

FIAP

NABA

ADELAIDE ALVES DE OLIVEIRA

PROFESSORA



profadelaide.alves@fiap.com.br

Formação Acadêmica

- Bacharel em Estatística – UNICAMP
- Mestre em Ciências – FSP/USP

Atividades Profissionais

- Diretora Técnica Estatística da empresa **SD&W** - www.sdw.com.br
- Professora de Fundamentos Estatísticos, DataMining, Análise Preditiva e Machine Learning na FIAP dos cursos MBA: Big Data, Data Science, Business Intelligence & Analytics, Digital Data Marketing, IA & ML e Engenharia de Dados e nos Shift: People Analytics e Python Journey

NAÏVE BAYES



PROBABILIDADE CONDICIONAL

INFERÊNCIA BAYESIANA



PROBABILIDADES SUBJETIVAS

O fenômeno aleatório pode ser separado em etapas. A informação que ocorreu em uma determinada etapa pode influenciar nas probabilidades de ocorrências das etapas sucessivas.

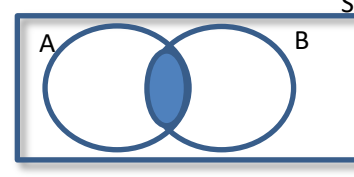
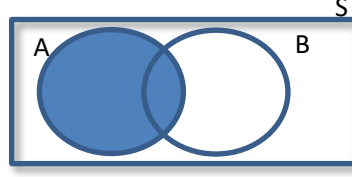
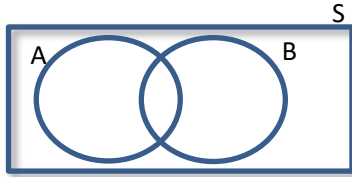
Definição:

Dados dois eventos A e B, a probabilidade condicional de A dado que ocorreu B é representado por $P(A/B)$ e dada por:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

PROBABILIDADE CONDICIONAL

Dados dois eventos A e B, com $P(A) \neq 0$:



$$\Rightarrow P(B | A) = \frac{P(A \cap B)}{P(A)}.$$

para o cálculo de $P(A \cap B)$:

$$P(A \cap B) = P(A) \cdot P(B | A)$$

$$P(A \cap B) = P(B) \cdot P(A | B)$$

Teoria

PROBABILIDADE CONDICIONAL

A técnica de Basket Analysis utiliza a probabilidade condicional para encontrar cestas de produtos. Algoritmo Apriori .

Um Exemplo de Sucesso!



- ✓ Descobriu-se que homens entre trinta e quarenta e cinco anos, que compram cervejas, nas sextas-feiras, após as dezesseis horas, também compram fraldas!
- ✓ Resultado: apenas mudando os produtos de lugar, colocando as fraldas ao lado de cervejas nos pontos de venda, obteve-se um aumento de mais de quarenta por cento nas vendas de fraldas.
- ✓ O que acha de possuir uma informação como essa?

A Wall-Mart soube tirar bom proveito dela!

- Support (frequência)

$$(A \cap B \Rightarrow C) = \%$$

- Confidence (probabilidade condicional)

$$(A \cap B \Rightarrow C) = \frac{P(A \cap B \cap C)}{P(A \cap B)}$$

- Lift(associação)

$$(A \& B \Rightarrow C) = \frac{P(A \cap B \cap C)}{P(A \cap B)P(C)}$$

TEOREMA DE BAYES

O Teorema de Bayes trata de problemas em que se deseja determinar a probabilidade de um evento ocorrer dada uma condição: (Probabilidade de ocorrer A, na condição de que B já tenha ocorrido – Probabilidade Condicional).

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

Onde,

- $P(A)$ e $P(B)$ são as probabilidades a priori de A e B.
- $P(B|A)$ e $P(A|B)$ são as probabilidades posteriores de B condicional a A e de A condicional a B, respectivamente.

