Aula 04

Métodos probabilísticos de classificação. Naïve Bayes.

Aula passada

Árvores de decisão: divisão e conquista

id	Estudante	Idade	ÍndiceCredito	CompraPC
x_1	Sim	Jovem	Regular	Sim
x_2	Sim	Meia-idade	Excelente	Sim
x_3	Não	Jovem	Excelente	Não
x_4	Não	Sênior	Regular	Não
x_5	Sim	Sênior	Excelente	Sim

Idade?

Jovem

id	Estudante	Idade	ÍndiceCredito	CompraPC
x_1	Sim	Jovem	Regular	Sim
x_3	Não	Jovem	Excelente	Não

Meia-Idade

id	Estudante	Idade	ÍndiceCredito	CompraPC
x_2	Sim	Meia-idade	Excelente	Sim

Sênior

id	Estudante	Idade	ÍndiceCredito	CompraPC
x_4	Não	Sênior	Regular	Não
x_5	Sim	Sênior	Excelente	Sim

Ideia: recursivamente tentar melhorar a pureza de uma partição

Aula passada

Algoritmos ID3, C4.5 e CART

	Critério de seleção	Tipos de atributos suportados	Lida com valores desconhecidos?	Estratégia de poda?
ID3	Ganho de Informação	Apenas categóricos	Não	Não adota
C4.5	Razão de Ganho	Categóricos e numéricos	Sim	Pessimista
CART	Índice Gini	Categóricos e numéricos	Sim	Custo-complexidade

Ganho de Informação

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i),$$

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{\|D|} \times Info(D_j).$$

$$Gain(A) = Info(D) - Info_A(D).$$

Razão de Ganho

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo_A(D)}.$$

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right).$$

Índice Gini

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2$$

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2).$$

Métodos probabilísticos de classificação

Aprendizado Bayesiano

Introdução ao Algoritmo Naïve Bayes

Métodos probabilísticos de classificação

- Objetivo da tarefa de classificação: aprender a função* $y = f(\mathbf{x})$
  - $\mathbf{x}$ é a entrada, um vetor com valores dos atributos preditivos
  - y é a saída, um valor categórico em $\{c_1, c_2, \ldots, c_m\}$

Aprender um modelo com base nos dados de treinamento que determine **a melhor classe** para uma nova instância dado o seu vetor de atributos

- Métodos probabilísticos:
  - Especialmente úteis quando há incerteza sobre o domínio

Aprender um modelo com base nos dados de treinamento que determine **a classe mais provável** para uma nova instância dado o seu vetor de atributos

* Assumiremos função = hipótese = modelo

Métodos probabilísticos de classificação

- Objetivo da tarefa de classificação: aprender a função* $y = f(\mathbf{x})$
  - $\mathbf{x}$ é a entrada, um vetor com valores dos atributos preditivos
  - y é a saída, um valor categórico em $\{c_1, c_2, \ldots, c_m\}$

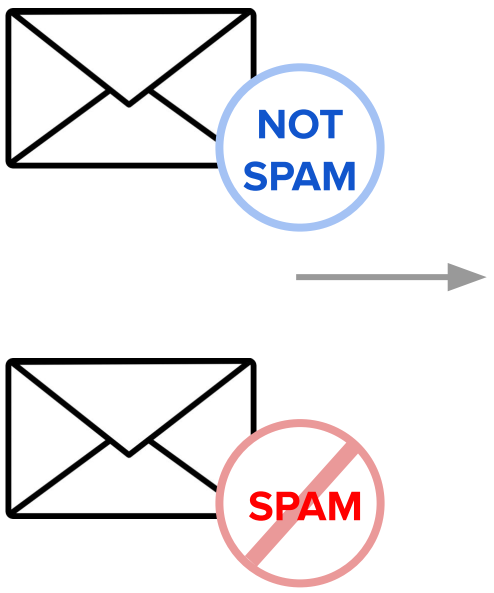
Encontrar **a melhor classe** y a partir de $\mathbf{x}$

Encontrar a **classe mais provável** y a partir de $\mathbf{x}$

* Assumiremos função = hipótese = modelo

"Classe mais provável"

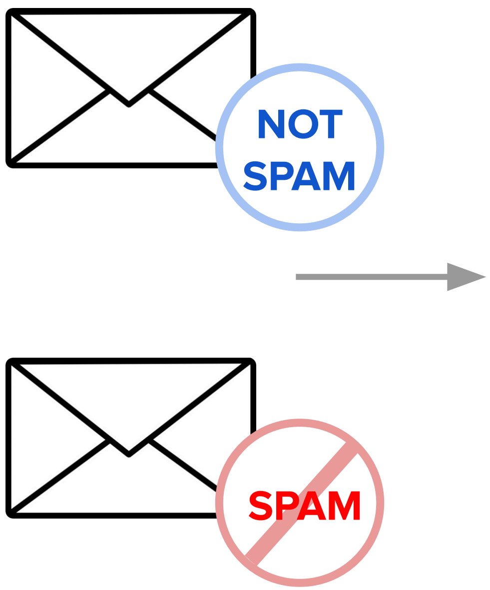
- Intuição básica



e-mail	Classe
1	SPAM
2	SPAM
3	NOT SPAM
4	SPAM
5	SPAM
6	SPAM
7	SPAM
8	NOT SPAM
9	NOT SPAM
10	SPAM

"Classe mais provável"

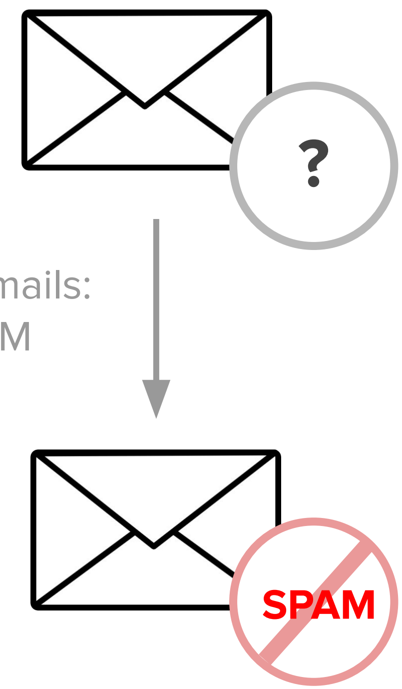
- Intuição básica



e-mail	Classe
1	SPAM
2	SPAM
3	NOT SPAM
4	SPAM
5	SPAM
6	SPAM
7	SPAM
8	NOT SPAM
9	NOT SPAM
10	SPAM

7/10 e-mails: SPAM

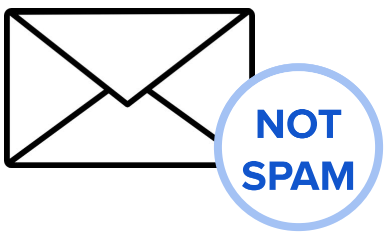
Novo email

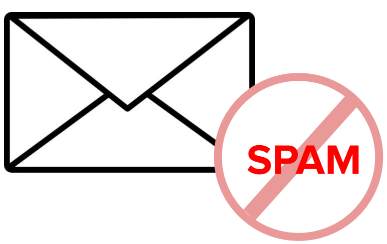


"Supondo um novo e-mail recebido, é mais provável que ele seja SPAM ou NOT SPAM?"

