

Naive Bayes e Redes Bayesianas

Prof. Bruno Brandoli

bruno.brandoli@ufms.br

Campus de Ponta Porã (CPPP)
Universidade Federal de Mato Grosso do Sul (UFMS)

2º Semestre/2017



Sumário

- 1 Introdução
- 2 Teoria da Probabilidade
- 3 Mapa de Bayes
- 4 Decisão Bayesiana para Classificação de Dados
- 5 Exemplos
- 6 Redes Bayesianas



Sumário

- 1 Introdução
- 2 Teoria da Probabilidade
- 3 Mapa de Bayes
- 4 Decisão Bayesiana para Classificação de Dados
- 5 Exemplos
- 6 Redes Bayesianas



- Paradigma probabilístico
- Naive Bayes é um algoritmo que usa o Teorema de Bayes e toma como hipótese que os atributos são independentes entre si
- Revisão sobre teoria da probabilidade
 - Probabilidade Conjunta
 - Eventos Independentes
 - Probabilidade Condicional
- Redes Bayesianas



Sumário

- 1 Introdução
- 2 Teoria da Probabilidade**
- 3 Mapa de Bayes
- 4 Decisão Bayesiana para Classificação de Dados
- 5 Exemplos
- 6 Redes Bayesianas



Probabilidade conjunta é a probabilidade de dois eventos A e B ocorrerem, apresentando os seguintes casos:

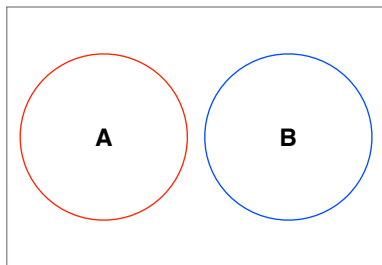
- $P(A)$: Probabilidade do evento A ocorrer
- $P(B)$: Probabilidade do evento B ocorrer
- $P(A \cup B) = P(A) * P(B)$ se A e B forem **eventos independentes**

Exemplo: Qual a probabilidade de sair o número seis em um jogo de dados, sabendo que em outro dado jogado foi cinco?

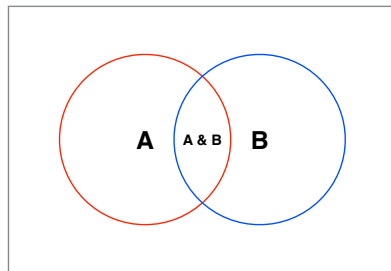
- $P(A \cap B)$ ou $P(A, B)$: **Probabilidade condicional** de A e B ocorrerem **Exemplo:** Qual a probabilidade de um paciente ter dengue, sabendo-se que ele têm sintomas de febre alta e dor nos olhos?



Teoria de Conjuntos



(a) Eventos Independentes



(b) Independência Condicional

Caso A e B não forem eventos independentes, temos que:

$$\begin{aligned} P(A \cap B) &= P(A) * P(B|A) \\ P(B|A) &= \frac{P(A \cap B)}{P(A)} \end{aligned} \quad (1)$$

é a probabilidade condicional de que B ocorra dado que A ocorreu.

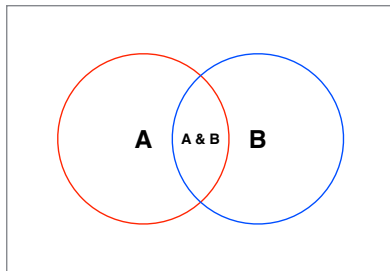


Teorema de Bayes

Sabe-se que $P(A \cap B) = P(B \cap A)$. Assim:

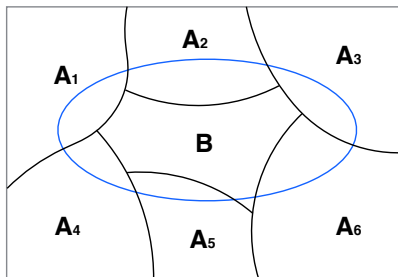
$$P(B|A) * P(A) = P(A|B) * P(B) \quad (2)$$

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)} \quad (3)$$



Teorema de Bayes

O raciocínio é análogo para mais de duas variáveis aleatórias A_1, A_2, \dots, A_n , e B .



