

Estimação, Detecção e Aprendizagem II

Fundamentos Gerais

Catarina Oliveira



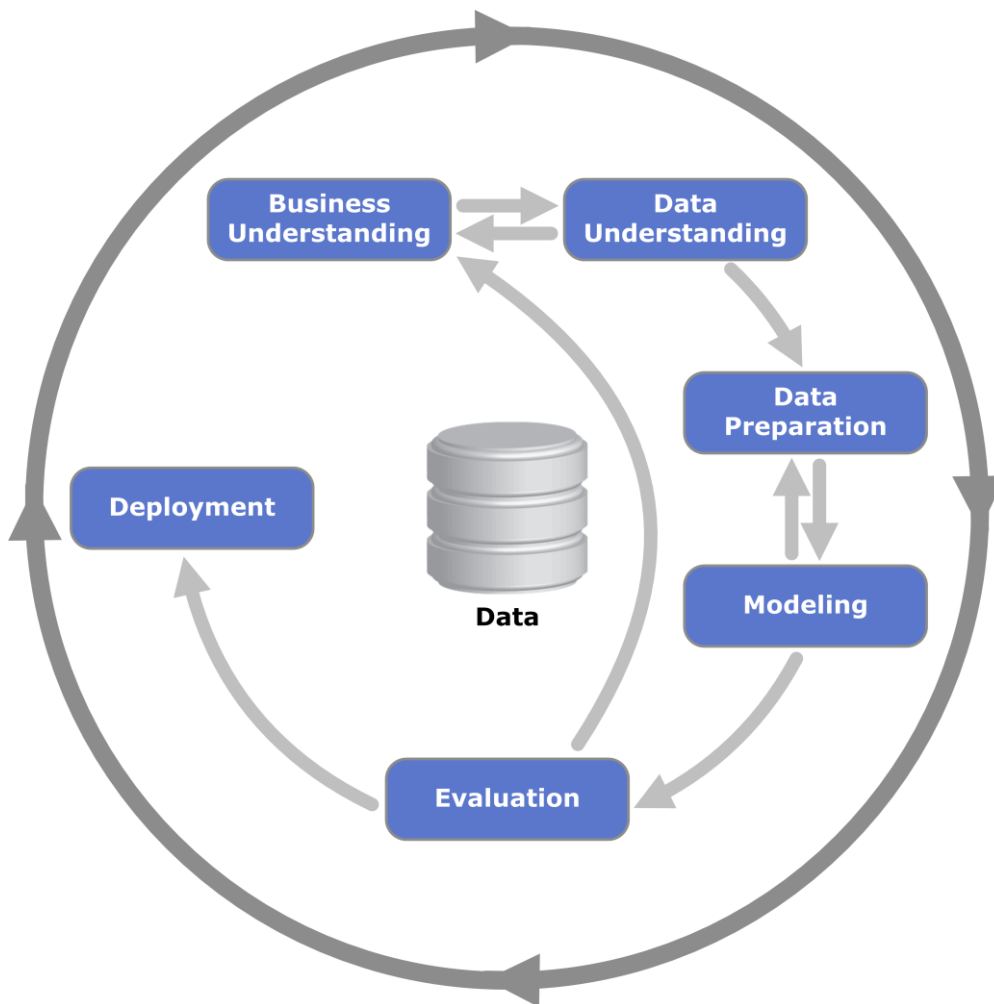
DEPARTAMENTO CIÊNCIA
E TECNOLOGIA



CONTEÚDO

1. Introdução
2. *Data mining* descritivo
 1. Regras de Associação
3. *Data mining* preditivo
 1. Classificação
 2. Regressão

Modelo CRISP-DM



Business Understanding:

- Compreensão dos objetivos e requisitos do projeto
- Conversão desse conhecimento em:
 - definição de problema de *data mining*
 - plano preliminar

Data Understanding:

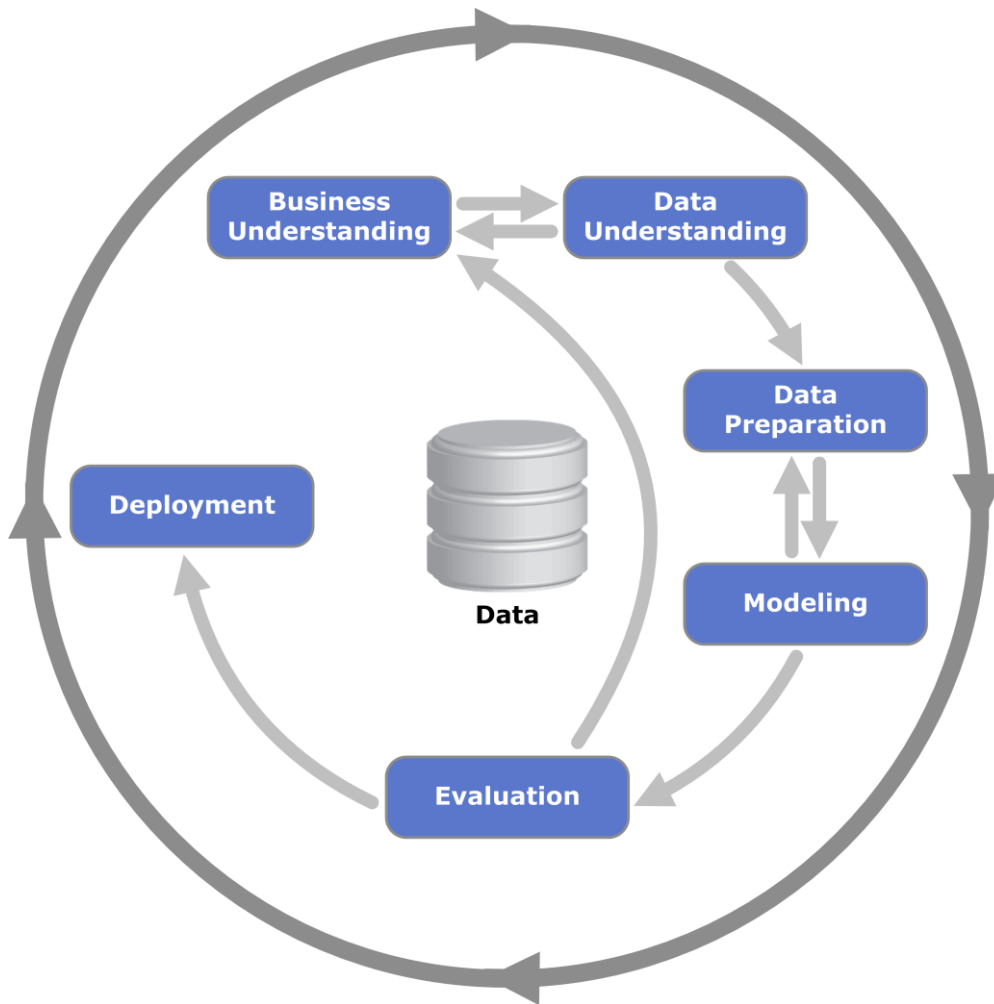
- Recolha de dados inicial
- Familiarização com os dados
 - identificar problemas de qualidade de dados
 - descobrir os primeiros insights sobre os dados
 - detetar subconjuntos interessantes

Data Preparation:

- Construir o conjunto de dados final a partir dos dados brutos iniciais.

<https://www.datasciencecentral.com/profiles/blogs/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome>

Modelo CRISP-DM



Modeling:

- Selecionar e aplicar técnicas de modelação
 - Algumas técnicas têm requisitos específicos quanto à forma dos dados
 - Possível necessidade de voltar à preparação

Evaluation:

- Teste dos modelos criados
 - Com base em medidas de performance para avaliar
 - Generalização
 - resposta ao objetivo
- Escolha do modelo “vencedor”

Deployment:

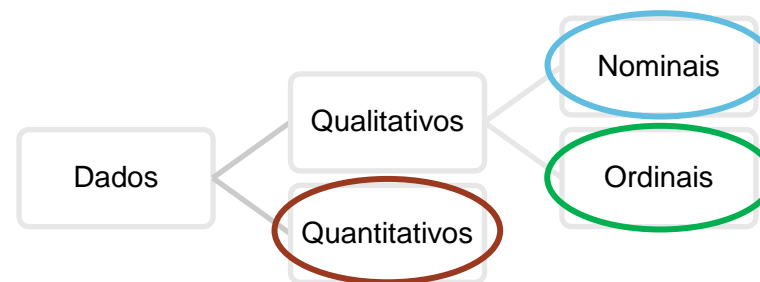
- Publicação do modelo escolhido de forma a poder aplicar o modelo aos novos dados

<https://www.datasciencecentral.com/profiles/blogs/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome>

