Naive Bayes

Teorema de Bayes

O teorema de Bayes é uma ferramenta matemática que permite calcular a probabilidade de um evento dado que outro evento ocorreu. Ele é baseado na ideia de que a probabilidade de um evento pode ser calculada a partir de informações sobre outros eventos relacionados.

A fórmula básica do teorema de Bayes é a seguinte:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Onde:

- P(A|B) é a probabilidade de A dado que B ocorreu (probabilidade condicional)
- P(B|A) é a probabilidade de B dado que A ocorreu (probabilidade condicional)
- P(A) é a probabilidade de A ocorrer (probabilidade marginal)
- P(B) é a probabilidade de B ocorrer (probabilidade marginal)

Exemplo:
Imagine que você tem um teste de câncer com 90% de acurácia e 1% da população tem essa doença.

Se um indivíduo testa positivo para o câncer, qual é a probabilidade de que ele realmente tenha o câncer?

Para responder a essa pergunta, precisamos calcular a probabilidade de ter câncer dado que o teste deu positivo. Isso é dado por:

$$P(Cancer|Positivo) = \frac{P(Positivo|Cancer) \times P(Cancer)}{P(Positivo)}$$

Para calcular essa probabilidade, precisamos conhecer as probabilidades condicionais e marginais.

- $P(Positivo|Cancer)$ é a probabilidade de testar positivo para o câncer dado que a pessoa tem a doença. Nesse caso, é dado como 90%.
- $P(Cancer)$ é a probabilidade de ter câncer na população, que é dada como 1%.
- $P(Positivo)$ é a probabilidade de testar positivo independentemente de ter câncer ou não. Essa probabilidade é dada pela probabilidade de ter câncer e testar positivo $(P(Cancer) \times P(Positivo|Cancer))$ mais a probabilidade de não ter câncer e testar positivo $(P(Sem\ Câncer) \times P(Positivo|Sem\ Câncer))$

então:

$$P(Positivo) = P(Câncer) \times P(Positivo|Câncer) + P(Sem\ Câncer) \times P(Positivo|Sem\ Câncer)$$

substituindo os valores:

$P(Positivo) = 0.01 \times 0.9 + 0.99 \times 0.1$ (considerando que a acurácia do teste para detectar pessoas sem câncer é de 10%)

Agora que temos todos os valores, podemos calcular a probabilidade de ter câncer dado que o teste deu positivo:

$$P(Câncer|Positivo) = \frac{P(Positivo|Câncer) \times P(Câncer)}{P(Positivo)}$$

$$P(Câncer|Positivo) = \frac{0.9 \times 0.01}{0.01 \times 0.9 + 0.99 \times 0.1}$$

$$P(Câncer|Positivo) = 0.0833\ ou\ 8.33\%$$

Portanto, mesmo que o teste tenha dado positivo, existe apenas uma chance de 8.33% de que a pessoa tenha realmente câncer. É importante lembrar que testes médicos, mesmo os mais precisos, nunca são 100% precisos e é importante considerar essas probabilidades ao interpretar os resultados.

Naive Bayes

O algoritmo Naive Bayes é um algoritmo de classificação probabilístico baseado na Teoria de Bayes. Ele é chamado de "naive" porque assume que todas as características presentes no conjunto de dados são independentes entre si, o que nem sempre é o caso na vida real.

A idéia básica do Naive Bayes é calcular a probabilidade de uma determinada classe dado um conjunto de características, e então escolher a classe com a maior probabilidade. Isso é feito utilizando a equação de Bayes, que relaciona a probabilidade de um evento dado outro evento.

Para usar o Naive Bayes para classificação, primeiro é preciso treinar o algoritmo com um conjunto de dados de treinamento. Isso envolve calcular a probabilidade de cada classe, bem como a probabilidade de cada característica para cada classe. Essas probabilidades são então armazenadas para uso posterior.