"Classe mais provável"

- Intuição básica

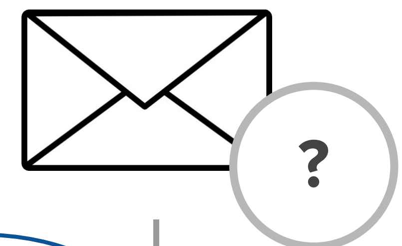


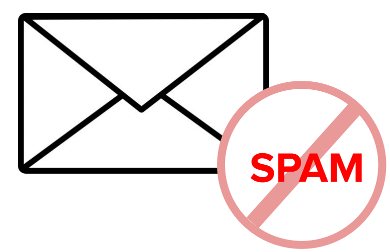


e-mail	Classe
1	SPAM
2	SPAM
3	NOT SPAM
4	SPAM
5	SPAM
6	SPAM
7	SPAM
8	NOT SPAM
9	NOT SPAM
10	SPAM

7/10 e-mails: SPAM

Novo email



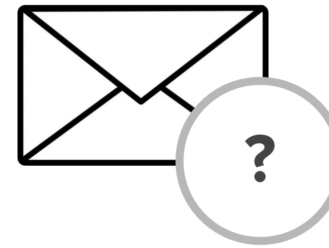


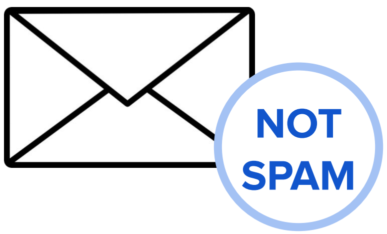
Probabilidade (SPAM) = 0.7

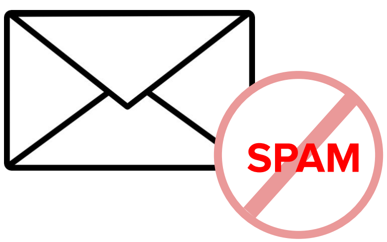
"Classe mais provável"

- Intuição básica

Novo email





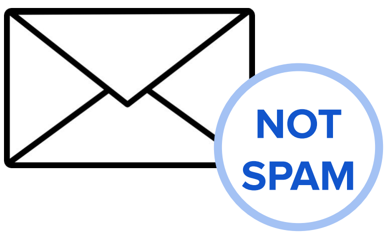


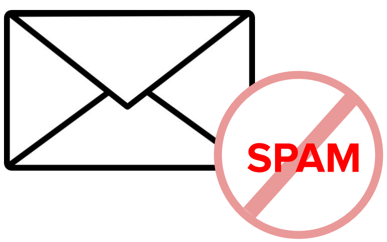
e-mail	Conteúdo	Classe
1	"password expired"	SPAM
2	"send password"	SPAM
3	"review conference"	NOT SPAM
4	"password review"	SPAM
5	"review account"	SPAM
6	"account password"	SPAM
7	"send account"	SPAM
8	"conference paper"	NOT SPAM
9	"send paper"	NOT SPAM
10	"expired account"	SPAM

"Supondo um novo e-mail recebido, com conteúdo "*review conference paper*", é mais provável que ele seja SPAM ou NOT SPAM?"

"Classe mais provável"

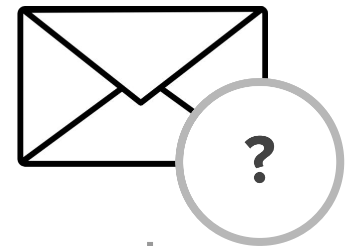
- Intuição básica

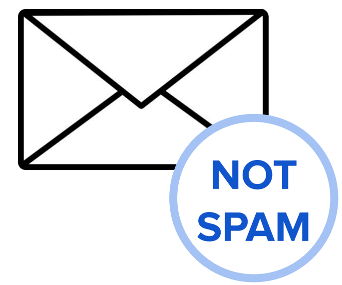




e-mail	Conteúdo	Classe
1	"password expired"	SPAM
2	"send password"	SPAM
3	**"review conference"**	NOT SPAM
4	**"password review"**	SPAM
5	**"review account"**	SPAM
6	"account password"	SPAM
7	"send account"	SPAM
8	**"conference paper"**	NOT SPAM
9	**"send paper"**	NOT SPAM
10	"expired account"	SPAM

Novo email

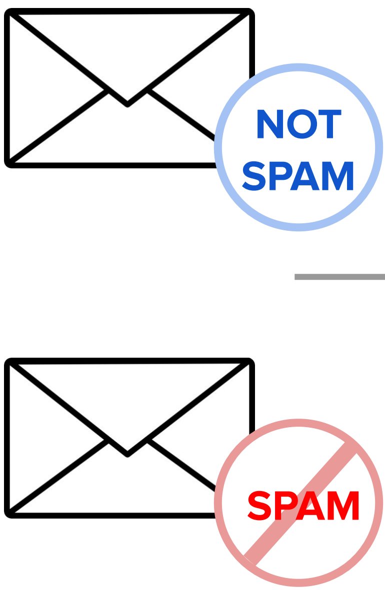




"Supondo um novo e-mail recebido, com conteúdo "*review conference paper*", é mais provável que ele seja SPAM ou NOT SPAM?"

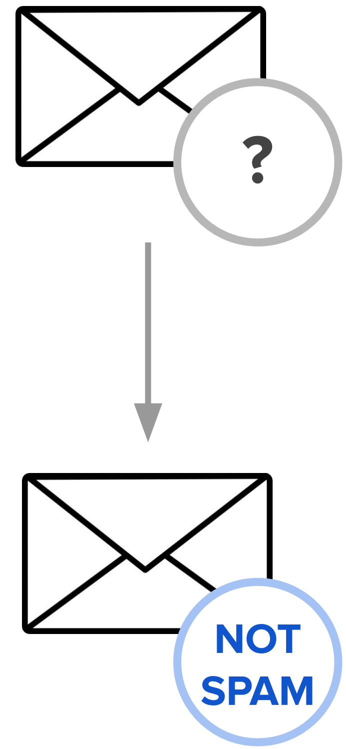
"Classe mais provável"

- Intuição básica



e-mail	Conteúdo	Classe
1	"password expired"	SPAM
2	"send password"	SPAM
3	**"review conference"**	NOT SPAM
4	**"password review"**	SPAM
5	**"review account"**	SPAM
6	"account password"	SPAM
7	"send account"	SPAM
8	**"conference paper"**	NOT SPAM
9	**"send paper"**	NOT SPAM
10	"expired account"	SPAM

Novo email

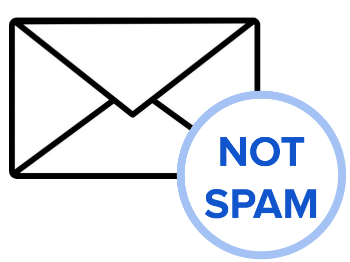


Qual a classe mais provável dado o conteúdo do e-mail?

Aprendizado Bayesiano

- Assumindo:
  - $\mathbf{x}$ é a entrada, um vetor com valores dos atributos preditivos
    - Isto é, seu conteúdo (as palavras contidas no e-mail)
  - y é a saída, um valor categórico em $\{c_1, c_2, \ldots, c_m\}$
    - Isto é, a classe definindo se é SPAM ou NOT SPAM
- Aprendizado Bayesiano visa encontrar a classe y_i que possui a <u>maior probabilidade</u> de explicar o exemplo de entrada $\mathbf{x}$

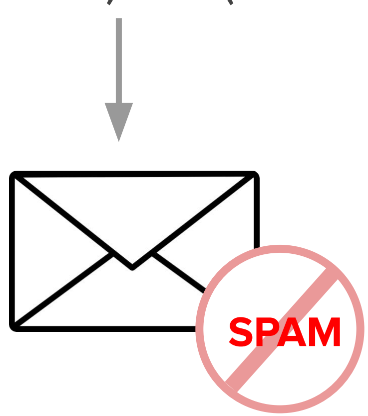
IF P("SPAM" | Conteúdo) < P("NOT SPAM" | Conteúdo)



Aprendizado Bayesiano

- Assumindo:
  - $\mathbf{x}$ é a entrada, um vetor com valores dos atributos preditivos
    - Isto é, seu conteúdo (as palavras contidas no e-mail)
  - y é a saída, um valor categórico em $\{c_1, c_2, \ldots, c_m\}$
    - Isto é, a classe definindo se é SPAM ou NOT SPAM
- Aprendizado Bayesiano visa encontrar a classe y_i que possui a <u>maior probabilidade</u> de explicar o exemplo de entrada $\mathbf{x}$

IF P("SPAM" | Conteúdo) **>** P("NOT SPAM" | Conteúdo)



Aprendizado Bayesiano

- Assumindo:
  - $\mathbf{x}$ é a entrada, um vetor com valores dos atributos preditivos
    - Isto é, seu conteúdo (as palavras contidas no e-mail)
  - y é a saída, um valor categórico em $\{c_1, c_2, \ldots, c_m\}$
    - Isto é, a classe definindo se é SPAM ou NOT SPAM
- Aprendizado Bayesiano visa encontrar a classe y_i que possui a <u>maior probabilidade</u> de explicar o exemplo de entrada $\mathbf{x}$
  - Formalmente:

$$y_{MAP} = \underset{i}{\operatorname{argmax}}\, P(y_i|\mathbf{x})$$

  - Retorna a classe y_i com maior probabilidade de estar associada a $\mathbf{x}$, isto é, aquela que possui <u>valor máximo</u> de $P(y_i \mid \mathbf{x})$