Teorema de Bayes

O raciocínio é análogo para mais de duas variáveis aleatórias A_1, A_2, \dots, A_n , e B . Dado que:

$$P(B|A) * P(A) = P(A|B) * P(B) \quad (4)$$

$$P(B|A_1, A_2, \dots, A_n) = \frac{P(A_1, A_2, \dots, A_n|B) * P(B)}{P(A_1, A_2, \dots, A_n)} \quad (5)$$

$$P(B|A_1, \dots, A_n) = \frac{[P(A_1|B) * P(A_2|B) * \dots * P(A_n|B)] * P(B)}{P(A_1) * P(A_2) * \dots * P(A_n)} \quad (6)$$

$$P(B|A_1, \dots, A_n) = \frac{\prod_{i=1}^n P(A_i|B) * P(B)}{\prod_{i=1}^n P(A_i)} \quad (7)$$

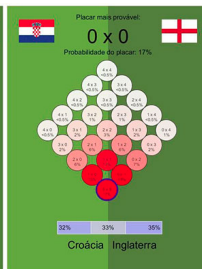
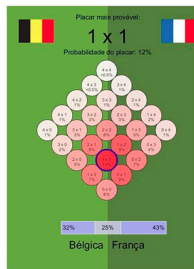
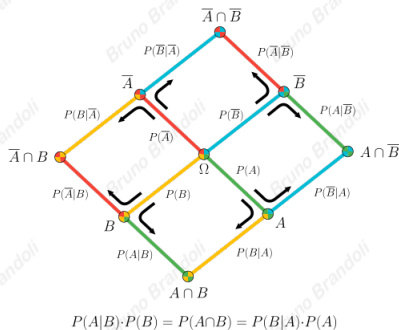


Sumário

- 1 Introdução
- 2 Teoria da Probabilidade
- 3 Mapa de Bayes**
- 4 Decisão Bayesiana para Classificação de Dados
- 5 Exemplos
- 6 Redes Bayesianas



Mapa de Bayes



Sumário

- 1 Introdução
- 2 Teoria da Probabilidade
- 3 Mapa de Bayes
- 4 Decisão Bayesiana para Classificação de Dados**
- 5 Exemplos
- 6 Redes Bayesianas



Decisão Bayesiana

- O padrão a ser classificado é representado pelo vetor:
 $\mathbf{x} = [x_1, x_2, x_3, \dots, x_n]^T$
- Supõe-se que existam duas classes ω_1 ω_2 nas quais um dado padrão pode pertencer.
- As probabilidades à priori de ambas são: $P(\omega_1)$ e $P(\omega_2)$
- Assume-se que as a probabilidades condicionais de cada classe $P(\mathbf{x}|\omega_i), i = 1, 2$
- Como $P(\mathbf{x}|\omega_i), i = 1, 2$, são desconhecidas, nesse caso elas podem ser estimadas.
- Assim, sabendo-se da probabilidade $P(\mathbf{x}|\omega_i)$ e $P(\omega_i), i = 1, 2$, pode-se determinar a probabilidade condicional do padrão descrito por \mathbf{x} pertencer à classe ω_i .



Decisão Bayesiana

- Pela lei da probabilidade total:

$$P(\mathbf{x}) = \sum_{i=1}^2 P(\mathbf{x}|\omega_i)P(\omega_i) \quad (8)$$

- Logo, é possível calcular para cada classe:

$$P(\omega_i|\mathbf{x}) = \frac{P(\mathbf{x}|\omega_i) * P(\omega_i)}{P(\mathbf{x})} \quad (9)$$

$\text{Posteriori} = \frac{\text{Verossimilhança} * \text{Priori}}{\text{Marginal}}$
--



A classificação Bayesiana será decidida pela regra:

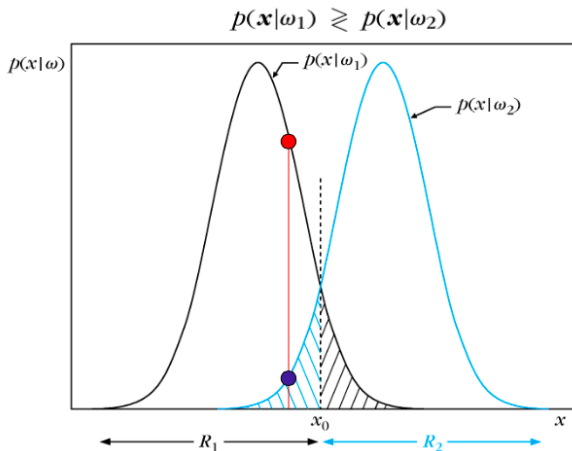
$$Regra = \begin{cases} \chi \in \omega_1, & \text{se } P(\omega_1|\chi) > P(\omega_2|\chi) \\ \chi \in \omega_2, & \text{se caso contrário} \end{cases}$$

Logo:

$$P(\chi|\omega_1) * P(\omega_1) \geq P(\chi|\omega_2) * P(\omega_2)$$



Decisão Bayesiana



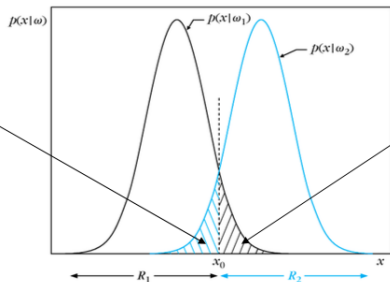
$$R_1(\rightarrow \omega_1) \text{ e } R_2(\rightarrow \omega_2)$$

Decisão Bayesiana

$$P_e = \frac{1}{2} \int_{-\infty}^{x_0} p(x|\omega_2) dx + \frac{1}{2} \int_{x_0}^{+\infty} p(x|\omega_1) dx$$

$$\frac{1}{2} \int_{-\infty}^{x_0} p(x|\omega_2) dx$$

$$\frac{1}{2} \int_{x_0}^{+\infty} p(x|\omega_1) dx$$



$l = 1d$



Sumário

- 1 Introdução
- 2 Teoria da Probabilidade
- 3 Mapa de Bayes
- 4 Decisão Bayesiana para Classificação de Dados
- 5 Exemplos**
- 6 Redes Bayesianas



Exemplo 1

Outlook	Temperature	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Exemplo 1

Outlook (A ₁)			Temperature (A ₂)			Humidity (A ₃)			Windy (A ₄)			Play (B)	
Yes	No		Yes	No		Yes	No		Yes	No		Yes	No
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	???

$$P(\text{Yes}|\text{Sunny, Cool, High, True}) = (2/9 * 3/9 * 3/9 * 3/9 * 9/14) / P(\text{Sunny, Cool, High, True})$$

$$P(\text{No}|\text{Sunny, Cool, High, True}) = (3/5 * 1/5 * 4/5 * 3/5 * 5/14) / P(\text{Sunny, Cool, High, True})$$

$$P(\text{Yes}|\text{Sunny, Cool, High, True}) = 0,0053$$

$$P(\text{No}|\text{Sunny, Cool, High, True}) = 0,0206$$

$$P(\text{Yes}|\text{Sunny, Cool, High, True}) < P(\text{No}|\text{Sunny, Cool, High, True})$$



Exemplo 1

Outlook (A_1)			Temperature (A_2)			Humidity (A_3)			Windy (A_4)			Play (B)	
Yes	No		Yes	No		Yes	No		Yes	No		Yes	No
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

Outlook	Temp.	Humidity	Windy	Play
Sunny	Cool	High	True	???

A probabilidade estimada é:

$$P(\text{Yes}|\text{Sunny}, \text{Cool}, \text{High}, \text{True}) = 0,0053 / (0,0053 + 0,0206) = 20\%$$

$$P(\text{No}|\text{Sunny}, \text{Cool}, \text{High}, \text{True}) = 0,0206 / (0,0053 + 0,0206) = 79.45\%$$