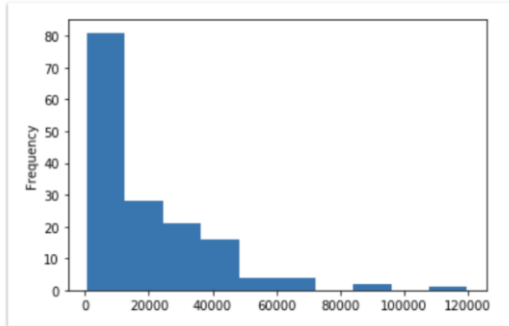
Tipos de Dados

CLIENTE	IDADE	ALTURA	PESO	GÊNERO	TIPO
ANDRÉ	25	77	175	M	Bom
BÁRBARA	31	110	195	M	Bom
CARLOS	15	70	172	F	Mau
DIANA	20	85	180	M	Bom
EDUARDO	10	65	168	F	Mau
FÁTIMA	12	75	173	M	Bom
GUILHERME	16	75	180	F	Mau
HELENA	26	63	165	F	Mau

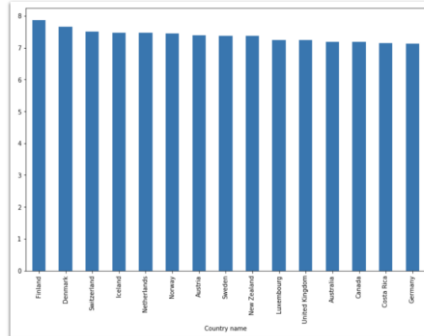
Variáveis independentes
 Variável dependente
 Variável objetivo
 Target



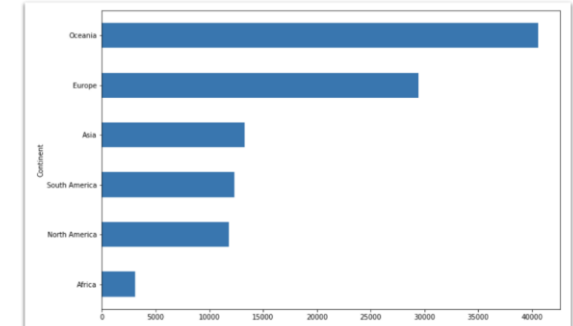
Visualização de dados



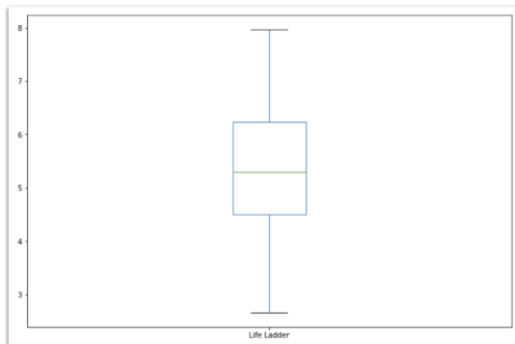
Histograma



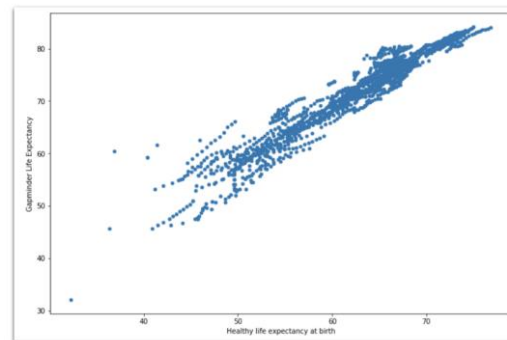
Barras verticais



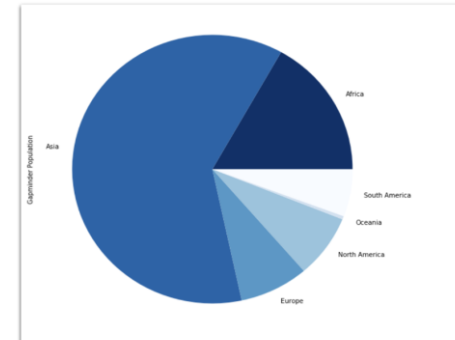
Barras horizontais



Box plot



Scatter plot



Pie chart

Plots em Python: <https://towardsdatascience.com/plotting-with-python-c2561b8c0f1f>

Exploração de dados em Python: <https://towardsdatascience.com/exploring-univariate-data-e7e2dc8fde80>

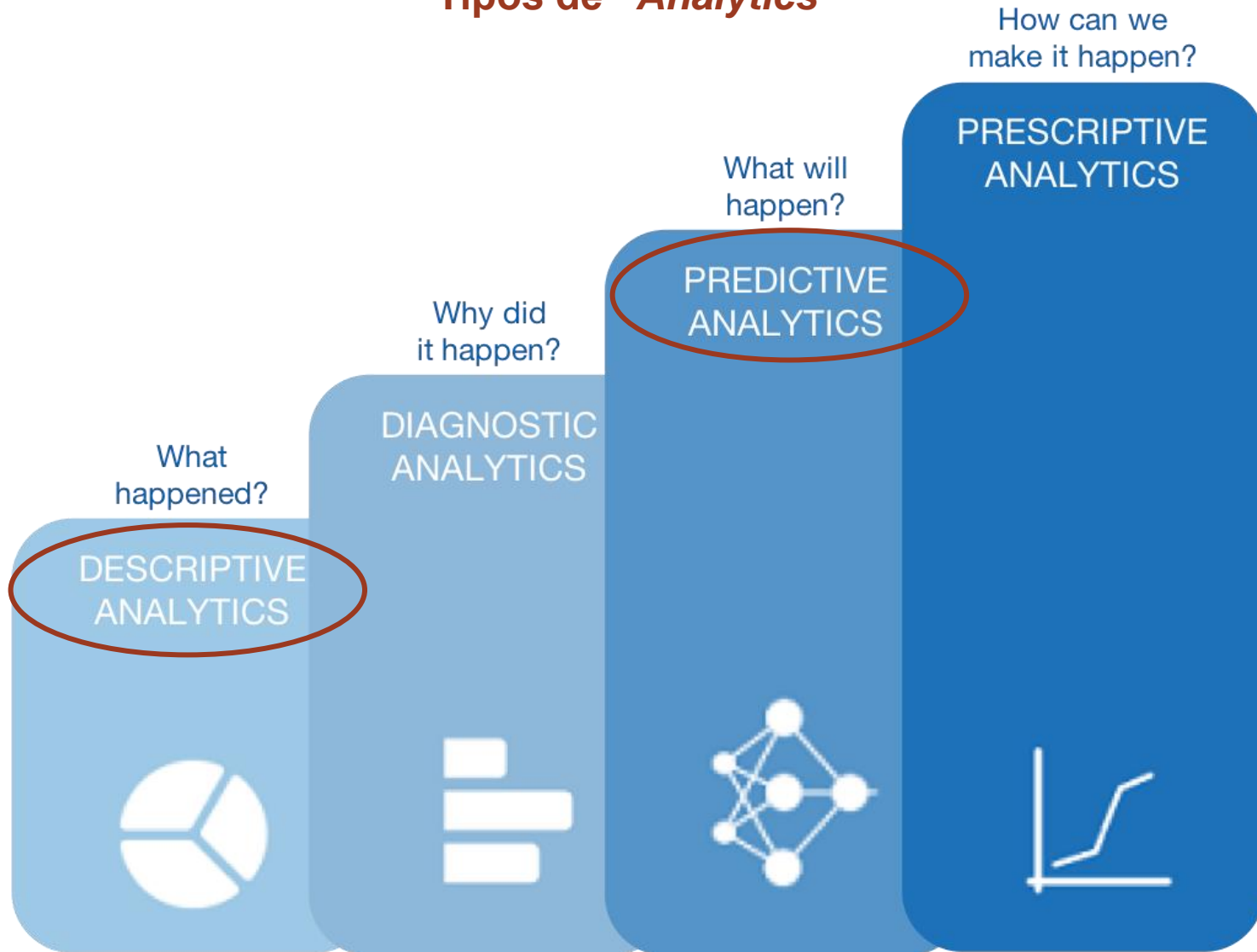
(Possíveis) Problemas com dados / Soluções

Problemas	Exemplos	Soluções:
Campos não preenchidos		<ul style="list-style-type: none"> • Representar outra informação (ex: ND, NA) • Preenchimento semiautomático (valor pré-definido, esperado [média, moda, mediana], previsto)
Erros/ruído	localidade="Prto" idade=-1	<ul style="list-style-type: none"> • Substituir valores • Representar outra informação (ex: ND, NA, Erro)
Erros sistemáticos	idade=99 (porque sistema obriga a inserir idade mesmo quando é desconhecida)	<ul style="list-style-type: none"> • Representar outra informação (ex: ND, NA, Erro) • Preenchimento semiautomático (valor pré-definido, "esperado" [média, moda, mediana], previsto)
Inconsistências	localidade="V.N. de Gaia" e "Gaia"	<ul style="list-style-type: none"> • Substituir valores
Outliers	despesas em automóveis com cartões bancários de valor 100.000€	<ul style="list-style-type: none"> • Substituir por valor extremo da distribuição