Aprendizado Bayesiano

- Assumindo:
  - $\mathbf{x}$ é a entrada, um vetor com valores dos atributos preditivos
    - Isto é, seu conteúdo (as palavras contidas no e-mail)
  - y é a saída, um valor categórico em $\{c_1, c_2, \ldots, c_m\}$
    - Isto é, a classe definindo se é SPAM ou NOT SPAM
- Aprendizado Bayesiano visa encontrar a classe y_i que possui a maior probabilidade de explicar o exemplo de entrada $\mathbf{x}$
  - Formalmente:

$$\underbrace{y_{MAP}}_{\text{Maximum a Posteriori}} = \underset{i}{\operatorname{argmax}} \underbrace{P(y_i|\mathbf{x})}_{\text{Probabilidade a posteriori}}$$

  - Retorna a classe y_i com maior probabilidade de estar associada a $\mathbf{x}$, isto é, aquela que possui valor máximo de $P(y_i \mid \mathbf{x})$

Aprendizado Bayesiano

- Assumindo:
  - $\mathbf{x}$ é a entrada, um vetor com valores dos atributos preditivos
    - Isto é, seu conteúdo (as palavras contidas no e-mail)
  - y é a saída, um valor categórico em $\{c_1, c_2, \ldots, c_m\}$
    - Isto é, a classe definindo se é SPAM ou NOT SPAM

Como calcular $P(y_i|\mathbf{x})$?

$$\underbrace{y_{MAP}}_{\text{Maximum a Posteriori}} = \underset{i}{\operatorname{argmax}}\ \underbrace{P(y_i|\mathbf{x})}_{\text{Probabilidade a posteriori}}$$

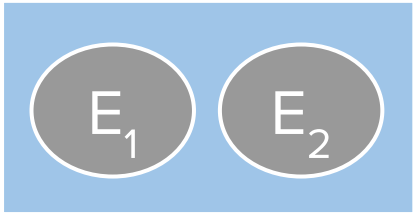
Teoria da probabilidade

Teorema de Bayes

Teoria da Probabilidade
Axiomas de Kolmogorof

- Considere um evento E no espaço amostral finito (**Ω**).
  - P(E) é a probabilidade do evento E ocorrer, dada pelo número de ocorrências (pontos) favoráveis a E dividido pelo número total de ocorrências (pontos) no espaço amostral **Ω**
- Probabilidade P(E) satisfaz os axiomas:

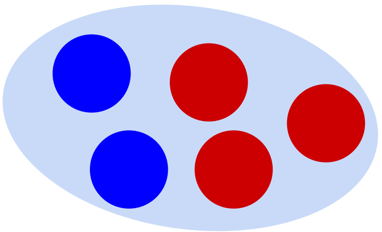
1. 0 ≤ P(E)
2. Se **Ω** é o espaço de todos os possíveis eventos, então P(**Ω**) = 1
3. Se E_1 e E_2 são eventos disjuntos, então:



$$P(E_1 \cup E_2) = \underbrace{P(E_1)}_{\text{Probabilidade a priori (marginal)}} + P(E_2)$$

Teoria da Probabilidade

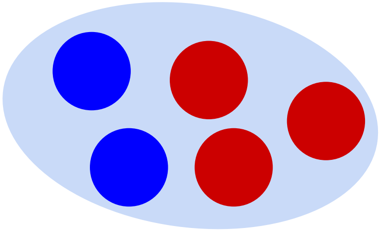
Probabilidade Condicional



- Suponha os eventos:
  - A: a primeira bola sorteada é azul
  - B: a segunda bola sorteada é azul

- P(A) = 2/5
- P(B): 1/4 se A ocorre; 2/4 caso contrário. P(B|A) = 1/4

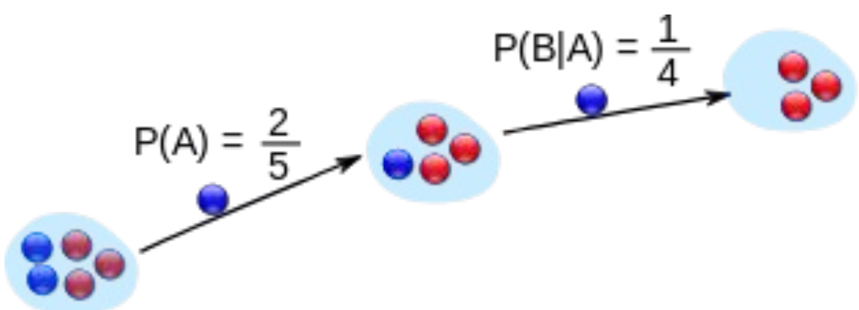
Teoria da Probabilidade
Probabilidade Condicional



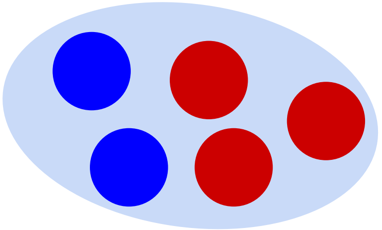
- Suponha os eventos:
  - A: a primeira bola sorteada é azul
  - B: a segunda bola sorteada é azul

- P(A) = 2/5
- P(B): 1/4 se A ocorre; 2/4 caso contrário. P(B|A) = 1/4

- Suponha agora o evento:
  - A e B: as duas primeiras bolas sorteadas são azuis
  - Probabilidade conjunta: $P(A \cap B) = ?$



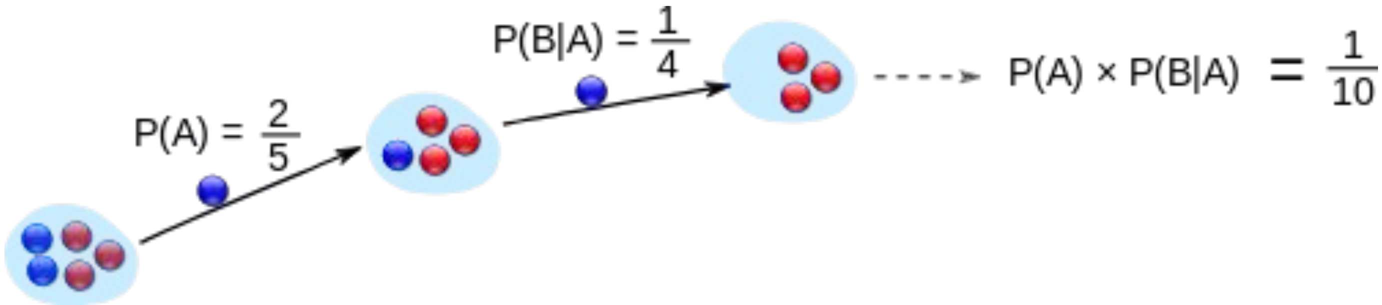
Teoria da Probabilidade
Probabilidade Condicional



- Suponha os eventos:
  - A: a primeira bola sorteada é azul
  - B: a segunda bola sorteada é azul