PROBABILIDADE CONDICIONAL

Exemplo :

- Uma determinada peça é manufaturada por 3 fábricas: A, B e C. Sabe-se que A produz o dobro de peças que B e que B e C produzem o mesmo número de peças. Sabe-se ainda que 2% das peças produzidas por A e por B são defeituosas, enquanto que 4% das produzidas por C são defeituosas. Todas as peças produzidas são misturadas e colocadas em um depósito. **Se do depósito for retirada uma peça ao acaso, qual a probabilidade de que ela seja defeituosa?**

$D = \{ \text{A peça é defeituosa} \},$

$A = \{ \text{A peça provém da fábrica A} \},$

$B = \{ \text{A peça é a da fábrica B} \}$ e

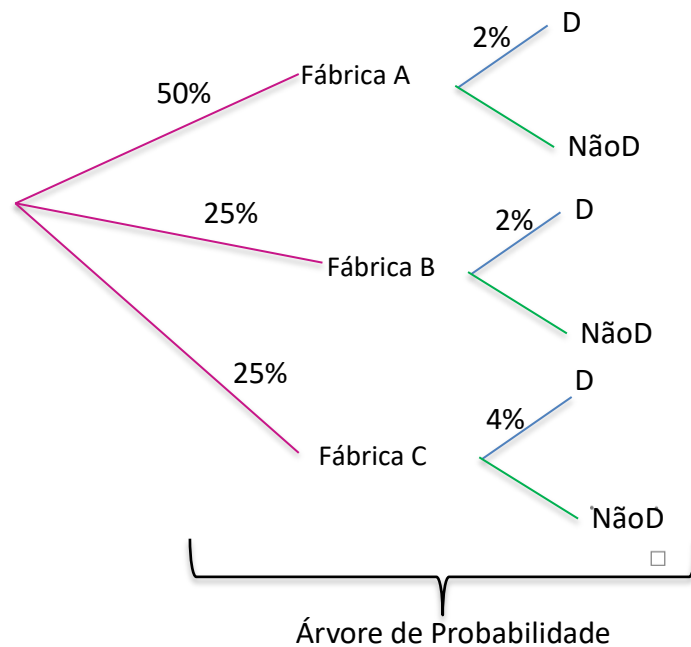
$C = \{ \text{A peça é da fábrica C} \}.$

→ $P(A) = 50\%, P(B) = P(C) = 25\%.$

→ $P(D|A) = P(D|B) = 2\%$ e que $P(D|C) = 4\%.$

Pelo teorema da probabilidade total:

→ $P(D) = P(A).P(D/A) + P(B).P(D/B) + P(C).P(D/C) =$
 $0,5*0,02 + 0,25*0,02 + 0,25*0,04 = 2,50\%,$

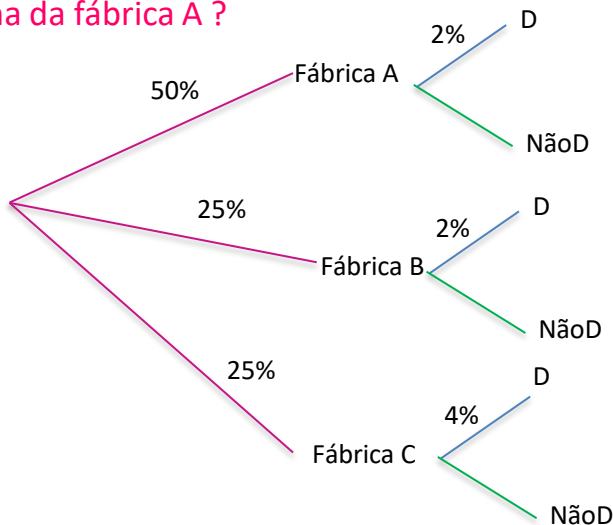


PROBABILIDADE CONDICIONAL

Exemplo :

- Uma determinada peça é manufaturada por 3 fábricas: A, B e C. Sabe-se que A produz o dobro de peças que B e que B e C produzem o mesmo número de peças. Sabe-se ainda que 2% das peças produzidas por A e por B são defeituosas, enquanto que 4% das produzidas por C são defeituosas. Todas as peças produzidas são misturadas e colocadas em um depósito. Se do depósito for retirada uma peça ao acaso, qual a probabilidade de que ela seja defeituosa?