Quando se deseja classificar um novo conjunto de dados, o algoritmo usa as probabilidades armazenadas para calcular a probabilidade de cada classe dado o conjunto de características. A classe com a maior probabilidade é então escolhida como a classificação para esse conjunto de dados.

O algoritmo Naive Bayes é bastante rápido e fácil de implementar, e é frequentemente usado em aplicações de classificação de texto, como filtragem de spam de e-mail. É uma boa escolha quando se tem poucos dados de treinamento e se deseja uma classificação rápida.

Suavização de LaPlace

O problema de probabilidade zero é um problema que ocorre quando uma determinada combinação de eventos tem uma probabilidade de ocorrência de exatamente zero em um conjunto de dados de treinamento. Isso pode causar problemas em modelos de aprendizado de máquina, pois eles não podem calcular probabilidades para esses eventos, o que pode levar a previsões imprecisas.

A suavização de Laplace é uma técnica usada para resolver o problema de probabilidade zero, adicionando uma pequena quantidade de "ruído" (ou fração) às contagens de eventos em um conjunto de dados de treinamento. Isso é feito para evitar que qualquer contagem tenha um valor exato de zero e, assim, evitar problemas de probabilidade zero.

A fórmula para a suavização de Laplace é dada por:

$$\theta_i = \frac{x_i + \alpha}{N + \alpha \times d}, \ (i = 1, \ldots, d)$$

onde:

- Theta-i é o novo parâmetro de probabilidade (suavizado)
- xi são as observações desse parâmetro
- alpha é o suavizador (default=1)
- N é o total de ocorrências (tabela de frequências)
- d é o número de palavras/valores diferentes no corpus

Ela é usada em ambas as funções de Bernoulli Naive Bayes e Multinomial Naive Bayes, mas não é necessária na Gaussian Naive Bayes.

MultinomialNB

A função MultinomialNB é uma implementação do algoritmo de classificação Bayesiana Multinomial, que é usado para classificar documentos em diferentes categorias ou classes. Ele é baseado no princípio de que uma palavra ou termo em um documento é independente de outras palavras ou termos no mesmo documento, desde que a classe do documento esteja dada.

O algoritmo funciona da seguinte maneira:

1. Primeiro, é necessário treinar o modelo com um conjunto de dados de treinamento, que contém exemplos de documentos e suas respectivas classes.

2. Para cada classe, é calculada a probabilidade condicional de cada termo dado a classe (P(termo | classe)) usando a fórmula abaixo:

   $$P(termo \mid classe) = \frac{Número\ de\ vezes\ que\ o\ termo\ aparece\ na\ classe + \alpha}{número\ total\ de\ termos\ na\ classe + \alpha \times d}$$

3. Em seguida, é calculada a probabilidade a priori de cada classe (P(classe)) usando a fórmula abaixo:

   $$P(classe) = \frac{número\ de\ documentos\ na\ classe}{número\ total\ de\ documentos}$$

4. Quando um novo documento é dado para classificação, é calculado o valor de probabilidade para cada classe usando a fórmula abaixo:

   $$P(classe \mid documento) = P(classe) \times \Pi\, P(termo \mid classe)$$

5. O documento é classificado na classe com a maior probabilidade.

Exemplo:

Vamos supor que temos um conjunto de dados de treinamento com três classes: "esportes", "política" e "tecnologia". E temos os seguintes documentos de treinamento:

Documento	Classe
"jogadores de futebol jogam em campo"	esportes
"políticos discutem leis em parlamento"	política
"novos smartphones são lançados a cada ano"	tecnologia
"jogadores de basquete jogam em quadra"	esportes

Vamos calcular a probabilidade condicional de cada termo dado a classe "esportes":

Termo	Número de vezes que o termo aparece na classe	P(termo|classe)
jogadores	2	$\frac{2+1}{12+8} = 0.15$
futebol	1	$\frac{1+1}{12+8} = 0.1$
campo	1	$\frac{1+1}{12+8} = 0.1$

basquete	1	$\frac{1+1}{12+8} = 0.1$
quadra	1	$\frac{1+1}{12+8} = 0.1$
jogam	2	$\frac{2+1}{12+8} = 0.15$
de	2	$\frac{2+1}{12+8} = 0.15$
em	2	$\frac{2+1}{12+8} = 0.15$