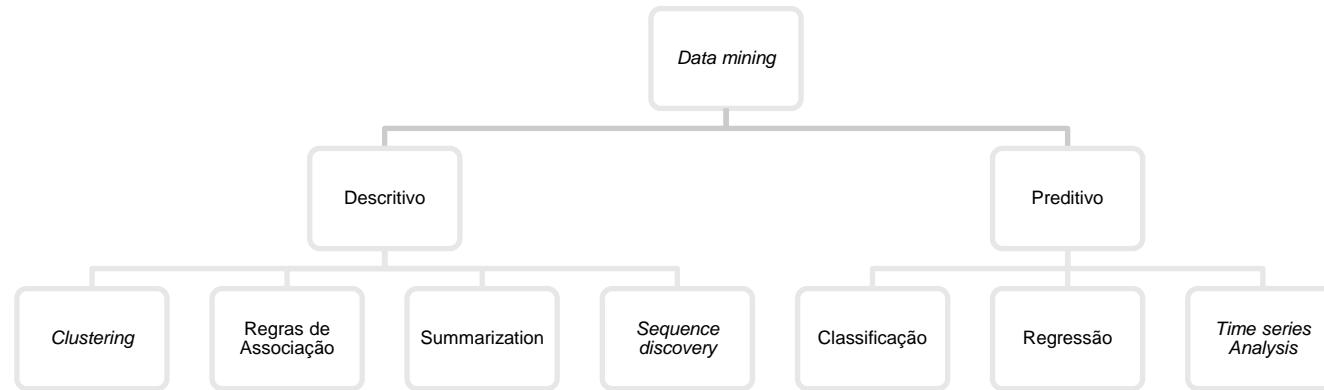
Solução limite: eliminar linhas/colunas

- impacto depende do número de linhas afetadas pelo problema

Tipos de “Analytics”

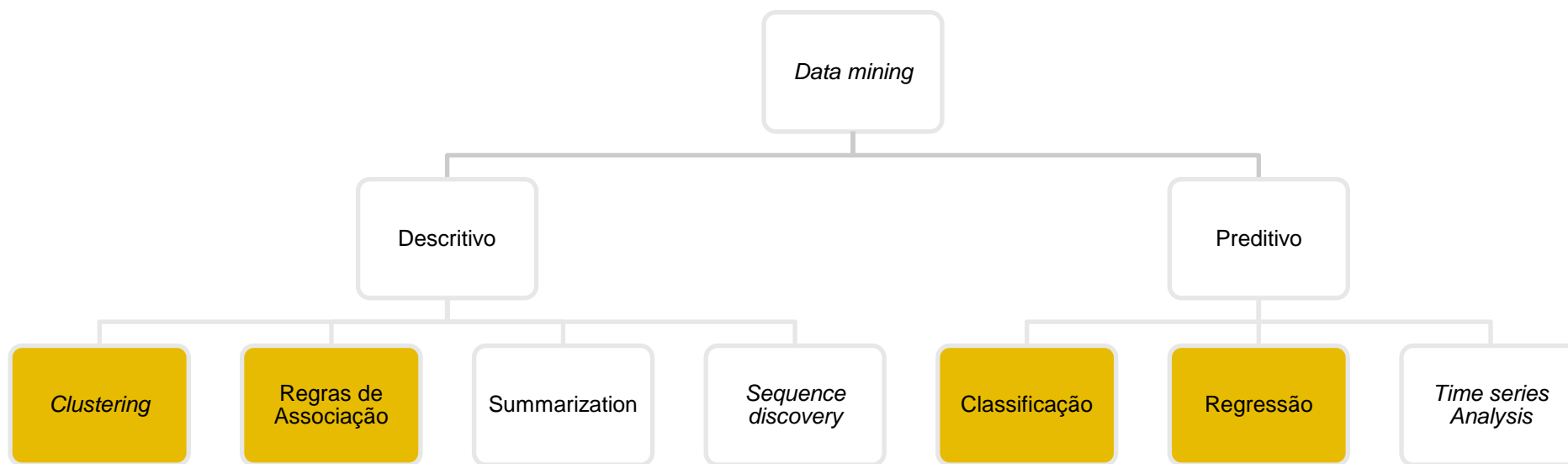


Data mining descritivo e preditivo

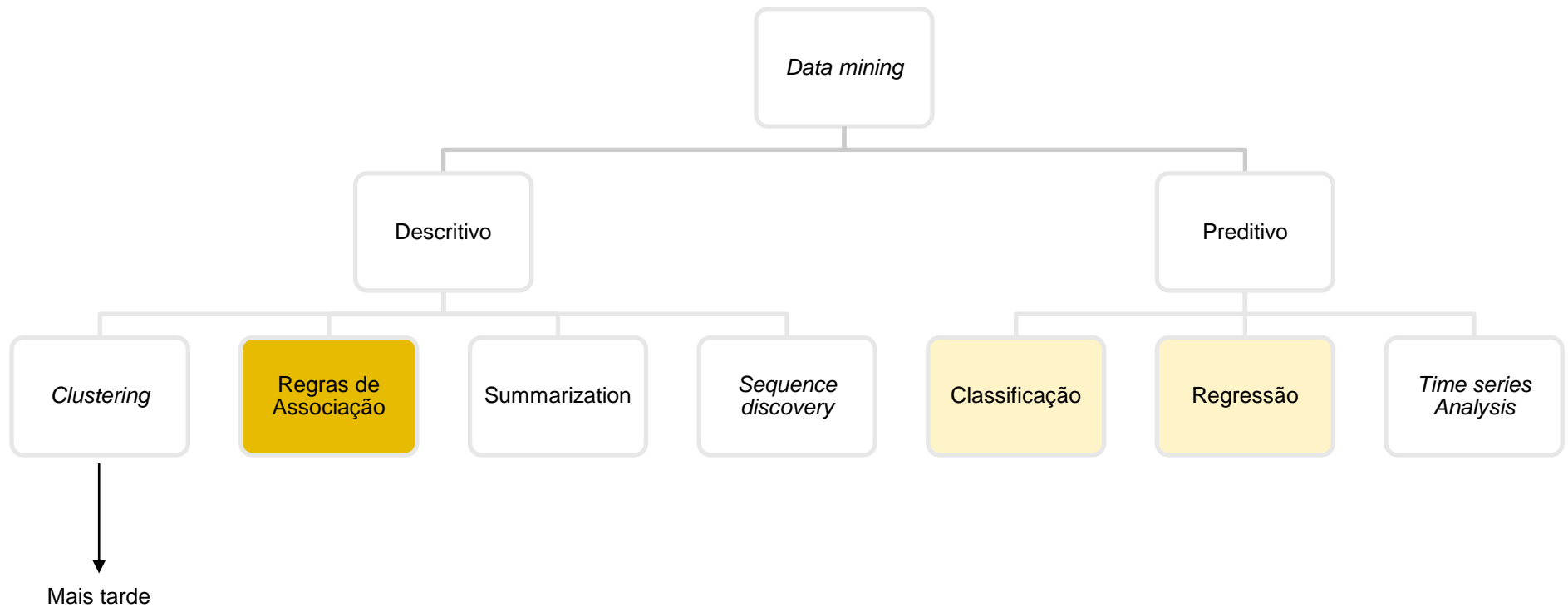


Determina	O que aconteceu no passado	O que poderá acontecer no futuro
Precisão	Resultados precisos	Não garantida
Métodos de análise prática	Relatórios padrão, consulta	Modelação preditiva, previsão, simulação e alertas.
Requisitos	Agregação de dados e data mining	Métodos estatísticos e de previsão
Abordagem	Reativa	Proativa
Descreve	Características dos dados	Indução sobre dados presentes e passados para efetuar previsões
Questões	<ul style="list-style-type: none"> • O que aconteceu? • Qual é o problema? • Com que frequência acontece o problema? 	<ul style="list-style-type: none"> • O que vai acontecer? • Qual é o resultado se a tendência se mantiver? • Que ações são necessárias?