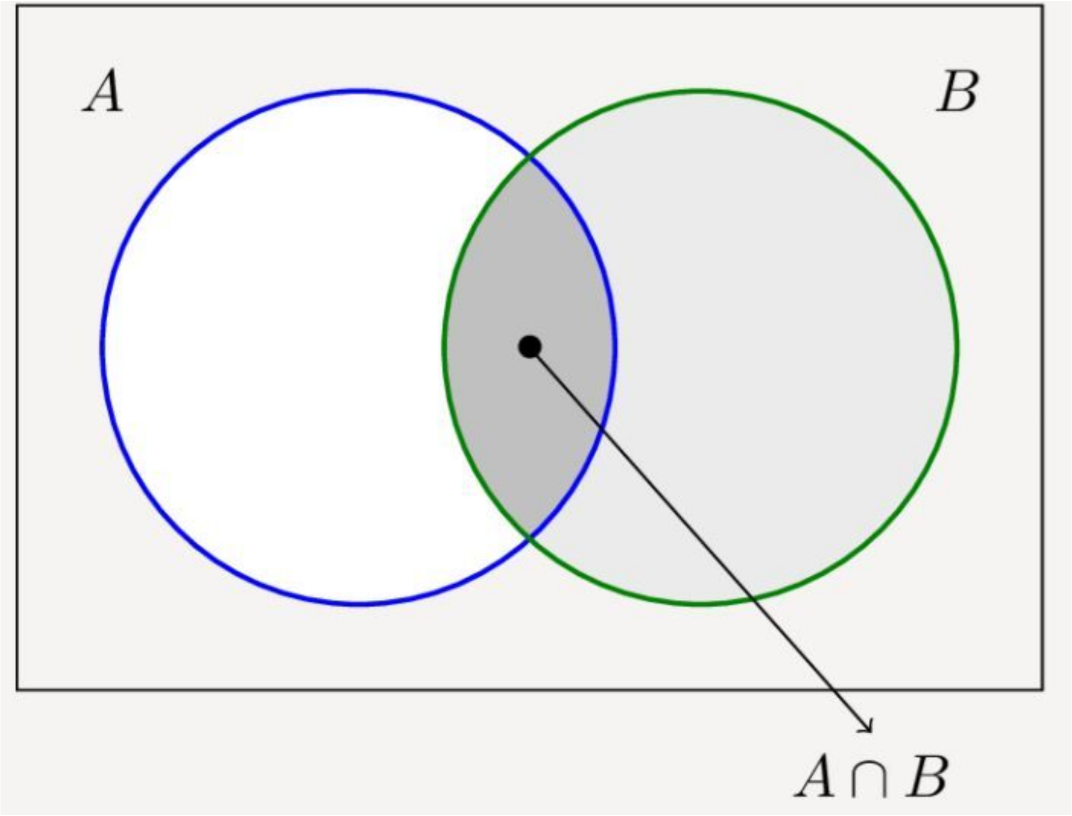
- P(A) = 2/5
- P(B): 1/4 se A ocorre; 2/4 caso contrário. P(B|A) = 1/4

- Suponha agora o evento:
  - A e B: as duas primeiras bolas sorteadas são azuis
  - Probabilidade conjunta: $P(A \cap B) = P(A) \times P(B|A)$



Teoria da Probabilidade
Lei da Probabilidade Condicional

- Supondo A e B genéricos



Teoria da Probabilidade

Lei da Probabilidade Condicional

- Supondo as probabilidades conjuntas $P(A \cap B)$ e $P(B \cap A)$, define-se as respectivas probabilidades condicionais $P(A|B)$ e $P(B|A)$:

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

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}} \qquad\qquad \boxed{P(B|A) = \frac{P(B \cap A)}{P(A)}}$$

- A seguinte igualdade é verdadeira

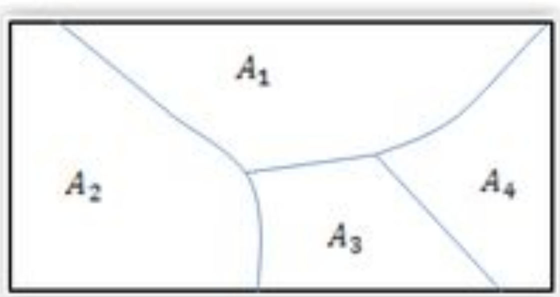
$$P(A \cap B) = P(B \cap A)$$

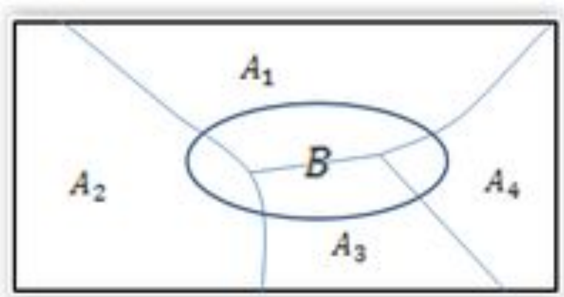
Teoria da Probabilidade

Lei da Probabilidade Total

- Se os eventos A_1, A_2, ...A_n são eventos disjuntos mensuráveis cuja união forma uma partição em **Ω** , então para qualquer evento B no mesmo espaço de probabilidades, temos:

$$P(B) = \sum_{j=1}^{n} \underbrace{P(B|A_j)}_{\textbf{Probabilidade condicional}} \overbrace{P(A_j)}^{\textbf{Probabilidade a priori}}$$





Exemplo de trabalho:
Diagnóstico médico de Malária

- A probabilidade de um médico observar um paciente com malária em Porto Alegre é de 8%
- Um exame para diagnóstico laboratorial da malária possui um grau de incerteza.
  - Em 75% dos casos em que o resultado do exame foi positivo, a doença foi confirmada no paciente (SENSIBILIDADE)
  - Em 96% dos casos em que o resultado do exame foi negativo, o paciente não tinha a doença (ESPECIFICIDADE)

O que podemos dizer sobre um paciente se o teste deu positivo?
Ele está doente?

Exemplo de trabalho:
Diagnóstico médico de Malária

- A probabilidade de um médico observar um paciente com malária em Porto Alegre é de 8%
- Um exame para diagnóstico laboratorial da malária possui um grau de incerteza.
  - Em 75% dos casos em que o resultado do exame foi positivo, a doença foi confirmada no paciente (SENSIBILIDADE)
  - Em 96% dos casos em que o resultado do exame foi negativo, o paciente não tinha a doença (ESPECIFICIDADE)

Probabilidade a priori

P(Malaria=presente) = **0.08**

P(Malaria=ausente) = 1 - P(Malaria=presente) = **0.92**

Exemplo de trabalho:
Diagnóstico médico de Malária

- A probabilidade de um médico observar um paciente com malária em Porto Alegre é de 8%
- Um exame para diagnóstico laboratorial da malária possui um grau de incerteza.
  - Em 75% dos casos em que o resultado do exame foi positivo, a doença foi confirmada no paciente (SENSIBILIDADE)
  - Em 96% dos casos em que o resultado do exame foi negativo, o paciente não tinha a doença (ESPECIFICIDADE)

Probabilidades condicionais

P(Exame=positivo | Malaria=presente) = **0.75**

P(Exame=negativo | Malaria=ausente) = **0.96**

Exame	Doença: presente	Doença: ausente
positivo	75%	?
negativo	?	96%

Exemplo de trabalho:

Diagnóstico médico de Malária

- A probabilidade de um médico observar um paciente com malária em Porto Alegre é de 8%
- Um exame para diagnóstico laboratorial da malária possui um grau de incerteza.
  - Em 75% dos casos em que o resultado do exame foi positivo, a doença foi confirmada no paciente (SENSIBILIDADE)
  - Em 96% dos casos em que o resultado do exame foi negativo, o paciente não tinha a doença (ESPECIFICIDADE)

Probabilidades condicionais

P(Exame=positivo | Malaria=presente) = **0.75**

P(Exame=negativo | Malaria=ausente) = **0.96**

Eventos complementares. Somas = 1

Exame	Doença: presente	Doença: ausente
positivo	75%	?
negativo	?	96%

Exemplo de trabalho:
Diagnóstico médico de Malária

- A probabilidade de um médico observar um paciente com malária em Porto Alegre é de 8%
- Um exame para diagnóstico laboratorial da malária possui um grau de incerteza.
  - Em 75% dos casos em que o resultado do exame foi positivo, a doença foi confirmada no paciente (SENSIBILIDADE)
  - Em 96% dos casos em que o resultado do exame foi negativo, o paciente não tinha a doença (ESPECIFICIDADE)

Probabilidades condicionais

P(Exame=positivo | Malaria=presente) = **0.75**

P(Exame=negativo | Malaria=ausente) = **0.96**

Doença

Exame	presente	ausente
positivo	75%	4%
negativo	25%	96%

Exemplo de trabalho:

Diagnóstico médico de Malária

- Pela Lei da Probabilidade Total:

$$P(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j)$$

Doença

Exame	presente	ausente
positivo	75%	4%
negativo	25%	96%

- Podemos calcular a probabilidade a priori da variável Exame:

P(Exame=positivo) = P(Exame=positivo | Malaria=presente) * P(Malaria=presente)
+ P(Exame=positivo | Malaria=ausente) * P(Malaria=ausente)
= 0.75 * 0.08 + 0.04*0.92 = **0.0968**

P(Exame=negativo) = P(Exame=negativo | Malaria=presente) * P(Malaria=presente)
+ P(Exame=negativo | Malaria=ausente) * P(Malaria=ausente)
= 0.25 * 0.08 + 0.96*0.92 = **0.9032**

Exemplo de trabalho:

Diagnóstico médico de Malária

- Com base nos cálculos e no enunciado, temos:

Probabilidades a priori

P(Malaria=presente) = 0.08
P(Malaria=ausente) = 0.92

P(Exame=positivo) = 0.0968
P(Exame=negativo) = 0.9032

Probabilidades condicionais (verossimilhança)

Doença

Exame		presente	ausente
	positivo	75%	4%
	negativo	25%	96%

O que podemos dizer sobre o paciente se o teste deu positivo?
Ele está doente? Qual a probabilidade P(Malaria=presente | exame=positivo)?