Vamos calcular a probabilidade condicional de cada termo dado a classe "política":

Termo	Número de vezes que o termo aparece na classe	P(termo|classe)
políticos	1	$\frac{1+1}{5+5} = 0.2$
discutem	1	$\frac{1+1}{5+5} = 0.2$
leis	1	$\frac{1+1}{5+5} = 0.2$
parlamento	1	$\frac{1+1}{5+5} = 0.2$
em	1	$\frac{1+1}{5+5} = 0.2$

E, finalmente, vamos calcular a probabilidade condicional de cada termo dado a classe "tecnologia":

Termo	Número de vezes que o termo aparece na classe	P(termo|classe)
novos	1	$\frac{1+1}{7+7} = 0.14$
smartphones	1	$\frac{1+1}{7+7} = 0.14$
lançados	1	$\frac{1+1}{7+7} = 0.14$
ano	1	$\frac{1+1}{7+7} = 0.14$
são	1	$\frac{1+1}{7+7} = 0.14$

a	1	$\frac{1+1}{7+7} = 0.14$
cada	1	$\frac{1+1}{7+7} = 0.14$

Agora, vamos calcular a probabilidade a priori de cada classe:

$P(esportes) = \frac{2}{4} = 0.5$

$P(política) = \frac{1}{4} = 0.25$

$P(tecnologia) = \frac{1}{4} = 0.25$

Finalmente, vamos supor que temos um novo documento **"jogadores de basquete usam smartphones nas quadras"**, e queremos classificá-lo em uma das classes.

Calcularemos a probabilidade de cada classe dado o documento:
p1 = 0.5 * 0.15 * 0.1 * 1/20 * 1/20 * 1/20 * 0.1 = 0.000009%
p2 = 0.25 * 1/10 * 1/10 * 1/10 * 1/10 * 1/10 * 1/10 * 1/10 = 0.000002%
p3 = 0.25 * 1/14 * 1/14 * 1/14 * 1/14 * 0.14 * 1/14 * 1/14 = 0.000004%

P(esportes | documento) = P(esportes) * P(jogadores | esportes) * P(de | esportes) * P(basquete | esportes) * P(usam | esporte) * P(smartphones | esporte) * P(nas | esporte) * P(quadras | esporte) = 0.5 * 0.15 * 0.1 * 1/20 * 1/20 * 1/20 * 0.1 = 0.000009%

P(política | documento) = P(política) * P(jogadores | política) * P(de | política) * P(basquete | política) * P(usam | política) * P(smartphones | política) * P(nas |política) * P(quadras | política) = 0.25 * 1/10 * 1/10 * 1/10 * 1/10 * 1/10 * 1/10 * 1/10 = 0.000002%

P(tecnologia | documento) = P(tecnologia) * P(jogadores | tecnologia) * P(de | tecnologia) * P(basquete | tecnologia) * P(usam | tecnologia) * P(smartphones | tecnologia) * P(nas | tecnologia) * P(quadras | tecnologia) = 0.25 * 1/14 * 1/14 * 1/14 * 1/14 * 0.14 * 1/14 * 1/14 = 0.000004%

Portanto, o novo documento é classificado como "esportes", pois é a classe com maior probabilidade.

BernoulliNB

Bernoulli Naive Bayes: assume que todos os nossos recursos são binários, de modo que eles levam apenas dois valores. Significa que 0s podem representar “palavra não ocorre no documento” e 1s como “palavra ocorre no documento”.

Aqui está um exemplo prático de como usar o algoritmo Naive Bayes de Bernoulli para classificar animais como mamíferos ou não-mamíferos com base em alguns atributos, como pelos e outros.