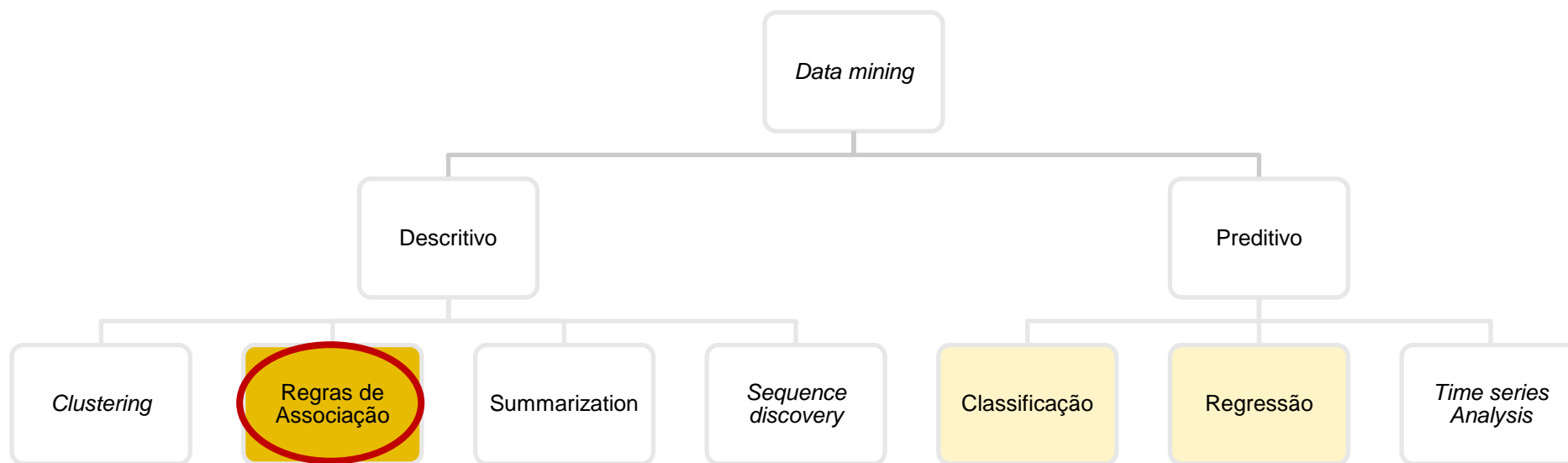
Estimação, Detecção e Aprendizagem II



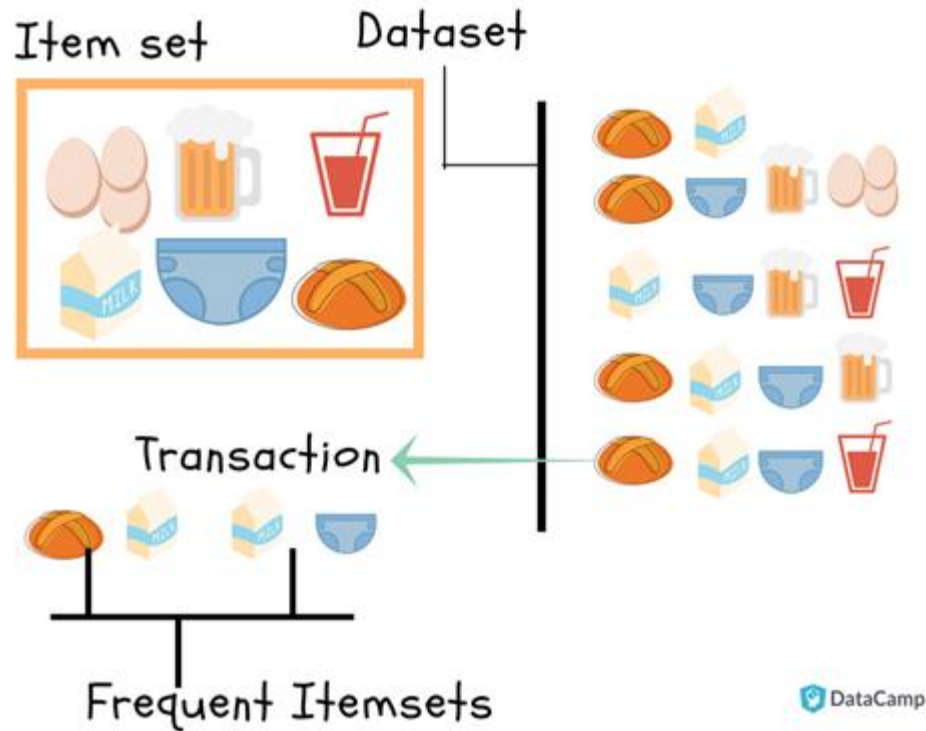
Este capítulo



Este capítulo



Regras de Associação



Dado um conjunto de transações (*Transactions*)

Identificar:

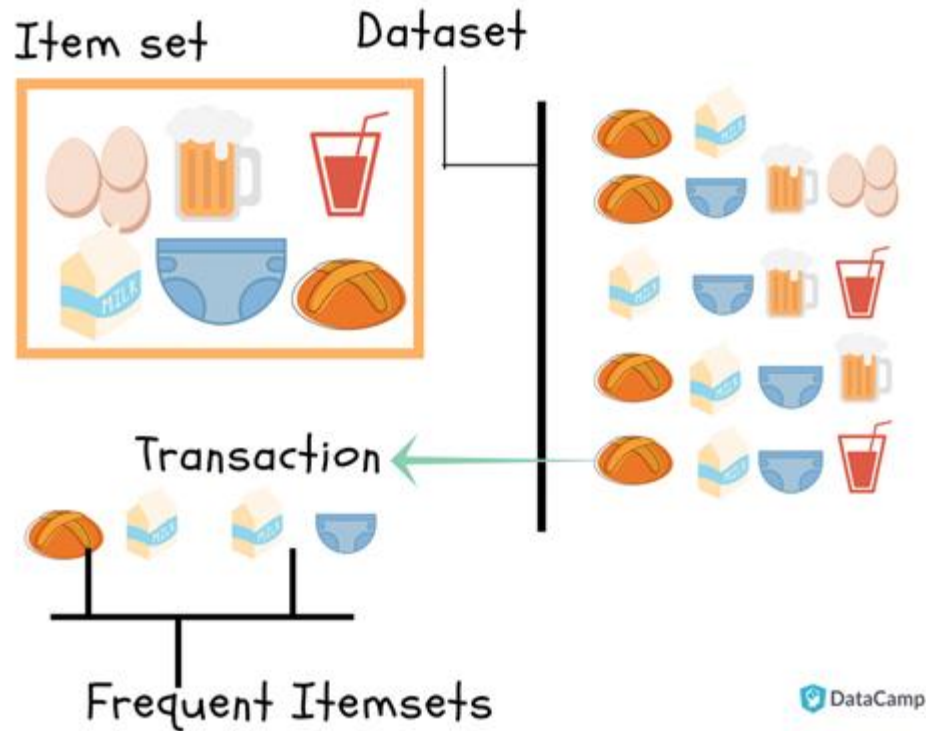
- coocorrências frequentes (*Frequent Itemsets*)
- *Itemsets* que originam outros *itemsets* (Regras de Associação)

Exemplos:

- Cestos de compras
- Transações de crédito
- Clickstreams
- Sistemas de recomendação



Itemset mining: Definição



Dado:

- Um conjunto de transações $D = \{t_1, t_2, \dots, t_n\}$

$$D = \{\{\text{pão, leite}\}, \\ \{\text{pão, fraldas, cerveja, ovos}\}, \\ \{\text{leite, fraldas, cerveja, refrigerantes}\}, \\ \{\text{pão, leite, fraldas, cerveja}\}, \\ \{\text{pão, leite, fraldas, refrigerantes}\}\}$$

- Um suporte mínimo $sup_{min} \in [0,1]$

Encontrar os itemsets $X: Support(X) > sup_{min}$,

Suporte: Frequência relativa (probabilidade) de X em D

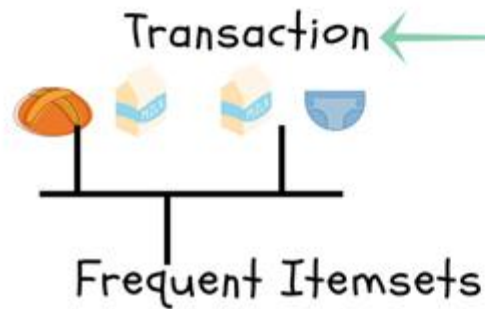
$$Support(X) = P(X)$$



Itemset mining: Exemplo

Produtos = {ovos, cerveja, refrigerante, leite, fraldas, pão}
 = {O, C, R, L, F, P}

Item set Dataset



Possíveis Itemsets:

1	2	3	4	5	6
{O}	{O, C}	{O, C, R}	{O, C, R, L}	{O, C, R, L, F}	{O, C, R, L, F, P}
{C}	{O, R}	{O, C, L}	{O, C, R, F}	{O, C, R, L, P}	
{R}	{O, L}	{O, C, F}	{O, C, R, P}	{O, C, R, F, P}	
{L}	{O, F}	{O, C, P}	{O, C, L, F}	{O, C, L, F, P}	
{F}	{O, P}	{O, R, L}	{O, C, L, P}	{O, R, L, F, P}	
{P}	{C, R}	{O, R, F}	{O, C, F, P}	{C, R, L, F, P}	
	{C, L}	{O, R, P}	{O, R, L, F}		
	{C, F}	{O, L, F}	{O, R, L, P}		
	{C, P}	{O, L, P}	{O, R, F, P}		
	{R, L}	{O, F, P}	{O, L, F, P}		
	{R, F}	{C, R, L}	{C, R, L, F}		
	{R, P}	{C, R, F}	{C, R, L, P}		
	{L, F}	{C, R, P}	{C, R, F, P}		
	{L, P}	{C, L, F}	{C, L, F, P}		
	{F, P}	{C, L, P}	{R, L, F, P}		
		{C, F, P}			
		{R, L, F}			
		{R, L, P}			
		{R, F, P}			
		{L, F, P}			

Itemset mining: Exemplo



Frequência dos Itemsets:

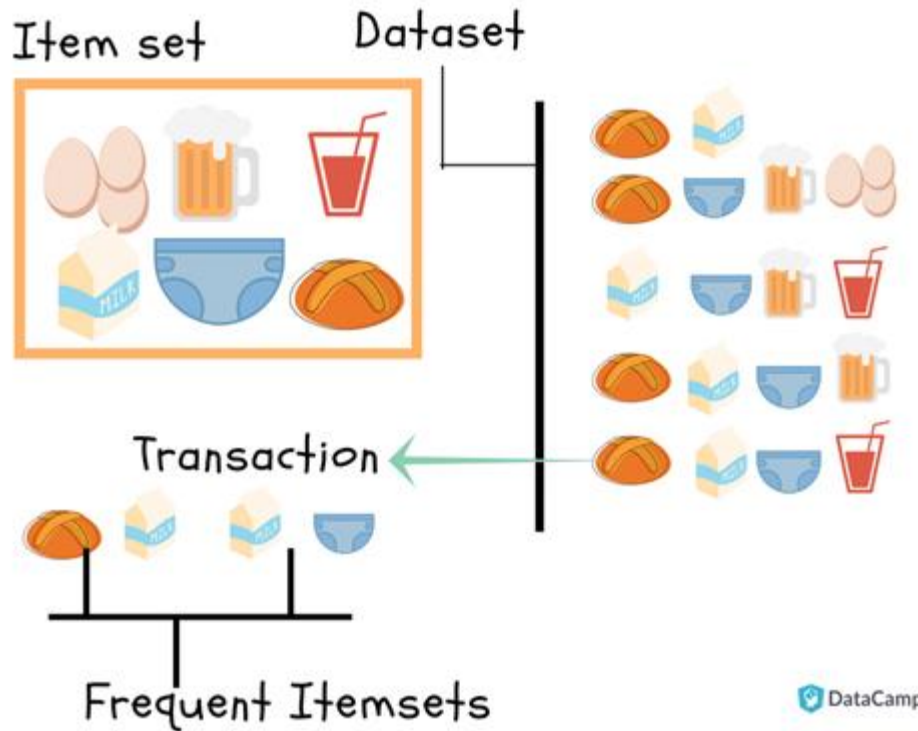
1	#	2	#	3	#	4	#	5	#	6	#
{O}	1	{O, C}	1	{O, C, R}	0	{O, C, R, L}	0	{O, C, R, L, F}	0	{O, C, R, L, F, P}	0
{C}	3	{O, R}	0	{O, C, L}	0	{O, C, R, F}	0	{O, C, R, L, P}	0		
{R}	2	{O, L}	0	{O, C, F}	1	{O, C, R, P}	0	{O, C, R, F, P}	0		
{L}	4	{O, F}	1	{O, C, P}	1	{O, C, L, F}	0	{O, C, L, F, P}	0		
{F}	4	{O, P}	1	{O, R, L}	0	{O, C, L, P}	0	{O, R, L, F, P}	0		
{P}	4	{C, R}	1	{O, R, F}	0	{O, C, F, P}	1	{C, R, L, F, P}	0		
		{C, L}	2	{O, R, P}	0	{O, R, L, F}	0				
		{C, F}	3	{O, L, F}	0	{O, R, L, P}	0				
		{C, P}	2	{O, L, P}	0	{O, R, F, P}	0				
		{R, L}	1	{O, F, P}	1	{O, L, F, P}	0				
		{R, F}	1	{C, R, L}	1	{C, R, L, F}	1				
		{R, P}	1	{C, R, F}	1	{C, R, L, P}	0				
		{L, F}	3	{C, R, P}	0	{C, R, F, P}	0				
		{L, P}	3	{C, L, F}	2	{C, L, F, P}	1				
		{F, P}	3	{C, L, P}	1	{R, L, F, P}	1				
				{C, F, P}	2						
				{R, L, F}	1						
				{R, L, P}	1						
				{R, F, P}	1						
				{L, F, P}	2						

Frequent Itemsets (Sup > 50%):

{C}
 {L}
 {F}
 {P}
 {C, F}
 {L, F}
 {L, P}
 {F, P}

Com $sup_{min} = 50\%$
 ($\# > 2,5$)

Regras de Associação: Conceitos



Formato: **Antecedente → Consequente**

“Quando o antecedente é observado, o consequente também deverá (provavelmente) ser observado”

Exemplo:

{A, B} → {C, D}

“Quando são observados os itens A e B, os itens C e D também deverão (provavelmente) ser observados”

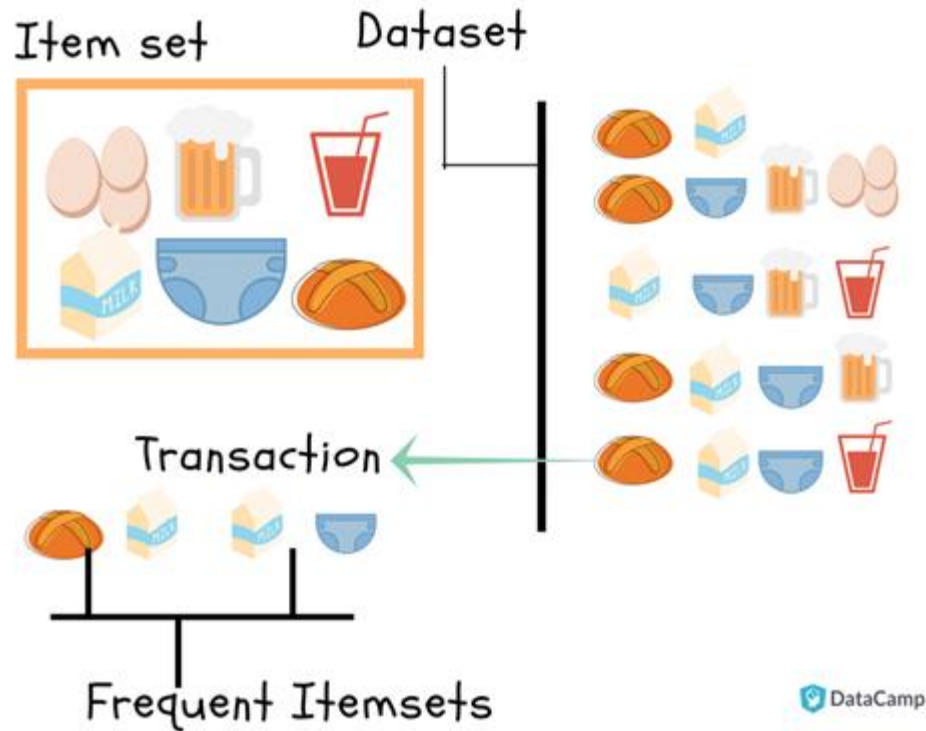
Suporte: percentagem de transações em que a coocorrência é observada

$$\text{Support}(\{A, B\} \rightarrow \{C, D\}) = P(\{A, B, C, D\})$$

Confiança: percentagem de transações em que a ocorrência do antecedente corretamente antevê a ocorrência do consequente

$$\text{Confidence}(\{A, B\} \rightarrow \{C, D\}) = P(\{C, D\}|\{A, B\}) = \frac{\text{freq}(\{A, B, C, D\})}{\text{freq}(\{A, B\})}$$

Mining Association Rules: Definição



Dado:

- Um conjunto de transações $D = \{t_1, t_2, \dots, t_n\}$
- Um suporte mínimo $sup_{min} \in [0,1]$
- Uma confiança mínima $conf_{min} \in [0,1]$

Encontrar todas as regras $A \rightarrow C$ tal que:

- $Support(A \rightarrow C) \geq sup_{min}$,
- $Confidence(A \rightarrow C) \geq conf_{min}$,

Mining Association Rules: Exemplo



1	#	2	#	3	#	4	#	5	#	6	#
{O}	1	{O, C}	1	{O, C, R}	0	{O, C, R, L}	0	{O, C, R, L, F}	0	{O, C, R, L, F, P}	0
{C}	3	{O, R}	0	{O, C, L}	0	{O, C, R, F}	0	{O, C, R, L, P}	0		
{R}	2	{O, L}	0	{O, C, F}	1	{O, C, R, P}	0	{O, C, R, F, P}	0		
{L}	4	{O, F}	1	{O, C, P}	1	{O, C, L, F}	0	{O, C, L, F, P}	0		
{F}	4	{O, P}	1	{O, R, L}	0	{O, C, L, P}	0	{O, R, L, F, P}	0		
{P}	4	{C, R}	1	{O, R, F}	0	{O, C, F, P}	1	{C, R, L, F, P}	0		
		{C, L}	2	{O, R, P}	0	{O, R, L, F}	0				
		{C, F}	3	{O, L, F}	0	{O, R, L, P}	0				
		{C, P}	2	{O, L, P}	0	{O, R, F, P}	0				
		{R, L}	1	{O, F, P}	1	{O, L, F, P}	0				
		{R, F}	1	{C, R, L}	1	{C, R, L, F}	1				
		{R, P}	1	{C, R, F}	1	{C, R, L, P}	0				
		{L, F}	3	{C, R, P}	0	{C, R, F, P}	0				
		{L, P}	3	{C, L, F}	2	{C, L, F, P}	1				
		{F, P}	3	{C, L, P}	1	{R, L, F, P}	1				
				{C, F, P}	2						
				{R, L, F}	1						
				{R, L, P}	1						
				{R, F, P}	1						
				{L, F, P}	2						