Teorema de Bayes

Cálculo da Probabilidade a Posteriori

[ou "como calcular P(A | B) se eu sei P(B | A)"]

- Dada a igualdade

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

- Da definição de probabilidade conjunta, temos que:

$$P(A|B)P(B) = P(B|A)P(A)$$

$$\downarrow$$

Teorema de Bayes

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

Teorema de Bayes

Cálculo da Probabilidade a Posteriori

[ou "como calcular P(A | B) se eu sei P(B | A)"]

Probabilidade
a priori
(conhecimento prévio)

+

Evidências
(observações)

→

Probabilidade
a posteriori

Teorema de Bayes

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

Aprendizado Bayesiano

na classificação

- Assumindo:
  - $\mathbf{x}$ é a entrada, um vetor com valores dos atributos preditivos
  - y é a saída, um valor categórico em $\{c_1, c_2, \ldots, c_m\}$
- Objetivo:
  - Encontrar a classe que maximize a probabilidade a posteriori

$$y_{MAP} = \underset{i}{\operatorname{argmax}}\, \underline{P(y_i|\mathbf{x})}$$

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

Teorema de Bayes

$$\boxed{P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{P(\mathbf{x})}}$$

Aprendizado Bayesiano
na classificação

- Componentes:

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{P(\mathbf{x})}$$

$\mathbf{x}$: sintomas de um paciente
y: paciente possui malária?

- P(y): probabilidade a priori de cada classe
  - Probabilidade estimada antes de qualquer evidência extra, com base em conhecimento do domínio
  - Quais classes são comuns e quais são raras?
  - *A priori, é mais provável que um paciente atendido em Porto Alegre tenha Gripe do que Malária, por exemplo..*

Aprendizado Bayesiano
na classificação

- Componentes:

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{P(\mathbf{x})}$$

$\mathbf{x}$: sintomas de um paciente
y: paciente possui malária?

- $P(x|y)$: modelo condicionado à classe
  - Probabilidade condicional de $\mathbf{x}$ dado y
  - Descreve o quão provável é ter uma observação $\mathbf{x}$ para a classe y
  - *Assumindo que o paciente tem Malária, os sintomas são plausíveis?*

Aprendizado Bayesiano
na classificação

- Componentes:

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{P(\mathbf{x})}$$

Teoria da Probabilidade Total

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{\sum_{i=1}^{m} P(\mathbf{x}|y_i)P(y_i)}$$

$\mathbf{x}$: sintomas de um paciente
y: paciente possui malária?

- P(x): probabilidade a priori de uma observação
    - Normaliza as probabilidades entre as observações
    - Não está conectada a nenhuma classe específica: é uma constante
    - Como não está conectado a uma classe específica, não afeta a predição de qual classe é mais provável ($\arg\max$) e por isso é normalmente ignorado no cálculo

Exemplo de trabalho:

Diagnóstico médico de Malária

- Com base nos cálculos e no enunciado, temos:

Probabilidades a priori

P(Malaria=presente) = 0.08
P(Malaria=ausente) = 0.92

P(Exame=positivo) = 0.0968
P(Exame=negativo) = 0.9032

Probabilidades condicionais (verossimilhança)

Doença

Exame	presente	ausente
positivo	75%	4%
negativo	25%	96%

Probabilidades condicionais (Verossimilhança)

Retomando....

O que podemos dizer sobre o paciente se o teste deu positivo?
Ele está doente? Qual a probabilidade P(Malaria=presente | exame=positivo)?

Exemplo de trabalho:

Diagnóstico médico de Malária

O que podemos dizer sobre o paciente se o teste deu positivo?
Ele está doente? Qual a probabilidade P(Malaria=presente | exame=positivo)?

Probabilidades a priori

P(Malaria=presente) = 0.08
P(Malaria=ausente) = 0.92

P(Exame=positivo) = 0.0968
P(Exame=negativo) = 0.9032

		Doença	
		presente	ausente
Exame	positivo	75%	4%
Exame	negativo	25%	96%

Probabilidades condicionais (Verossimilhança)

P(Malaria=presente | Exame=positivo)
= P(Exame = positivo | Malaria = presente) * P (Malaria=presente)
= 0.75 * 0.08 = 0.06

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{\cancel{P(\mathbf{x})}}$$

Teorema de Bayes

P(Malaria = ausente | Exame=positivo)
= P(Exame = positivo | Malaria = ausente) * P (Malaria=ausente)
= 0.04 * 0.92 = 0.0368 = 0.0368

Exemplo de trabalho:

Diagnóstico médico de Malária

O que podemos dizer sobre o paciente se o teste deu positivo?
Ele está doente? Qual a probabilidade P(Malaria=presente | exame=positivo)?

Probabilidades a priori

P(Malaria=presente) = 0.08
P(Malaria=ausente) = 0.92

P(Exame=positivo) = 0.0968
P(Exame=negativo) = 0.9032

Exame \ Doença	presente	ausente
positivo	75%	4%
negativo	25%	96%

Probabilidades condicionais (Verossimilhança)

P(Malaria=presente | Exam
= P(Exame = positivo | Mala
= 0.75 * 0.08 = **0.06**

0.06 > 0.0368:
Por inferência Bayesiana, a classe mais provável é "Malaria presente":

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{\cancel{P(\mathbf{x})}}$$

Teorema de Bayes

P(Malaria = ausente | Exame=positivo)
= P(Exame = positivo | Malaria = ausente) * P (Malaria=ausente)
= 0.04 * 0.92 = 0.0368 = **0.0368**

Exemplo de trabalho:

Diagnóstico médico de Malária

O que podemos dizer sobre o paciente se o teste deu positivo?
Ele está doente? Qual a probabilidade P(Malaria=presente | exame=positivo)?

Probabilidades a priori

P(Malaria=presente) = 0.08
P(Malaria=ausente) = 0.92

P(Exame=positivo) = 0.0968
P(Exame=negativo) = 0.9032

Doença

Exame	presente	ausente
positivo	75%	4%
negativo	25%	96%

Probabilidades condicionais (Verossimilhança)

P(Malaria=presente | Exame=positivo)
= P(Exame = positivo | Malaria = presente) * P (Malaria=presente)
= 0.75 * 0.08 = 0.06

Divisão por P(Exame=positivo)

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{P(\mathbf{x})}$$

Teorema de Bayes

P(Malaria = ausente | Exame=positivo)
= P(Exame = positivo | Malaria = ausente) * P (Malaria=ausente)
= 0.04 * 0.92 = 0.0368 = 0.0368

Exemplo de trabalho:

Diagnóstico médico de Malária

O que podemos dizer sobre o paciente se o teste deu positivo?
Ele está doente? Qual a probabilidade P(Malaria=presente | exame=positivo)?

Probabilidades a priori

P(Malaria=presente) = 0.08
P(Malaria=ausente) = 0.92

P(Exame=positivo) = 0.0968
P(Exame=negativo) = 0.9032

Doença

Exame	presente	ausente
positivo	75%	4%
negativo	25%	96%

Probabilidades condicionais (Verossimilhança)

P(Malaria=presente | Exame=positivo)
= P(Exame = positivo | Malaria = presente) * P (Malaria=presente)
= 0.75 * 0.08 = 0.06
= 0.06 / 0.0968 = **0.6198**

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{P(\mathbf{x})}$$

Teorema de Bayes

P(Malaria = ausente | Exame=positivo)
= P(Exame = positivo | Malaria = ausente) * P (Malaria=ausente)
= 0.04 * 0.92 = 0.0368 = 0.0368
= 0.0368 / 0.0968 = **0.3802**

Probabilidades a posteriori (Soma = 1)

Problema do dia

- Um restaurante deseja descobrir sob que condições climáticas se vende mais Feijoada ou Filé à Parmegiana.
- Objetivo: preparar melhor a *mise en place* da cozinha

?