Se uma peça escolhida ao acaso está fora das especificações(defeituosa), qual é a probabilidade que venha da fábrica A ?



$$P(A / D) = ?$$

$$P(A / D) = \frac{P(A) \cdot P(D/A)}{P(A) \cdot P(D/A) + P(B) \cdot P(D/B) + P(C) \cdot P(D/C)}$$

$$P(A / D) = \frac{0,50 \cdot 0,02}{(0,50 \cdot 0,02 + 0,25 \cdot 0,02 + 0,25 \cdot 0,04)}$$

$$P(A / D) = 0,40$$

PROBABILIDADE CONDICIONAL

INFERÊNCIA BAYESIANA



PROBABILIDADES SUBJETIVAS

Exemplo :

Considere o evento A = chover em SP no dia 12 de janeiro do próximo ano.

Suponha que uma pessoa morando em Fortaleza tenha que calcular essa probabilidade. Se ela não tiver informação sobre o tempo em São Paulo, poderá atribuir a probabilidade de $\frac{1}{2}$.

Já o morador de São Paulo tem informações adicionais, como por exemplo, ele saberá que janeiro, fevereiro e março são os meses mais chuvosos e poderá arriscar uma probabilidade de $\frac{2}{3}$ de ocorrer o evento A.

PROBABILIDADE CONDICIONAL

Exercício :

O São Paulo Futebol Clube ganha com probabilidade 0,7 se chove e com 0,8 se não chove. Em Setembro a probabilidade de chuva é de 0,3. O São Paulo ganhou uma partida em Setembro, qual a probabilidade de ter chovido nesse dia?

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

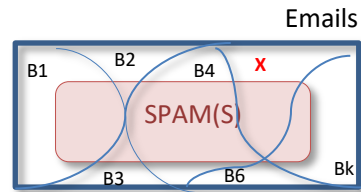
Eventos:
C: Chover
G: São Paulo ganhar um jogo

$$P(C/G) = \frac{P(C) * P(G/C)}{P(C) * P(G/C) + P(NC) * P(G/NC)}$$

$$P(C/G) = \frac{0,30 * 0,70}{0,30 * 0,70 + 0,70 * 0,80} = \frac{0,21}{0,21 + 0,56} = 0,273$$

USANDO TEOREMA DE BAYES

A técnica Naïve Bayes é um dos algoritmos de aprendizado supervisionado mais simples. Tem como premissa a suposição de independência entre as variáveis do problema (teorema de Bayes). Isto é, o classificador Naïve Bayes assume que o efeito de uma determinada variável é independente de outras variáveis. Mesmo que essas características dependam umas das outras, todas contribuem independentemente para a probabilidade e, é por isso que ele é conhecido como “naïve” (ingênuo).



USANDO TEOREMA DE BAYES

detecção de spam:

Dado uma mensagem que contém uma determinada palavra “X” queremos computar a probabilidade dela ser um spam. A maioria das pessoas acostumadas a receber e-mails sabe que essa mensagem provavelmente é spam, no entanto, como podemos “indicar” ao “gerenciador” para uma tomada de decisão?