Frequent Itemsets (Sup > 50%):

- {C}
- {L}
- {F}
- {P}
- {C, F}
- {L, F}
- {L, P}
- {F, P}

Regra	Ant	Cons	freg(Ant)	Sup(Ant)	freg(Cons)	Sup(Cons)	freg(Ant→Cons)	Sup(Ant→Cons)	Conf(Ant→Cons)
C→F	C	F	3	60%	4	80%	3	60%	100%
F→C	F	C	4	80%	3	60%	3	60%	75%
L→F	L	F	4	80%	4	80%	3	60%	75%
F→L	F	L	4	80%	4	80%	3	60%	75%
L→P	L	P	4	80%	4	80%	3	60%	75%
P→L	P	L	4	80%	4	80%	3	60%	75%
F→P	F	P	4	80%	4	80%	3	60%	75%
P→F	P	F	4	80%	4	80%	3	60%	75%

Com $sup_{min} = 50\%$

e $conf_{min} = 90\%$

Regras de Associação: Exercício

Em Python:

```
records = [['p', 'l'],
            ['p', 'f', 'c', 'o'],
            ['l', 'f', 'c', 'r'],
            ['p', 'l', 'f', 'c'],
            ['p', 'l', 'f', 'r']]
```

```
from apyori import apriori
```

```
rules = apriori(records, min_support=0.5, min_confidence=0.9)
litrules = list(rules)
for item in litrules:
    pair = item[0]
    items = [x for x in pair]
    ant = str(list(item[2][0][0]))[1:-1]
    cons = str(list(item[2][0][1]))[1:-1]
    print("Rule: {" + ant + "} -> {" + cons + "}")
    print("Support: " + str(item[1]))
    print("Confidence: " + str(item[2][0][2]))
    print("Lift: " + str(item[2][0][3]))
    print("=====")
```

Com $sup_{min} = 50\%$
e $conf_{min} = 90\%$

Regra	Ant	Cons	freq(Ant)	Sup(Ant)	freq(Cons)	Sup(Cons)	freq(Ant--Cons)	Sup(Ant--Cons)	Conf(Ant--Cons)
C→F	C	F	3	60%	4	80%	3	60%	100%
F→C	F	C	4	80%	3	60%	3	60%	75%
L→F	L	F	4	80%	4	80%	3	60%	75%
F→L	F	L	4	80%	4	80%	3	60%	75%
L→P	L	P	4	80%	4	80%	3	60%	75%
P→L	P	L	4	80%	4	80%	3	60%	75%
F→P	F	P	4	80%	4	80%	3	60%	75%
P→F	P	F	4	80%	4	80%	3	60%	75%

Rule: {'c'} -> {'f'}
Support: 0.6
Confidence: 1.0
Lift: 1.25

=====

Avaliação de Regras de Associação: Interesse

Para uma regra de associação ser interessante, tem de ser:

- Inesperada (desviar-se do esperado)
- Útil (com benefício espectável)

Exemplo: numa bomba de gasolina, {jornal} \rightarrow {combustível} não inesperada nem útil

Geralmente, uma regra $A \rightarrow C$ é interessante se A e C não são estatisticamente independentes

- A e C são estatisticamente independentes se:
 - $Support(A \cup C) \approx Support(C) \times Support(C)$
 - $Confidence(A \rightarrow C) \approx Confidence(\emptyset \rightarrow C)$

Avaliação de Regras de Associação

Lift: mede a importância de uma regra:

- Lift > 1: o antecedente e o consequente aparecem mais frequentemente juntos do que o esperado
 - a ocorrência do antecedente tem um efeito positivo na ocorrência do consequente.
- Lift < 1: o antecedente e o consequente aparecem com menos frequência juntos do que o esperado
 - a ocorrência do antecedente tem um efeito negativo na ocorrência do consequente.
- Lift ≈ 1 o antecedente e o consequente aparecem quase tão frequentemente juntos quanto o esperado
 - a ocorrência do antecedente quase não tem efeito na ocorrência do consequente.

$$Lift(A \rightarrow C) = \frac{Confidence(A \rightarrow C)}{Support(C)}$$

Conviction: mede a “implicação” (frequência em que o antecedente ocorre sem o consequente)

- Conviction ≈ 1 se o antecedente e o consequente não são relacionados

$$Conviction(A \rightarrow C) = \frac{1 - Support(C)}{1 - Confidence(A \rightarrow C)}$$

Leverage: mede a proporção de elementos adicionais cobertos pela regra (em relação ao esperado se fossem independentes)

- Leverage ≈ 0 se são independentes

$$Leverage(A \rightarrow C) = Support(A \rightarrow C) - Support(A) \times Support(C)$$

Avaliação de Regras de Associação

Table 5: Interestingness Measures for Association Patterns.

#	Measure	Formula
1	ϕ -coefficient	$\frac{P(A,B) - P(A)P(B)}{\sqrt{P(A)P(B)(1-P(A))(1-P(B))}}$
2	Goodman-Kruskal's (λ)	$\frac{\sum_j \max_k P(A_j, B_k) + \sum_k \max_j P(A_j, B_k) - \max_j P(A_j) - \max_k P(B_k)}{2 - \max_j P(A_j) - \max_k P(B_k)}$
3	Odds ratio (α)	$\frac{P(A,B)P(\bar{A},\bar{B})}{P(A,\bar{B})P(\bar{A},B)}$
4	Yule's Q	$\frac{P(A,B)P(\bar{A}\bar{B}) - P(A,\bar{B})P(\bar{A},B)}{P(A,B)P(\bar{A}\bar{B}) + P(A,\bar{B})P(\bar{A},B)} = \frac{\alpha - 1}{\alpha + 1}$
5	Yule's Y	$\frac{\sqrt{P(A,B)P(\bar{A}\bar{B})} - \sqrt{P(A,\bar{B})P(\bar{A},B)}}{\sqrt{P(A,B)P(\bar{A}\bar{B})} + \sqrt{P(A,\bar{B})P(\bar{A},B)}} = \frac{\sqrt{\alpha} - 1}{\sqrt{\alpha} + 1}$
6	Kappa (κ)	$\frac{P(A,B) + P(\bar{A},\bar{B}) - P(A)P(B) - P(\bar{A})P(\bar{B})}{1 - P(A)P(B) - P(\bar{A})P(\bar{B})}$
7	Mutual Information (M)	$\frac{\sum_i \sum_j P(A_i, B_j) \log \frac{P(A_i, B_j)}{P(A_i)P(B_j)}}{\min(-\sum_i P(A_i) \log P(A_i), -\sum_j P(B_j) \log P(B_j))}$
8	J-Measure (J)	$\max(P(A, B) \log(\frac{P(B A)}{P(B)}) + P(\bar{A}\bar{B}) \log(\frac{P(\bar{B} \bar{A})}{P(\bar{B})}),$ $P(A, B) \log(\frac{P(A B)}{P(A)}) + P(\bar{A}\bar{B}) \log(\frac{P(\bar{A} \bar{B})}{P(\bar{A})}))$
9	Gini index (G)	$\max(P(A)[P(B A)^2 + P(\bar{B} \bar{A})^2] + P(\bar{A})[P(B \bar{A})^2 + P(\bar{B} \bar{A})^2]$ $- P(B)^2 - P(\bar{B})^2,$ $P(B)[P(A B)^2 + P(\bar{A} \bar{B})^2] + P(\bar{B})[P(A \bar{B})^2 + P(\bar{A} \bar{B})^2]$ $- P(A)^2 - P(\bar{A})^2)$
10	Support (s)	$P(A, B)$
11	Confidence (c)	$\max(P(B A), P(A B))$
12	Laplace (L)	$\max(\frac{NP(A,B)+1}{NP(A)+2}, \frac{NP(A,B)+1}{NP(B)+2})$
13	Conviction (V)	$\max(\frac{P(A)P(\bar{B})}{P(A\bar{B})}, \frac{P(B)P(\bar{A})}{P(\bar{B}\bar{A})})$
14	Interest (I)	$\frac{P(A,B)}{P(A)P(B)}$
15	cosine (IS)	$\frac{P(A,B)}{\sqrt{P(A)P(B)}}$
16	Piatetsky-Shapiro's (PS)	$P(A, B) - P(A)P(B)$
17	Certainty factor (F)	$\max(\frac{P(B A) - P(B)}{1 - P(B)}, \frac{P(A B) - P(A)}{1 - P(A)})$
18	Added Value (AV)	$\max(P(B A) - P(B), P(A B) - P(A))$
19	Collective strength (S)	$\frac{P(A,B) + P(\bar{A}\bar{B})}{P(A)P(B) + P(\bar{A})P(\bar{B})} \times \frac{1 - P(A)P(B) - P(\bar{A})P(\bar{B})}{1 - P(A,B) - P(\bar{A}\bar{B})}$
20	Jaccard (ζ)	$\frac{P(A,B)}{P(A) + P(B) - P(A,B)}$
21	Klogsen (K)	$\sqrt{P(A, B) \max(P(B A) - P(B), P(A B) - P(A))}$