1. Aqui está um exemplo prático de como usar o algoritmo Naive Bayes de Bernoulli para classificar animais como mamíferos ou não-mamíferos com base em alguns atributos, como pelos e outros.

Animal	Pelos	Ovos	Penas	Classificação
Cachorro	sim	não	não	Mamífero
Gato	sim	não	não	Mamífero
Pássaro	não	sim	sim	não-mamífero
Tartaruga	não	sim	não	não-mamífero

2. Calcular a probabilidade de cada classe (mamífero / não-mamífero) ocorrer:

$$P(classe) \ = \ \frac{número\ de\ instâncias\ da\ classe}{número\ total\ de\ instâncias\ de\ todas\ as\ classes}$$

P(mamífero) = 2/4 = 0,5
P(não-mamífero) = 2/4 = 0,5

3. Calcular a probabilidade de cada atributo ocorrer para cada classe:

$$P(atributo \mid class) \ = \frac{Número\ de\ vezes\ que\ os\ recursos\ aparecem\ na\ classe + 1}{número\ total\ de\ instâncias\ da\ classe + 2}$$

$$P(pêlos\ =\ sim \mid mamífero)\ = \frac{2+1}{2+2} = \frac{3}{4}$$
$$P(pêlos\ =\ não \mid mamífero)\ = \frac{0+1}{2+2} = \frac{1}{4}$$
$$P(ovos\ =\ sim \mid mamífero)\ = \frac{0+1}{2+2} = \frac{1}{4}$$
$$P(ovos\ =\ não \mid mamífero)\ = \frac{2+1}{2+2} = \frac{3}{4}$$
$$P(penas\ =\ sim \mid mamífero)\ = \frac{0+1}{2+2} = \frac{1}{4}$$
$$P(penas\ =\ não \mid mamífero)\ = \frac{2+1}{2+2} = \frac{3}{4}$$

$$P(pêlos\ =\ sim \mid não\ mamífero)\ = \frac{0+1}{2+2} = \frac{1}{4}$$
$$P(pêlos\ =\ não \mid não\ mamífero)\ = \frac{2+1}{2+2} = \frac{3}{4}$$
$$P(ovos\ =\ sim \mid não\ mamífero)\ = \frac{2+1}{2+2} = \frac{3}{4}$$
$$P(ovos\ =\ não \mid não\ mamífero)\ = \frac{0+1}{2+2} = \frac{1}{4}$$

$P(penas\ =\ sim\ |\ não\ mamífero)\ =\frac{1+1}{2+2}=\frac{1}{2}$

$P(penas\ =\ não\ |\ não\ mamífero)\ =\frac{1+1}{2+2}=\frac{1}{2}$

4. Usar as probabilidades calculadas para classificar novos animais com base em seus atributos:

Exemplo: animal novo com **pelos = sim**, **ovos = sim**, **penas = não (Ornitorrinco)**

P(mamífero | pelos = sim, ovos = sim, penas = não) = P(pelos = sim | mamífero) * P(ovos = sim | mamífero) * P(penas = não | mamífero) * P(mamífero)
P(não-mamífero | pelos = sim, ovos = sim, penas = não) = P(pelos = sim | não-mamífero) * P(ovos = sim | não-mamífero) * P(penas = não | não-mamífero) * P(não-mamífero)

P(mamífero | pelos = sim, ovos = sim, penas = não) = (3/4) * (1/4) * (3/4) * (1/2) = 7%
P(não-mamífero | pelos = sim, ovos = sim, penas = não) = (1/4) * (3/4) * (1/2) * (1/2) = 4%

Como P(mamífero | pelos = sim, ovos = sim, penas = não) é maior do que P(não-mamífero | pelos = sim, ovos = sim, penas = não), podemos classificar esse animal como um mamífero.