→ Sabemos que a palavra “X” está presente no e-mail.
Qual a chance deste e-mail ser um SPAM (S)? $P(S / X)$

$P(S)$ = Probabilidade de ter um SPAM → 10%

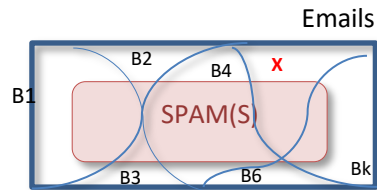
$P(NS)$ = Probabilidade de Não SPAM → 90%

$P(X/S)$ = Probabilidade de ter a palavra X aparecer em mensagens de SPAM's → 30%

$P(X/NS)$ = Probabilidade de ter a palavra X aparecer em mensagens sem SPAM's → 2%

→ $P(S / X) = P(X / S) * P(S) / (P(X / S) * P(S) + P(X/NS) * P(NS))$

→ $P(S / X) = 30\% * 10\% / (30\% * 10\% + 2\% * 90\%) = 62,5\%$



CLASSIFICADOR NAÏVE BAYES

Vamos acompanhar os cálculos seguindo um exemplo simples: a avaliação de concessão de um empréstimo (concessão: SIM ou NÃO). A tomada de decisão será em função das características do solicitante: seu salário, idade, local de moradia, histórico de empréstimos e transações anteriores. Mesmo que essas informações sejam interdependentes, eles ainda são considerados de forma independentes.



CLASSIFICADOR NAÏVE BAYES

O que precisaremos é classificar se o solicitante de crédito vai ter ou não o crédito liberado, com base em suas características. Usaremos para esse exemplo como variáveis explicativas: estado civil, idade, se tem casa própria, situação de emprego. Variável Target: Crédito (SIM/NÃO).

estadocivil	fx_etaria	casapropria	empregado	crédito
SOLTEIRO	26 a 34 anos	NAO	SIM	SIM
CASADO	26 a 34 anos	SIM	NAO	SIM
DEMAIS	26 a 34 anos	SIM	SIM	NAO
SOLTEIRO	26 a 34 anos	SIM	SIM	SIM
CASADO	35 e mais anos	NAO	NAO	SIM
...
DEMAIS	35 e mais anos	SIM	NAO	SIM

Quadro 1 – Exemplo da base de dados



CLASSIFICADOR NAÏVE BAYES

Queremos calcular qual a probabilidade de ser liberado crédito para um solicitante dado que suas características são: seu estado civil é casado, sua idade é 30 anos, não tem cada própria e está empregado.

O classificador Naïve Bayes calcula a probabilidade de um evento nos seguintes passos:

- Passo 1: Calcular a probabilidade das classes (Probabilidade a priori):

$$P(\text{crédito}=\text{SIM}) = 11/20=0,55$$

$$P(\text{crédito}=\text{NAO}) = 9/20=0,45$$



CLASSIFICADOR NAÏVE BAYES

Queremos calcular qual a probabilidade de ser liberado crédito para um solicitante dado que suas características são: seu estado civil é casado, sua idade é 30 anos, não tem cada própria e está empregado.

O classificador Naïve Bayes calcula a probabilidade de um evento nos seguintes passos:

- Passo 1: Calcular a probabilidade das classes (Probabilidade a priori):

$$P(\text{crédito}=\text{SIM}) = 11/20=0,55$$

$$P(\text{crédito}=\text{NAO}) = 9/20=0,45$$



CLASSIFICADOR NAÏVE BAYES

- Passo 2: Criar a tabela de “Probabilidade” (likelihood) encontrando as probabilidades para cada atributo:

	Crédito		
estado civil	NAO	SIM	Cálculo de Probabilidade
casado	2	5	$7/20 = 0,35$
demais	3	2	$5/20 = 0,25$
solteiro	4	4	$8/20 = 0,40$

Quadro 2 – Cálculo de Probabilidade Estado Civil

	Crédito		
Faixa Etária	NAO	SIM	Cálculo de Probabilidade
18 a 25 anos	3	2	$5/20 = 0,25$
26 a 34 anos	4	4	$8/20 = 0,40$
35 e mais anos	2	5	$7/20 = 0,35$

Quadro 3 – Cálculo de Probabilidade Faixa Etária

CLASSIFICADOR NAÏVE BAYES

	Crédito		
Casa Própria	NAO	SIM	Cálculo de Probabilidade
NAO	8	5	$13/20 = 0,65$
SIM	1	6	$7/20 = 0,35$

Quadro 4 – Cálculo de Probabilidade Ter Casa Própria

	Crédito		
Empregado	NAO	SIM	Cálculo de Probabilidade
NAO	4	2	$6/20 = 0,30$
SIM	5	9	$14/20 = 0,70$

Quadro 5 – Cálculo de Probabilidade estar empregado



CLASSIFICADOR NAÏVE BAYES

- Passo 3: Agora, usamos a equação de Bayes para calcular a probabilidade posterior de cada classe. A classe com a maior probabilidade posterior é o resultado da previsão.

