

Mineração de Dados

Trabalho 7

Arthur do Prado Labaki - 11821BCC017

03-05, 2023

GBC212

Estudo do Conceito

O Classificador Naive Bayes (também conhecido como Classificador Ingênuo de Bayes) é um algoritmo de aprendizado de máquina que usa o Teorema de Bayes para classificar dados. Ele é chamado de "ingênuo" porque assume que as características dos dados são independentes entre si, o que é geralmente uma simplificação bastante grosseira, mas muitas vezes útil na prática.

O algoritmo funciona calculando a probabilidade de uma instância de dados pertencer a uma determinada classe, dada a probabilidade de cada característica do dado ocorrer em cada classe. Ele usa o Teorema de Bayes para calcular essa probabilidade. O Teorema de Bayes afirma que a probabilidade de uma hipótese ser verdadeira (por exemplo, a probabilidade de uma instância de dados pertencer a uma determinada classe) pode ser calculada a partir da probabilidade a priori da hipótese (a probabilidade da classe antes de vermos os dados) e da probabilidade condicional dos dados, dado que a hipótese é verdadeira (a probabilidade da característica ocorrer dado que sabemos que a classe é verdadeira).

Resolução do Exercício 1)

Dado a base de dados abaixo, classifique os seguintes objetos usando o classificador ingênuo de Bayes:

1. Ensolarado, Fria, Alta, Forte
2. Nublado, Fria, Alta, Forte

Para resolvermos os exercícios, precisamos utilizar a fórmula do classificador ingênuo de Bayes, que é:

$$P(Classe \mid Atributos) = P(Atributos \mid Classe) * P(Classe) / P(Atributos)$$

Dia	Panorama	Temperatura	Umidade	Vento	Jogar Tênis
1	Ensolarado	Quente	Alta	Fraco	Não
2	Ensolarado	Quente	Alta	Forte	Não
3	Nublado	Quente	Alta	Fraco	Sim
4	Chuvoso	Intermediária	Alta	Fraco	Sim
5	Chuvoso	Fria	Normal	Fraco	Sim
6	Chuvoso	Fria	Normal	Forte	Não
7	Nublado	Fria	Normal	Forte	Sim
8	Ensolarado	Intermediária	Alta	Fraco	Não
9	Ensolarado	Fria	Normal	Fraco	Sim
10	Chuvoso	Intermediária	Normal	Fraco	Sim
11	Ensolarado	Intermediária	Normal	Forte	Sim
12	Nublado	Intermediária	Alta	Forte	Sim
13	Nublado	Quente	Normal	Fraco	Sim
14	Chuvoso	Intermediária	Alta	Forte	Não

Figura 1: Base de dados

Com isso, para o item 1, é necessário realizar a operação $P(JogarTênis \mid Panorama = Ensolarado, Temperatura = Fria, Umidade = Alta, Vento = Forte)$. Então:

$$\begin{aligned}
&P(JogarTênis \mid Panorama = Ensolarado, Temperatura = Fria, Umidade = Alta, \\
&Vento = Forte) = \\
&P(Panorama = Ensolarado \mid JogarTênis) * P(Temperatura = Fria \mid JogarTênis) * \\
&P(Umidade = Alta \mid JogarTênis) * P(Vento = Forte \mid JogarTênis) * \\
&P(JogarTênis) / P(Tempo = Ensolarado) * P(Temperatura = Fria) * P(Umidade = Alta) * \\
&P(Vento = Forte)
\end{aligned}$$

Separando a equação em partes, temos:

$$P(Panorama = Ensolarado \mid JogarTênis) = 2/9$$

$$P(Temperatura = Fria \mid JogarTênis) = 3/9$$

$$P(Umidade = Alta \mid JogarTênis) = 3/9$$

$$P(Vento = Forte \mid JogarTênis) = 3/9$$

Para esses, seria das probabilidades de jogar tênis, as que ocorrem o atributo.

$$P(JogarTênis) = 9/14$$

$$P(Panorama = Ensolarado) = 5/14$$

$$P(Temperatura = Fria) = 4/14$$

$$P(Umidade = Alta) = 7/14$$

$$P(Vento = Forte) = 6/14$$

Já para esses, seria dos dias disponíveis na base de dados, quais deles ocorreram o evento.

Realizando a conta, temos que $(2/9) * (3/9) * (3/9) * (3/9) * (9/14) / (5/14) * (4/14) * (7/14) * (6/14) = 0.241$

Agora é necessário refazer essas contas usando o atributo não jogar tênis, em vez do jogar.

Com isso temos: $P(\text{Panorama} = \text{Ensolarado} \mid \text{NãoJogarTênis}) = 3/5$

$P(\text{Temperatura} = \text{Fria} \mid \text{NãoJogarTênis}) = 1/5$

$P(\text{Umidade} = \text{Alta} \mid \text{NãoJogarTênis}) = 4/5$

$P(\text{Vento} = \text{Forte} \mid \text{NãoJogarTênis}) = 3/5$

$P(\text{NãoJogarTênis}) = 5/14$

Realizando a conta, temos que $(3/5) * (1/5) * (4/5) * (3/5) * (5/14) / (5/14) * (4/14) * (7/14) * (6/14) = 0.9408$

Comparando as duas probabilidades, não jogar é mais provável que jogar, então é mais provável que esse objeto pertence à classe "Não" (não jogar tênis) de acordo com o classificador ingênuo de Bayes.

Por fim, para o ultimo item, é necessário realizar a operação $P(\text{JogarTênis} \mid \text{Panorama} = \text{Nublado}, \text{Temperatura} = \text{Fria}, \text{Umidade} = \text{Alta}, \text{Vento} = \text{Forte})$. Como a única variável que foi alterada do item 1 foi o panorama, será calculado somente as variáveis atreladas a ela, que são:

$P(\text{Panorama} = \text{Nublado} \mid \text{JogarTênis}) = 4/9$

$P(\text{Panorama} = \text{Nublado}) = 4/14$

$P(\text{Panorama} = \text{Nublado} \mid \text{NãoJogarTênis}) = 0/5$

Com isso temos que:

Jogar: $(4/9) * (3/9) * (3/9) * (3/9) * (9/14) / (4/14) * (4/14) * (7/14) * (6/14) = 0.604$

Não jogar: $(0/5) * (1/5) * (4/5) * (3/5) * (5/14) / (4/14) * (4/14) * (7/14) * (6/14) = 0$

Comparando as duas probabilidades, jogar é mais provável que não jogar, então é mais provável que esse objeto pertence à classe "Sim" (jogar tênis) de acordo com o classificador ingênuo de Bayes. Ainda o não jogar tem probabilidade 0, indicando que é praticamente impossível de que esse evento se realize nessas condições (pois sempre que está nublado, de acordo com a base de dados, ele joga tênis).