Gaussian Naive Bayes

Devido à suposição da distribuição normal, Gaussian Naive Bayes é usado nos casos em que todos os nossos recursos são contínuos. Por exemplo, os recursos do conjunto de dados Iris são largura da sépala, largura da pétala, comprimento da sépala, comprimento da pétala. Portanto, suas características podem ter valores diferentes no conjunto de dados, pois a largura e o comprimento podem variar. Não podemos representar recursos em termos de suas ocorrências. Isso significa que os dados são contínuos. Portanto, usamos Gaussian Naive Bayes aqui.

Em um exemplo, para classificar as alturas de pessoas como "baixas" ou "altas", podemos seguir os seguintes passos:

1. Coletar dados de altura de pessoas classificadas como "baixas" e "altas".
2. Calcular a média e o desvio padrão para cada classe (baixas e altas).
3. Utilizar as seguintes fórmulas para calcular a probabilidade de uma altura dada pertencer a cada classe:

$$P(altura\ |\ baixa)\ =\frac{1}{desvio\_padrao\_baixas\times\sqrt{2\times pi}}\times e^{-\frac{(altura\ -\ media\_baixas)^2}{2\times desvio\_padrao\_baixas^2}}$$

$$P(altura \mid alta) \ = \frac{1}{desvio\_padrao\_altas \times \sqrt{2 \times pi}} \times e^{-\frac{(altura - media\_altas)^2}{2 \times desvio\_padrao\_altas^2}}$$

4. Calcular a probabilidade a priori de uma pessoa ser baixa ou alta.
5. Utilizar a regra de Bayes para calcular a probabilidade a posteriori de uma altura dada pertencer a cada classe:

$$P(baixa \mid altura) \ = \ \frac{P(altura \mid baixa) \times P(baixa)}{P(altura)}$$

$$P(alta \mid altura) \ = \ \frac{P(altura \mid alta) \times P(alta)}{P(altura)}$$

6. Classificar uma altura dada como "baixa" se P(baixa | altura) > P(alta | altura), e "alta" caso contrário.

Dados coletados:

Altura(cm)	Classe
160	Baixa
150	Baixa
170	Baixa
180	Alta
185	Alta
190	Alta

Calculando as médias e desvios padrões:

$$media\_baixas = \frac{160+150+170}{3} = 160$$

$$desvio\_padrao\_baixas = \sqrt{\frac{(160-160)^2+(150-160)^2+(170-160)^2}{3}} = 8.16$$

$$media\_altas = \frac{180+185+190}{3} = 185$$

$$desvio\_padrao\_altas = \sqrt{\frac{(180-185)^2+(185-185)^2+(190-185)^2}{3}} = 4.08$$

Calculando a probabilidade a priori:

$$P(baixa) = \frac{3}{6} = 0.5$$
$$P(alta) = \frac{3}{6} = 0.5$$

Vamos supor que queremos classificar uma altura de 177 cm.

Calculando as probabilidades a posteriori:

$$P(baixa|altura = 177) = \frac{\frac{1}{8.16 \times \sqrt{2\times pi}} \times e^{-\frac{(177-160)^2}{2\times 8.16^2}} \times 0.5}{\frac{1}{8.16 \times \sqrt{2\times pi}} \times e^{-\frac{(177-160)^2}{2\times 8.16^2}} \times 0.5 + \frac{1}{4.08 \times \sqrt{2\times pi}} \times e^{-\frac{(177-185)^2}{2\times 4.08^2}} \times 0.5} = 0.28$$

$$P(alta|altura = 177) = \frac{\frac{1}{4.08 \times \sqrt{2\times pi}} \times e^{-\frac{(177-185)^2}{2\times 4.08^2}} \times 0.5}{\frac{1}{8.16 \times \sqrt{2\times pi}} \times e^{-\frac{(177-160)^2}{2\times 8.16^2}} \times 0.5 + \frac{1}{4.08 \times \sqrt{2\times pi}} \times e^{-\frac{(177-185)^2}{2\times 4.08^2}} \times 0.5} = 0.72$$

Como P(alta | altura = 175) > P(baixa | altura = 175), classificamos a altura de 177 cm como "alta".