Calculando para o indivíduo com estas características: seu estado civil é casado, sua idade é 30 anos, não tem casa própria e está empregado.

→ $X = \{\text{estado civil} = \text{"casado"}; Fx_etaria = \text{"26 a 34 anos"}; \text{não tem casa própria e empregado} = \text{sim}\}$

Calculando na situação de Concessão de Crédito=SIM



CLASSIFICADOR NAÏVE BAYES

- Passo 3: Agora, usamos a equação de Bayes para calcular a probabilidade posterior de cada classe. A classe com a maior probabilidade posterior é o resultado da previsão.

Calculando para o indivíduo com estas características:

seu estado civil é casado, sua idade é 30 anos, não tem cada própria e está empregado.

- $P(\text{estado civil}=\text{"casado"} / \text{credito}=\text{SIM}) = 5/11$
- $P(\text{Fx_etaria}=\text{"26 a 34 anos"} / \text{credito}=\text{SIM}) = 4/11$
- $P(\text{ter cada própria}=\text{"Não"} / \text{credito}=\text{SIM}) = 5/11$
- $P(\text{empregado}=\text{"SIM"} / \text{credito}=\text{SIM}) = 9/11$

$P(\text{Sim} | \text{Casado}, 26\text{a}34\text{anos}, \text{NãoCasa}, \text{Empregado}) = ?$

⇒ $P(\text{Sim}/X) = P(\text{Casado}/\text{SIM})P(26\text{a}34\text{anos}/\text{SIM})P(\text{NãoCasa}/\text{SIM})P(\text{empregado}/\text{SIM}) = 0,0615$



CLASSIFICADOR NAÏVE BAYES

- Passo 3: Agora, usamos a equação de Bayes para calcular a probabilidade posterior de cada classe. A classe com a maior probabilidade posterior é o resultado da previsão.

Agora calculando na situação de Concessão de Crédito=NÃO

- $P(\text{estado civil}=\text{"casado"} / \text{credito}=\text{NAO}) = 2/9$
- $P(\text{Fx_etaria}=\text{"26 a 34 anos"} / \text{credito}=\text{NAO}) = 4/9$
- $P(\text{ter cada própria}=\text{"Não"} / \text{credito}=\text{NAO}) = 8/9$
- $P(\text{empregado}=\text{"SIM"} / \text{credito}=\text{NAO}) = 5/9$

$$\Rightarrow P(\text{Não} \mid \text{Casado}, 26\text{a}34\text{anos}, \text{NãoCasa}, \text{Empregado} = ?)$$

$$\Rightarrow P(\text{Não} \mid \text{Casado}, 26\text{a}34\text{anos}, \text{NãoCasa}, \text{Empregado}) =$$

$$\Rightarrow P(\text{Casado}/\text{NAO}) * P(26\text{a}34\text{anos}/\text{NAO}) * P(\text{NãoCasa}/\text{NAO}) * P(\text{empregado}/\text{NAO}) = 0,04877$$



CLASSIFICADOR NAÏVE BAYES

- E,

Como $P(\text{Sim}/X) > P(\text{Não}/X)$, classificamos esse solicitante com essas características como Crédito = SIM.



