(x_i)$  é um denominador comum para todas as probabilidades, como queremos obter a maior probabilidade, podemos omitir.
- Se um certo atributo categórico nunca apareceu na base de dados, sua probabilidade será 0 e teremos um resultado indefinido.
- Se nosso vetor de atributos é muito grande, o termo  $P(x_i | y = y_i)$  tende a zero.



Para lidarmos com um novo atributo utilizamos o *add-one smoothing*, e toda estimativa de probabilidade é calculada como:

$$P(x) = \frac{f(x) + 1}{n + d},$$

com  $f(x)$  sendo a frequência de  $x$  na base,  $n$  o número de amostras/objetos na base,  $d$  é o número de valores distintos daquele atributo.

Dessa forma um novo atributo terá sempre a probabilidade de  $1/(n + d)$ .  
Isso indica uma pequena probabilidade de um atributo novo aparecer em uma nova amostra.

Para evitar que alguma produtória se torne zero, também conhecido como *underflow*, podemos aplicar uma transformação na nossa fórmula da probabilidade de tal forma que a relação entre duas probabilidades se mantenha, ou seja:

$$P(A) < P(B) \rightarrow g(P(A)) < g(P(B))$$

$$P(A) > P(B) \rightarrow g(P(A)) > g(P(B))$$

$$P(A) = P(B) \rightarrow g(P(A)) = g(P(B))$$

Uma escolha para a função  $g(\cdot)$  é o logaritmo.

Aplicando o logaritmo em nossa função de probabilidade, temos:

$$\log\text{-prob}(y = \text{seguro} \mid x) = \sum_i \log(P(x_i \mid y = \text{seguro})) + \log(P(y = \text{seguro}))$$

Que é muito mais simples de ser calculado.

O algoritmo Naive Bayes deve ser implementado primeiro pré-calculando a tabela de frequências dos atributos condicionados a classe e das classes. A ideia é gerar tabelas de hash com a chave sendo o atributo desejado e o valor a frequência.

- Para a tabela de classes, a chave será cada uma das classes. Ex.: ("seguro", 10.0).
- Para a tabela de atributos, a chave será composta por índice do atributo, nome do atributo e classe. Ex.: ((2, "pequeno", "perigoso"), 15.0).
- Já sabemos calcular a frequência de valores categóricos. Basta implementarmos uma função para gerar as transformações da base na forma que desejamos:

# Naive Bayes

```
from collections import Counter

def NaiveBayes(X,y):
    xtuplas = [(j, xij, yi)
                for (xi,yi) in zip(X,y)
                for (j,xij) in enumerate(xi)
                ]
    n = len(y)
    fxy = Counter(xtuplas)
    fy = Counter(y)

    py = {k : v/n for k,v in fy.items()}
    pxy = {k : v/fy[k[2]] for k,v in fxy.items()}
    return py, pxy
```

Com essa informação, a classe de um dado exemplo  $x_i$  pode ser calculado como:

```
logp = {}  
for yi in fy.keys():  
    loglikelihood = sum([np.log(pxy[(j, xij, yi)])  
                        for (j,xij) in enumerate(xi)])  
    logprior = np.log(py[yi])  
    logp[yi] = loglikelihood + logprior  
  
yhat = argmaxDict(logp) # essa funcao nao existe no Python
```



Na próxima aula aprenderemos sobre:

- Árvore de Decisão para Regressão e Classificação.