Problema do dia

- Um restaurante deseja descobrir sob que condições climáticas se vende mais Feijoada ou Filé à Parmegiana.
- Objetivo: preparar melhor a *mise en place* da cozinha

Mise en place (posta no lugar): "pôr em ordem"; organização de utensílios, ingredientes, etc... que precede a preparação de uma receita

- A equipe de vendas coletou os seguintes dados *(próximo slide)*

Problema do dia

Classe

	x_k^1	x_k^2	x_k^3	x_k^4	y_k
	Previsão	**Temperatura**	**Umidade**	**Vento**	**Prato**
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

Classificação probabilística

Estimativa Maximum a Posteriori

- Considerando um exemplo desconhecido $\mathbf{x}$, para cada classe y_i em $\{c_1, c_2, \ldots, c_m\}$, calcule:

$$P(y_i|\mathbf{x}) = \frac{P(\mathbf{x}|y_i)P(y_i)}{P(\mathbf{x})}$$

- Retornando como saída:

$$y_{MAP} = \underset{i}{\operatorname{argmax}} P(y_i|\mathbf{x})$$

Classificação probabilística

Estimativa Maximum a Posteriori

- Considerando um exemplo desconhecido $\mathbf{x}$, para cada classe y_i em $\{c_1, c_2, \ldots, c_m\}$, calcule:

$$P(y_i|\mathbf{x}) = \frac{P(\mathbf{x}|y_i)P(y_i)}{\cancel{P(\mathbf{x})}}$$

Para encontrar a classe com maior probabilidade, este termo não precisa ser calculado, pois é constante para todas as classes

- Retornando como saída:

$$y_{MAP} = \underset{i}{\operatorname{argmax}} P(y_i|\mathbf{x})$$

Maximum a Posteriori

Classificação probabilística
Estimativa Maximum a Posteriori

- Considerando um exemplo desconhecido $\mathbf{x}$, para cada classe y_i em $\{c_1, c_2, \ldots, c_m\}$, calcule:

$$P(y_i|\mathbf{x}) = \frac{P(\mathbf{x}|y_i)\cancel{P(y_i)}}{\cancel{P(\mathbf{x})}}$$

Nos casos em que podemos assumir que esta probabilidade é uniforme (ou quando é difícil estimá-la), é comum ignorarmos este termo

- Retornando como saída:

$$y_{MAP} = \underset{i}{\operatorname{argmax}}\, P(y_i|\mathbf{x})$$

Maximum a Posteriori

Simplificação

$$y_{ML} = \underset{i}{\operatorname{argmax}}\, P(\mathbf{x}|y_i)$$

Maximum Likelihood

O método MAP se resume a encontrar a hipótese que maximiza a verossimilhança P(Dados | Hipótese): tem <u>aplicabilidade limitada</u> pois o cálculo $P(x_1,\ldots,x_d|y)$ para conjuntos de dados com muitos atributos é computacionalmente custoso ou impraticável

Naïve Bayes

- Ideia "naïve": assumir que os valores dos atributos de um exemplo $\mathbf{x}$ são independentes entre si dada a classe y
- Adota a suposição de independência condicional:

$$P(\mathbf{x}|y) = P(x^1, ..., x^d|y) = \underbrace{\prod_{j=1}^{d} P(x^j|x^1, ...x^{j-1}, y)}_{\text{Regra da Cadeia (exata)}} = \underbrace{\prod_{j=1}^{d} P(x^j|y)}_{\text{Independência condicional}}$$

Naïve Bayes

- Ideia "naïve": assumir que os valores dos atributos de um exemplo $\mathbf{x}$ são independentes entre si dada a classe y
- Adota a suposição de independência condicional:

$$P(\mathbf{x}|y) = P(x^1, ..., x^d|y) = \underbrace{\prod_{j=1}^{d} P(x^j|x^1, ...x^{j-1}, y)}_{\text{Regra da Cadeia (exata)}} = \underbrace{\prod_{j=1}^{d} P(x^j|y)}_{\text{Independência condicional}}$$

- Substituindo na fórmula do Teorema de Bayes:

$$P(y_i|\mathbf{x}) = \frac{P(\mathbf{x}|y_i)P(y_i)}{\cancel{P(\mathbf{x})}} \longrightarrow \boxed{P(y_i|\mathbf{x}) = P(y_i)\prod_{j=1}^{d} P(x^j|y_i)}$$

Naïve Bayes

- Ideia "naïve": assumir que os valores dos atributos de um exemplo $\mathbf{x}$ são independentes entre si dada a classe y
- Adota a suposição de independência condicional:

$$P(\mathbf{x}|y) = P(x^1, ..., x^d|y) = \prod^{d} P(x^j|x^1, ...x^{j-1}, y) = \prod^{d} P(x^j|y)$$

Como estimar as probabilidades envolvidas?

- Substituindo na fórmula do Teorema de Bayes:

$$P(y_i|\mathbf{x}) = \frac{P(\mathbf{x}|y_i)P(y_i)}{\cancel{P(\mathbf{x})}} \longrightarrow \boxed{P(y_i|\mathbf{x}) = P(y_i) \prod_{j=1}^{d} P(x^j|y_i)}$$

Naïve Bayes

- Todas as **probabilidades utilizadas são estimadas a partir dos dados de treinamento** $\mathbf{D} = \{(\mathbf{x}_k, f(\mathbf{x}_k)), k = 1,\ldots n\}$, com cada exemplo dado por um vetor de atributos $\mathbf{x_k} = \{x_k^1, x_k^2, \ldots, x_k^d\}$ formado pelo valor observado para os d atributos $A_1, \ldots, A_d$

$$P(y_i|\mathbf{x}) = P(y_i) \prod_{j=1}^{d} P(x^j|y_i)$$

  - $P(y_i)$: probabilidade da classe y_i no conjunto de treinamento
    - Calculada como a razão entre quantas instâncias pertencem à classe y_i e o número total de exemplos no conjunto D
  - $P(x^j|y_i)$: probabilidade do valor de atributo A_j assumir um valor x^j dada a distribuição de classe y_i
    - cálculo depende do tipo do atributo A_j, categórico ou numérico (contínuo). veremos mais adiante...

Aplicação de Naïve Bayes para classificação de dados estruturados

Estimativa das probabilidades com atributos categóricos, numéricos e mistos.

Algoritmo Naïve Bayes

Entradas:

Conjunto de treinamento, $\mathbf{D} = \{(\mathbf{x}_k, f(\mathbf{x}_k)), k = 1,...n\}$

Instância de teste, com rótulo desconhecido, $\boldsymbol{t} = \{\mathbf{x}_t, y_t = ?\}$

Saída:

Rótulo estimado para a entrada $\boldsymbol{t}$, y_t

1. Calcule a probabilidade a priori $P(y_i)$ para cada classe y_i em $\{c_1, c_2, \ldots, c_m\}$
2. Para cada classe y_i, faça:
3. Inicialize uma variável acumuladora para o produtório (prod_i) com valor 1
4. Para cada atributo preditivo x_t^j em $\mathbf{x}_t$, com $j = 1,...,d$ faça:
5. Recupere os exemplos de D com $x_k^j = x_t^j$
6. Organize os exemplos recuperados de acordo com seus rótulos y_k
7. Calcule $P(x_k^j|y_i)$, com $y_i = y_k$, e atualize prod_i como $\mathrm{prod}_i = \mathrm{prod}_i * P(x_k^j|y_i)$
8. Faça $P(y_i|\mathbf{x}_t) = P(y_i) * \mathrm{prod}_i$
9. Encontre a classe y_i que **maximize** $P(y_i|\mathbf{x}_t)$ e a retorne como rótulo y_t

Algoritmo Naïve Bayes

Entradas:

Conjunto de treinamento, $\mathbf{D} = \{(\mathbf{x}_k, f(\mathbf{x}_k)), \text{k} = 1,\ldots,\text{n}\}$

Instância de teste, com rótulo desconhecido, $t = \{\mathbf{x}_t, y_t = ?\}$

Todas as probabilidades podem ser pré-computadas e armazenadas em uma etapa de "treinamento" do modelo

1. Calcule a probabilidade a priori $\text{P}(y_i)$ para cada classe y_i em $\{c_1, c_2, \ldots, c_m\}$
2. Para cada classe y_i, faça:
3. Inicialize uma variável acumuladora para o produtório (prod_i) com valor 1
4. Para cada atributo preditivo x_t^j em $\mathbf{x}_t$, com $\text{j} = 1,\ldots,\text{d}$ faça:
5. Recupere os exemplos de D com $\text{x}_k^j = \text{x}_t^j$
6. Organize os exemplos recuperados de acordo com seus rótulos y_k
7. Calcule $\text{P}(\text{x}_k^j|\text{y}_i)$, com $\text{y}_i = \text{y}_k$, e atualize prod_i como $\text{prod}_i = \text{prod}_i * \text{P}(\text{x}_k^j|\text{y}_i)$
8. Faça $\text{P}(\text{y}_i|\mathbf{x}_t) = \text{P}(y_i) * \text{prod}_i$
9. Encontre a classe y_i que **maximize** $\text{P}(\text{y}_i|\mathbf{x}_t)$ e a retorne como rótulo y_t

Algoritmo Naïve Bayes
com atributos categóricos

- Um restaurante deseja descobrir sob que condições climáticas se vende mais Feijoada ou Filé à Parmegiana.
- Objetivo: preparar melhor a *mise en place* da cozinha

?

Algoritmo Naïve Bayes

com atributos categóricos

Classe

	Previsão (x_k^1)	Temperatura (x_k^2)	Umidade (x_k^3)	Vento (x_k^4)	Prato (y_k)
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

Objetivo: Encontrar a classe y_i que **maximize** $P(y_i|\mathbf{x}_t)$ e a retornar como rótulo y_t

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	sol	frio	normal	sim	?

$P(\mathit{feijoada} \mid x_t)$?

$P(\mathit{parmegiana} \mid x_t)$?

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

1. Calcule a probabilidade a priori $P(y_i)$ para cada classe y_i em $\{c_1, c_2, \ldots, c_m\}$:

14 instâncias ao total

5 instâncias da classe parmegiana

9 instâncias da classe feijoada

$$P(parmegiana) = 5/14 = 0.36$$

$$P(feijoada) = 9/14 = 0.64$$

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	sol	frio	normal	sim	?

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

$$\mathrm{P}(\textit{parmegiana}) = 5/14 = 0.36$$

$$\mathrm{P}(\textit{feijoada}) = 9/14 = 0.64$$

2. Inicializar os produtórios para cada classe

$$\mathrm{prod}_p = 1$$

$$\mathrm{prod}_f = 1$$

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	sol	frio	normal	sim	?

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

$$P(\textit{parmegiana}) = 5/14 = 0.36$$

$$P(\textit{feijoada}) = 9/14 = 0.64$$

$$\text{prod}_p = 1$$

$$\text{prod}_f = 1$$

4. Avaliar cada atributo preditivo, calculando sua probabilidade condicionada à cada classe seguindo os passos 5, 6 e 7 do algoritmo

- Encontrar todas as instâncias com mesmo valor para o atributo preditivo analisado na instância de teste
- Contar as ocorrências de cada classe para este valor de atributo, calculando $P(x^j \mid y_i)$

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	$\boxed{\text{sol}}$	frio	normal	sim	?

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

$$P(parmegiana) = 5/14 = 0.36$$

$$P(feijoada) = 9/14 = 0.64$$

$$prod_p = 1$$

$$prod_f = 1$$

- 5 exemplos possuem $x_k^1 = \text{sol}$
- $P(\text{sol} \mid \text{parmegiana}) = 3/5 = 0.6$
- $P(\text{sol} \mid \text{feijoada}) = 2/9 = 0.22$

$$prod_p = prod_p * P(\text{sol} \mid \text{parmegiana}) = 0.6$$

$$prod_f = prod_f * P(\text{sol} \mid \text{feijoada}) = 0.22$$

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	sol	frio	normal	sim	?

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

$P(\textit{parmegiana}) = 5/14 = 0.36$

$P(\textit{feijoada}) = 9/14 = 0.64$

$prod_p = 0.6$

$prod_f = 0.22$

- 4 exemplos possuem $x_k^2 = \text{frio}$
- $P(\text{frio} \mid \text{parmegiana}) = 1/5 = 0.2$
- $P(\text{frio} \mid \text{feijoada}) = 3/9 = 0.33$

$prod_p = prod_p * P(\text{frio} \mid \text{parmegiana}) = 0.12$

$prod_f = prod_f * P(\text{frio} \mid \text{feijoada}) = 0.0726$

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	sol	frio	normal	sim	?

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

$\mathrm{P}(\textit{parmegiana}) = 5/14 = 0.36$

$\mathrm{P}(\textit{feijoada}) = 9/14 = 0.64$

$\mathrm{prod}_p = 0.12$

$\mathrm{prod}_f = 0.0726$

- 7 exemplos possuem $x_k^{\;3} = \mathrm{normal}$
- $\mathrm{P}(\mathrm{normal} \mid \mathrm{parmegiana}) = 1/5 = 0.2$
- $\mathrm{P}(\mathrm{normal} \mid \mathrm{feijoada}) = 6/9 = 0.66$

$\mathrm{prod}_p = \mathrm{prod}_p * \mathrm{P}(\mathrm{normal} \mid \mathrm{parmegiana}) = 0.024$

$\mathrm{prod}_f = \mathrm{prod}_f * \mathrm{P}(\mathrm{normal} \mid \mathrm{feijoada}) = 0.0479$

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	sol	frio	normal	sim	?

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

$\mathrm{P}(\textit{parmegiana}) = 5/14 = 0.36$

$\mathrm{P}(\textit{feijoada}) = 9/14 = 0.64$

$\mathrm{prod}_p = 0.024$

$\mathrm{prod}_f = 0.0479$

- 6 exemplos possuem $x_k^4 = \mathrm{sim}$
- $\mathrm{P}(\mathrm{sim} \mid \mathrm{parmegiana}) = 3/5 = 0.6$
- $\mathrm{P}(\mathrm{sim} \mid \mathrm{feijoada}) = 3/9 = 0.33$

$\mathrm{prod}_p = \mathrm{prod}_p * \mathrm{P}(\mathrm{sim} \mid \mathrm{parmegiana}) = 0.0144$

$\mathrm{prod}_f = \mathrm{prod}_f * \mathrm{P}(\mathrm{sim} \mid \mathrm{feijoada}) = 0.0158$

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	sol	frio	normal	$\boxed{\text{sim}}$	?

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

$$P(\textit{parmegiana}) = 5/14 = 0.36$$

$$P(\textit{feijoada}) = 9/14 = 0.64$$

$$\text{prod}_p = 0.0144$$

$$\text{prod}_f = 0.0158$$

8. Calcular a probabilidade a posteriori para cada classe

$$P(\textit{feijoada} \mid x_t) = \text{prod}_f * P(\textit{feijoada}) = 0.0158 * 0.64 = 0.011$$

$$P(\textit{parmegiana} \mid x_t) = \text{prod}_p * P(\textit{parmegiana}) = 0.0144 * 0.36 = 0.005$$

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	sol	frio	normal	sim	?

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

$$P(parmegiana) = 5/14 = 0.36$$

$$P(feijoada) = 9/14 = 0.64$$

$$\text{prod}_p = 0.0144$$

$$\text{prod}_f = 0.0158$$

8. Calcular a probabilidade a posteriori para cada classe

$$P(feijoada \mid x_t) = \text{prod}_f * P(feijoada)$$
$$= 0.0158 * 0.64$$
$$= \mathbf{0.011}$$

$$P(parmegiana \mid x_t) = \text{prod}_p * P(parmegiana)$$
$$= 0.0144 * 0.36$$
$$= 0.005$$

A classe de maior probabilidade a posteriori é feijoada, assim $y_t = feijoada$

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	sol	frio	normal	sim	?