CLASSIFICADOR NAÏVE BAYES

Aplicações do Algoritmo Naïve Bayes:

- Em aplicações na análise de crédito, diagnósticos médicos, busca por falhas em sistemas mecânicos, detecção de SPAMs entre outras aplicações.
- Aplicações na saúde, como sistemas que determinam se alguém tem uma doença ou não.
- Como classificador de SPAM ele analisa e-mails e tenta avaliar se ele é spam ou não (classes definidas) com base em suas informações e estrutura.
- Classificadores de sentimento. Nesses casos, eles analisam os textos e tentam identificar a emoção expressada, geralmente entre opções específicas como “neutro”, “positivo” ou “negativo”. Nas redes sociais é utilizado para identificar se o usuário está feliz ou triste ao publicar determinado texto.
- Em classificação de textos, têm boas taxas de sucesso em comparação com outros algoritmos.



CLASSIFICADOR NAÏVE BAYES

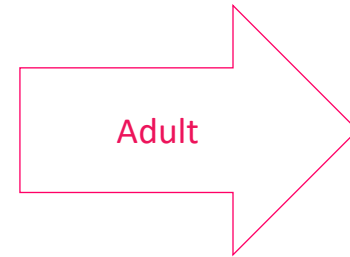
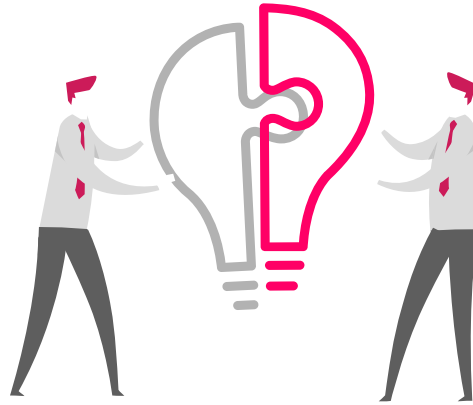
Mais aplicações do Algoritmo Naïve Bayes:

- Sistema de Recomendação: utilizando filtragem colaborativa constroem um sistema de recomendação. Nesse caso, o objetivo é analisar certas pessoas e tentar sugerir algo que possa interessar a elas, seja conteúdo ou produtos.
- Usado para fazer previsões em tempo real: Por possuir uma velocidade relativamente alta e precisar apenas de poucos dados para realizar a classificação, o Naïve Bayes pode ser utilizado para previsões em tempo real.
- Pode ser utilizado para prever a probabilidade de múltiplas classes de variáveis.



EXERCITANDO

NAIVE BAYES



BIBLIOGRAFIA

- KUHN, M. / JOHNSON K. **Applied Predictive Modeling**, 1st ed. 2013, Corr. 2nd printing 2018 Edition
- LESKOVEC, RAJAMARAM, ULLMAN. **Mining of Massive Datasets**, 2014. <http://mmds.org>.
- HAIR, J.F. / ANDERSON, R.E. / TATHAN, R.L. / BLACK, W.C. **Análise multivariada de dados**, 2009
- TORGO, L. **Data Mining with R: Learning with Case Studies**, 2.a ed. Chapman and Hall/CRC , 2007
- MINGOTI, S.A.; **Análise de dados através de métodos de estatística multivariada**, UFMG, 2005
- CARVALHO, L.A.V., **Datamining – A mineração de dados no marketing, medicina, economia, engenharia e administração**. Rio de Janeiro: Editora Ciência Moderna, 2005.
- BERRY, M.J.A., LINOFF, G. **Data Mining Techniques For Marketing, Sales and Customer Support**. 3a. ed. New York: John Wiley & Sons, Inc., 2011.
- DUNHAM, M.H. **Data Mining - Introductory and Advanced Topics**. Prentice Hall, 2002.
- DINIZ, C.A.R. , NETO F.L. **Data Mining: Uma Introdução**. São Paulo: XIV Simpósio Nacional de Probabilidade e Estatística. IME-USP, 2000.

OBRIGADO



/AdelaideAlves



profadelaide.alves@fiap.com.br

FIAP

Copyright © 2022 | Professor (a) Adelaide Alves de Oliveira

Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente proibido sem consentimento formal, por escrito, do professor/autor.