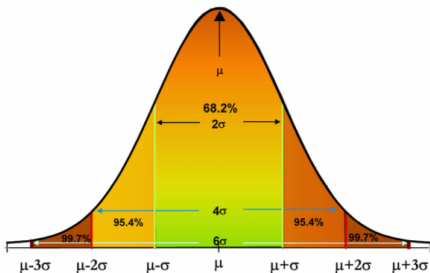
Exemplo 2

Outlook	Temperature	Humidity	Windy	Play
Sunny	85	85	False	No
Sunny	80	90	True	No
Overcast	83	86	False	Yes
Rainy	70	96	False	Yes
Rainy	68	80	False	Yes
Rainy	65	70	True	No
Overcast	64	65	True	Yes
Sunny	72	95	False	No
Sunny	69	70	False	Yes
Rainy	75	80	False	Yes
Sunny	75	70	True	Yes
Overcast	72	90	True	Yes
Overcast	81	75	False	Yes
Rainy	71	91	True	No

Exemplo 2

Para atributos numéricos, assume-se que os valores seguem uma distribuição.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i, \quad \sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2}, \quad p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



Exemplo 2

Outlook (A ₁)			Temperature (A ₂)			Humidity (A ₃)			Windy (A ₄)			Play (B)	
Yes	No		Yes	No		Yes	No		Yes	No		Yes	No
Sunny	2	3	Calcular média e desvio padrão para Yes e No			Calcular média e desvio padrão para Yes e No			False	6	2	9	5
Overcast	4	0							True	3	3		
Rainy	3	2											
Sunny	2/9	3/5	média	73	74,6	média			False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	desvio	6,2	7,9	desvio			True	3/9	3/5		
Rainy	3/9	2/5											

Qual a probabilidade de chover se a temperatura for 66 °F?

$$P(\text{Yes}|66^\circ F)? \text{ ou } P(\text{No}|66^\circ F)?$$

$$P(\text{Yes}|66^\circ F) = \frac{1}{6,2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{66-73}{6,2}\right)^2} = 0,0340$$

$$P(\text{No}|66^\circ F) = \frac{1}{7,9\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{66-74,6}{7,9}\right)^2} = 0,0130$$

Logo vai chover, pois $P(\text{Yes}|66^\circ F) > P(\text{No}|66^\circ F)$.



Exemplo 3

Para entregar na próxima aula, feito à mão. Calcular, para o caso $P(\text{Yes}, \text{No}, \text{Yes}, \text{No}, ???)$, (1) a probabilidade de ser um mamífero (*mammal*) ou não-mamífero (*not-mammal*), e a (2) probabilidade estimada deles.

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	reptiles
salmon	no	no	yes	no	fishes
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	amphibians
komodo	no	no	no	yes	reptiles
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	birds
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	fishes
turtle	no	no	sometimes	yes	reptiles
penguin	no	no	sometimes	yes	birds
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	fishes
salamander	no	no	sometimes	yes	amphibians
gila monster	no	no	no	yes	reptiles
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	birds
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	birds

Exemplo 4

Para entregar na próxima aula, feito à mão também. Calcular, para o caso $P(\text{No}, \text{Married}, \$120K, ???)$, (1) a probabilidade sonegar imposto (evade) ou não, e a (2) probabilidade estimada deles.

<i>Tid</i>	<i>Refund</i>	<i>Marital Status</i>	<i>Taxable Income</i>	<i>Evade</i>
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



Sumário

- 1 Introdução
- 2 Teoria da Probabilidade
- 3 Mapa de Bayes
- 4 Decisão Bayesiana para Classificação de Dados
- 5 Exemplos
- 6 Redes Bayesianas**



Redes Bayesianas (BBN)

- Modelo probabilístico baseado em grafo
 - Nós são variáveis não determinísticas
 - Arestas representam a interdependência entre as variáveis
- A construção do modelo não é trivial, porém é uma solução para o Naive Bayes



Redes Bayesianas (BBN)

Passo 1: Ordenar as variáveis.

Passo 2: A topologia da rede é criada seguindo os casos:

- Se um atributo não tem nó-pai $P(A)$, assume-se apenas a probabilidade a priori
- Se um atributo tem um pai $P(B, A)$, assume-se a probabilidade condicional
- Se um atributo tem vários pais $P(B, A_1, \dots, A_n)$, assume-se a probabilidade condicional

Passo 3: Calcular as probabilidades por meio das frequências.



Redes Bayesianas (BBN)

Exemplo:

Tarefa: Leia o capítulo 14 do livro e estude o exemplo do livro texto da disciplina.



Fim...

Alguma pergunta?