Algoritmo Naïve Bayes com atributos categóricos

	Previsão	Temperatura	Umidade	Vento	Prato
$\mathbf{x}_1$	chuva	frio	normal	sim	parmegiana
$\mathbf{x}_2$	chuva	moderado	alta	sim	parmegiana
$\mathbf{x}_3$	sol	quente	alta	não	parmegiana
$\mathbf{x}_4$	sol	quente	alta	sim	parmegiana
$\mathbf{x}_5$	sol	moderado	alta	não	parmegiana
$\mathbf{x}_6$	nublado	frio	normal	sim	feijoada
$\mathbf{x}_7$	nublado	quente	alta	não	feijoada
$\mathbf{x}_8$	nublado	quente	normal	não	feijoada
$\mathbf{x}_9$	nublado	moderado	alta	sim	feijoada
$\mathbf{x}_{10}$	chuva	frio	normal	não	feijoada
$\mathbf{x}_{11}$	chuva	moderado	alta	não	feijoada
$\mathbf{x}_{12}$	chuva	moderado	normal	não	feijoada
$\mathbf{x}_{13}$	sol	frio	normal	não	feijoada
$\mathbf{x}_{14}$	sol	moderado	normal	sim	feijoada

$$P(parmegiana) = 5/14 = 0.36$$

$$P(feijoada) = 9/14 = 0.64$$

$$\text{prod}_p = 0.0144$$

$$\text{prod}_f = 0.0158$$

8. Calcular a probabilidade a posteriori para cada classe

$$\begin{aligned} P(feijoada \mid x_t) &= \text{prod}_f * P(feijoada) \\ &= 0.0158 * 0.64 \\ &= \mathbf{0.011} \end{aligned}$$

$$\begin{aligned} P(parmegiana \mid x_t) &= \text{prod}_p * P(parmegiana) \\ &= 0.0144 * 0.36 \\ &= 0.005 \end{aligned}$$

A classe de maior probabilidade a posteriori é <u>feijoada</u>, assim $y_t = feijoada$

Instância de teste $\mathbf{x}_t$

$\mathbf{x}_t$	sol	frio	normal	sim	**feijoada**

O problema da frequência zero

- Para uma nova instância $x_t = \{ \text{chuva, quente, alta, não, ?} \}$, haveria chance de ser predita como Parmegiana?
- Dado que as probabilidades são estimadas a partir dos dados de treinamento, quando a frequência de um determinado valor de atributo é zero para uma classe, a Probabilidade $P(x^j|y_i) = 0$

Previsão	Temperatura	Umidade	Vento	Prato
chuva	frio	normal	sim	parmegiana
chuva	moderado	alta	sim	parmegiana
sol	quente	alta	sim	parmegiana
sol	quente	alta	sim	parmegiana
sol	moderado	alta	sim	parmegiana
nublado	frio	normal	sim	feijoada
nublado	quente	alta	não	feijoada
nublado	quente	normal	não	feijoada

O problema da frequência zero

- Para uma nova instância x_t = { chuva, quente, alta, não, ?}, haveria chance de ser predita como Parmegiana?
- Dado que as probabilidades são estimadas a partir dos dados de treinamento, quando a frequência de um determinado valor de atributo é zero para uma classe, a Probabilidade $P(x^j|y_i) = 0$
- x_t nunca seria predita como Parmegiana pois $P(\text{vento=não}|\mathit{Parmegiana})=0$
- Solução: nunca permitir zero!

Previsão	**Temperatura**	**Umidade**	**Vento**	**Prato**
chuva	frio	normal	sim	parmegiana
chuva	moderado	alta	sim	parmegiana
sol	quente	alta	sim	parmegiana
sol	quente	alta	sim	parmegiana
sol	moderado	alta	sim	parmegiana
nublado	frio	normal	sim	feijoada
nublado	quente	alta	não	feijoada
nublado	quente	normal	não	feijoada

Correção de Laplace

Solucionando o problema da frequência zero

- **Correção de Laplace**: assumimos ter uma instância a mais para cada valor possível do atributo com referência à classe analisada (efeito negligenciável para grandes conjuntos D)

$P(\text{vento=sim}|\mathit{Parmegiana}) = (5+1)/(\mathbf{5+2})$

$P(\text{vento=não}|\mathit{Parmegiana}) = (0+1)/(\mathbf{5+2})$

$P(\text{vento=sim}|\mathit{Feijoada}) = (1+1)/(\mathbf{3+2})$

$P(\text{vento=não}|\mathit{Feijoada}) = (2+1)/(\mathbf{3+2})$

5 exemplos

3 exemplos

Previsão	Temperatura	Umidade	Vento	Prato
chuva	frio	normal	sim	parmegiana
chuva	moderado	alta	sim	parmegiana
sol	quente	alta	sim	parmegiana
sol	quente	alta	sim	parmegiana
sol	moderado	alta	sim	parmegiana
nublado	frio	normal	sim	feijoada
nublado	quente	alta	não	feijoada
nublado	quente	normal	não	feijoada

Correção de Laplace

Solucionando o problema da frequência zero

- **Correção de Laplace**: assumimos ter uma instância a mais para cada valor possível do atributo com referência à classe analisada (efeito negligenciável para grandes conjuntos D)

$$P(\text{vento}=\text{sim}|\mathit{Parmegiana})=(5+1)/(\mathbf{5+2})$$

$$P(\text{vento}=\text{não}|\mathit{Parmegiana})=(0+1)/(\mathbf{5+2})$$

5 exemplos

$$P(\text{vento}=\text{sim}|\mathit{Feijoada})=(1+1)/(\mathbf{3+2})$$

$$P(\text{vento}=\text{não}|\mathit{Feijoada})=\underline{(2+1)}/\underline{(\mathbf{3+2})}$$

3 exemplos

Previsão	Temperatura	Umidade	Vento	Prato
chuva	frio	normal	sim	parmegiana
chuva	moderado	alta	sim	parmegiana
sol	quente	alta	sim	parmegiana
sol	quente	alta	sim	parmegiana
sol	moderado	alta	sim	parmegiana
nublado	frio	normal	sim	feijoada
nublado	quente	alta	não	feijoada
nublado	quente	normal	não	feijoada

2 instâncias com vento = não
+
1 instância "extra" para a este valor de atributo e a classe "feijoada"

3 instâncias ao total na classe "feijoada"
+
2 instâncias extras, uma para cada valor possível do atributo vento (i.e., "sim" e não")

Estimativa das probabilidades no algoritmo Naïve Bayes com atributos categóricos

- Aspectos principais
  - A **probabilidade a priori** de cada classe pode ser estimada facilmente como uma contagem do número de exemplos associados a classe y_i, $|y_i|$ e o número total de exemplos $|D|$:

$$P(y_i) = \frac{|y_i|}{|D|}, y_i \in \{c_1, ..., c_m\}$$

  - Para atributos categóricos, o conjunto de valores possíveis é **enumerável**. O cálculo da **probabilidade condicional** de se observar um determinado valor de atributo dado que o exemplo pertence a uma classe y_i envolve manter uma variável acumuladora do produtório para cada classe e contabilizar, para cada atributo x_k^j
    - Número de exemplos com $x_k^j = x_t^j$ e $y_k = y_i$, dividido por $|y_i|$

Estimativa das probabilidades no algoritmo Naïve Bayes com atributos categóricos

- Aspectos principais
  - A **probabilidade a priori** de cada classe pode ser estimada facilmente como uma contagem do número de exemplos associados a classe y_i, $|y_i|$ e o número total de exemplos $|D|$:

$$P(y_i) = \frac{|y_i|}{|D|}, y_i \in \{c_1, ..., c_m\}$$

  - Para atributos categóricos, o conjunto de valores possíveis é **enumerável**. O cálculo da **probabilidade condicional** de se observar um determinado valor de atributo dado que o exemplo pertence a uma classe y_i envolve manter uma variável acumuladora do produtório para cada classe e contabilizar, para cada atributo x_k^j
    - Número de exemplos com $x_k^j = x_t^j$ e $y_k = y_i$, dividido por $|y_i|$

Como proceder com o cálculo da probabilidade condicional quando o atributo é <u>numérico contínuo</u>? *Próxima aula!*