Tan, P. N., Kumar, V., & Srivastava, J. (2002, July). Selecting the right interestingness measure for association patterns. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 32-41). https://www.researchgate.net/publication/2829316_Selecting_the_Right_Interestingness_Measure_for_Association_Patterns

Algoritmo APRIORI

Apriori(T, ϵ)

$L_1 \leftarrow \{\text{large 1 - itemsets}\}$

$k \leftarrow 2$

while $L_{k-1} \neq \emptyset$

$C_k \leftarrow \{c = a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a, \{s \subseteq c \mid |s| = k-1\} \subseteq L_{k-1}\}$

for transactions $t \in T$

$D_t \leftarrow \{c \in C_k \mid c \subseteq t\}$

for candidates $c \in D_t$

$\text{count}[c] \leftarrow \text{count}[c] + 1$

$L_k \leftarrow \{c \in C_k \mid \text{count}[c] \geq \epsilon\}$

$k \leftarrow k + 1$

return $\bigcup_k L_k$

T: Conjunto de transações

ϵ : Confiança mínima

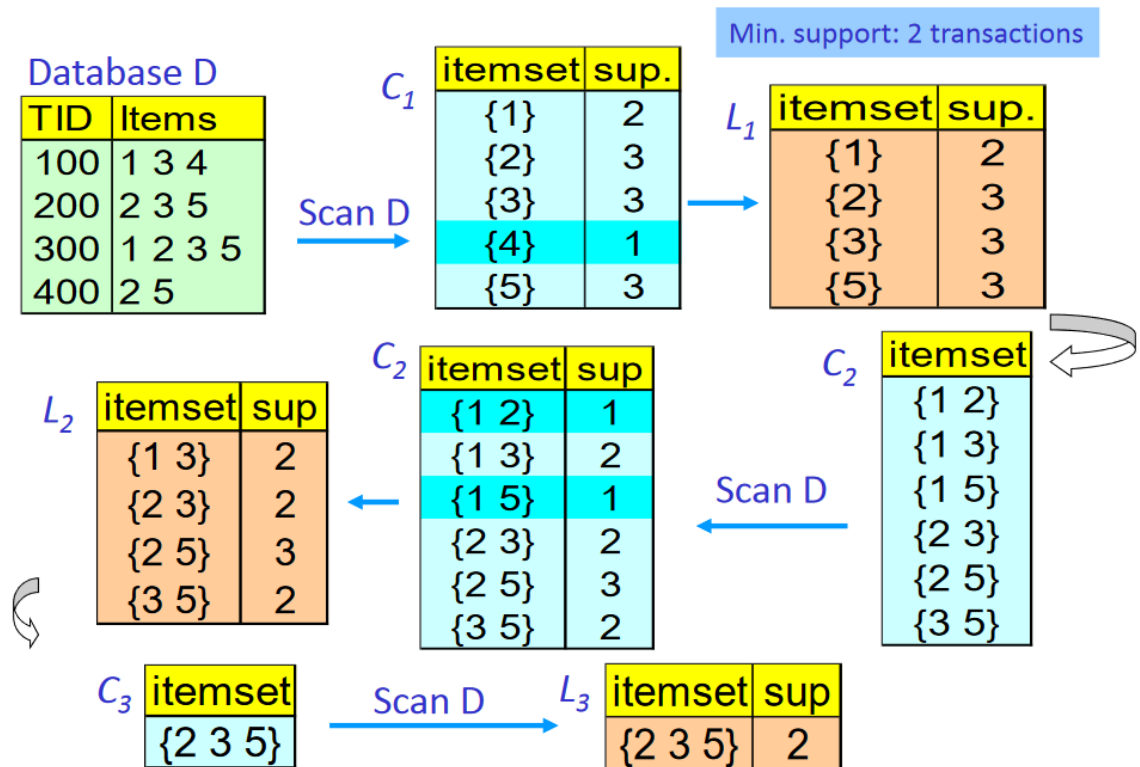
L: Lista de itemsets

k: Tamanho dos itemsets

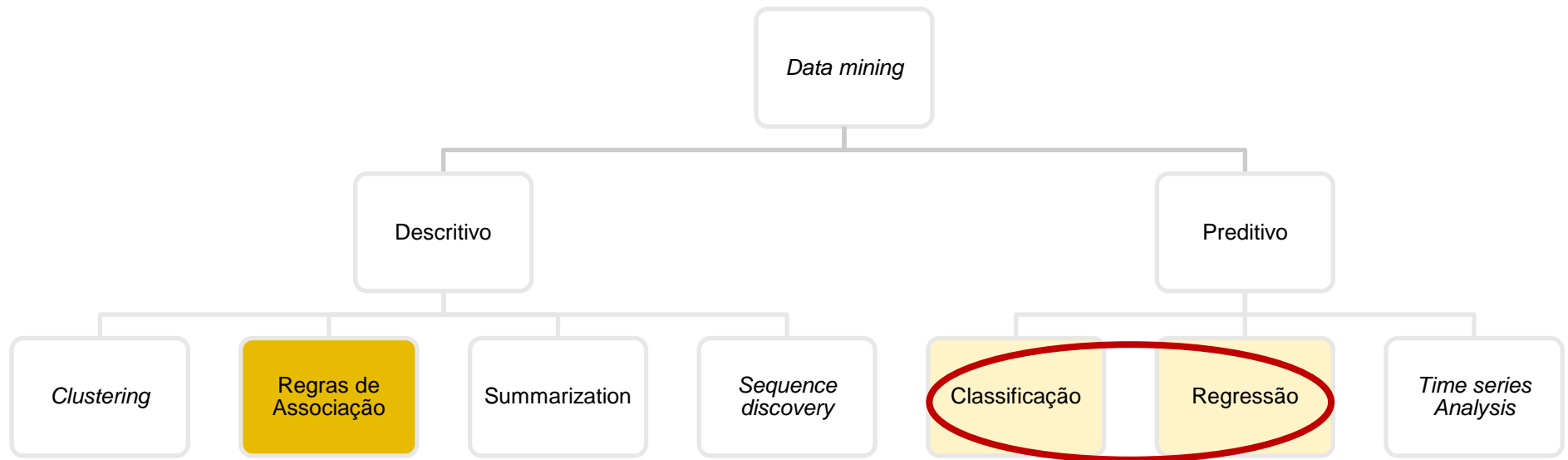
C: Itemsets candidatos

a, b: conjuntos de itens

D: itemsets contidos em T



Este capítulo

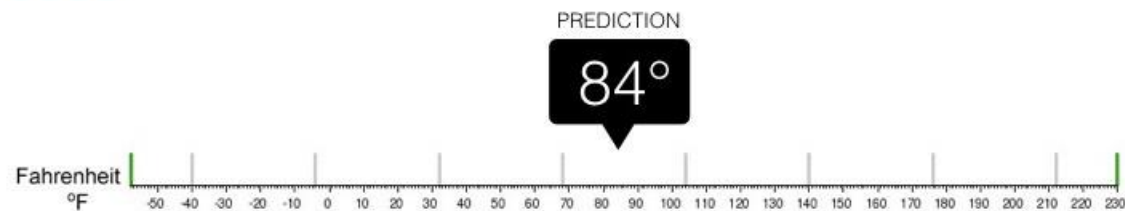


Classificação ou regressão?



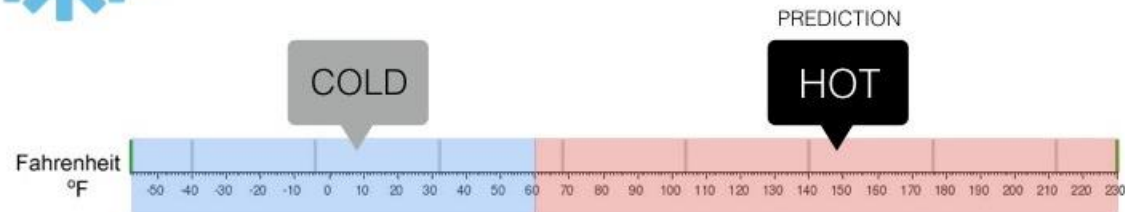
Regression

What is the temperature going to be tomorrow?

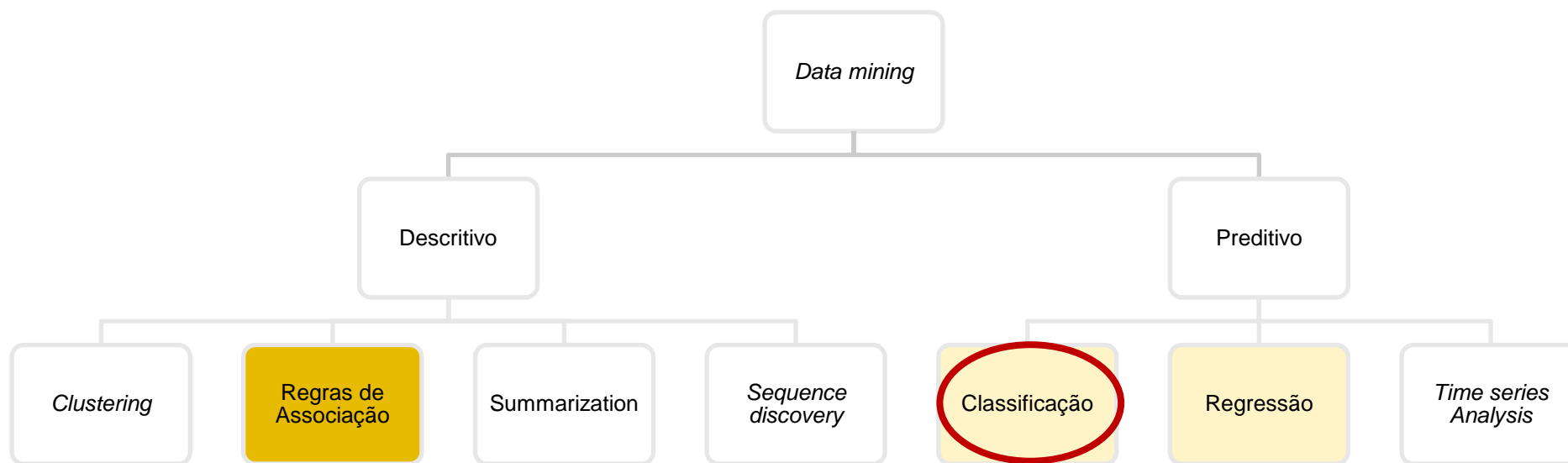


Classification

Will it be Cold or Hot tomorrow?



Este capítulo



Classificação

Dados os resultados da
última campanha, que
conhecemos (dados
históricos)

Variável dependente
Variável objetivo

Queremos prever os
potenciais aderentes

Comprou	Idade	Rendimento	Ag.fam	Vendas anteriores	Ultima Venda
nao	37	49000	2	1	42000
sim	43	68000	3	0	0
sim	42	61000	4	0	0
sim	26	52000	2	0	0
sim	40	64000	1	1	21000
sim	38	52000	1	0	0
sim	45	43000	4	1	47000
sim	35	45000	2	1	34000
nao	39	43000	2	0	0
sim	31	55000	3	1	46000
sim	34	57000	3	1	52000
nao	38	44000	1	0	0

Comprou	Idade	Rendimento	Ag.fam	Vendas anteriores	Ultima Venda
?	41	50000	2	1	0
?	39	68000	2	0	30000
?	58	61000	4	0	0
?	26	25000	3	0	0
?	21	50000	1	1	20000
?	38	43000	2	0	0
?	44	43000	4	1	47000
?	27	47000	2	1	21000
?	70	23000	2	0	25000

Variáveis independentes

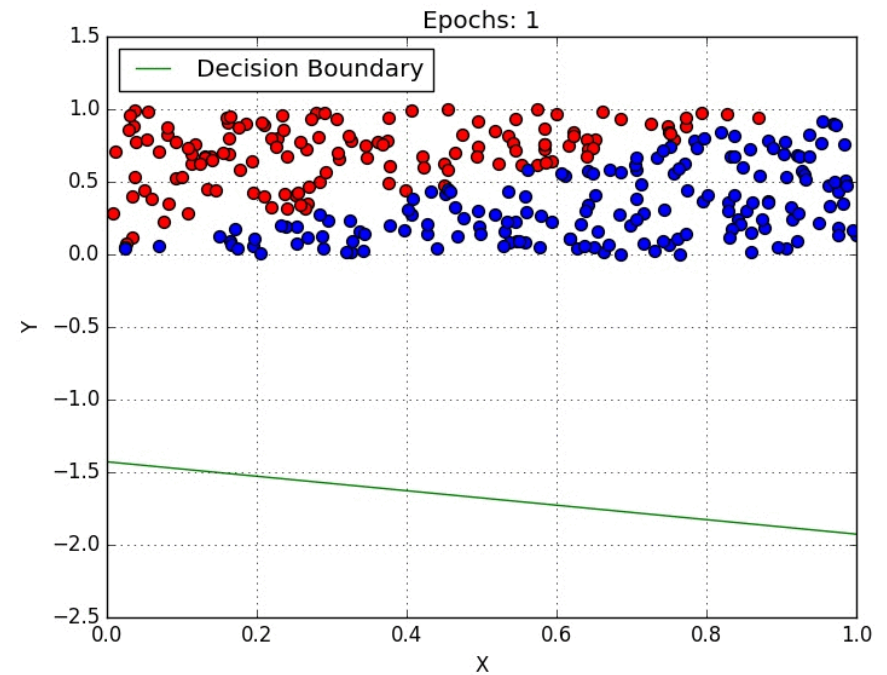
Algoritmos de classificação

- Logistic Regression
- Naive Bayes Classifier
- K-Nearest Neighbors
- Decision Tree
 - Random Forest
- Support Vector Machines
- ...

Logistic regression classifier

Calcula: $P(Y = 1|X)$ ou $P(Y = 0|X)$

A probabilidade de a variável dependente (Y) ter um determinado valor, dados os valores das variáveis independentes (X)



<https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148>

Naive Bayes classifier

Calcula: $P(Y|X) = \frac{P(X|Y) \times P(Y)}{P(X)}$

A probabilidade de a variável dependente (Y) ter um determinado valor, dados os valores das variáveis independentes (X)

Naive Bayes

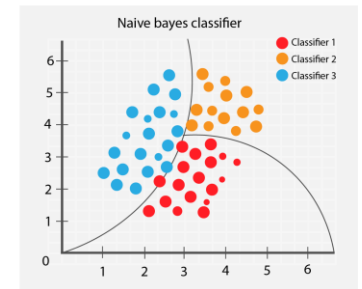
@thatware.co

In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

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

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

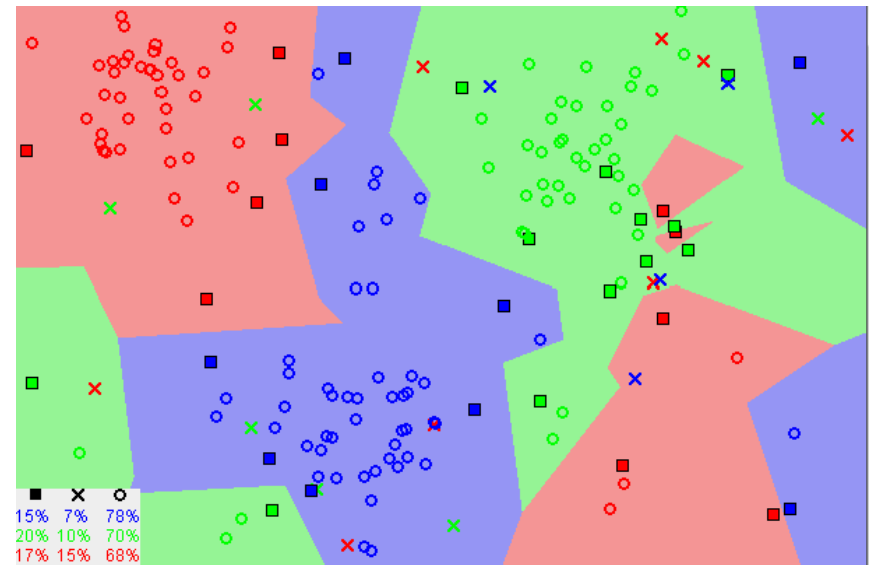


<https://towardsdatascience.com/introduction-to-na%C3%A5ve-bayes-classifier-fa59e3e24aaf>

K nearest neighbors classifier

Calcula a distância entre os elementos.

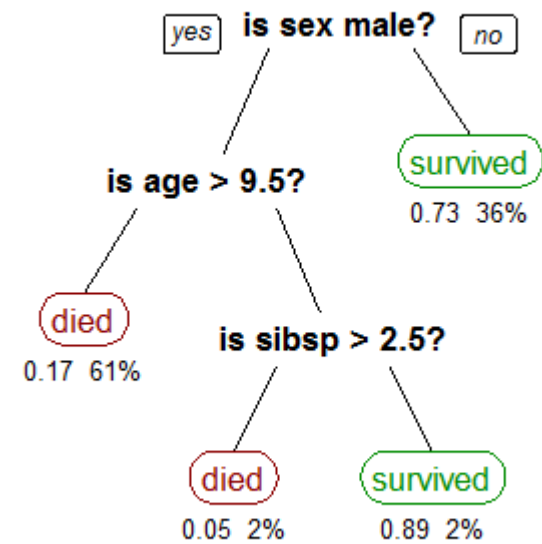
Assume que os elementos semelhantes se encontram próximos uns dos outros.



<https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>

Decision tree classifier

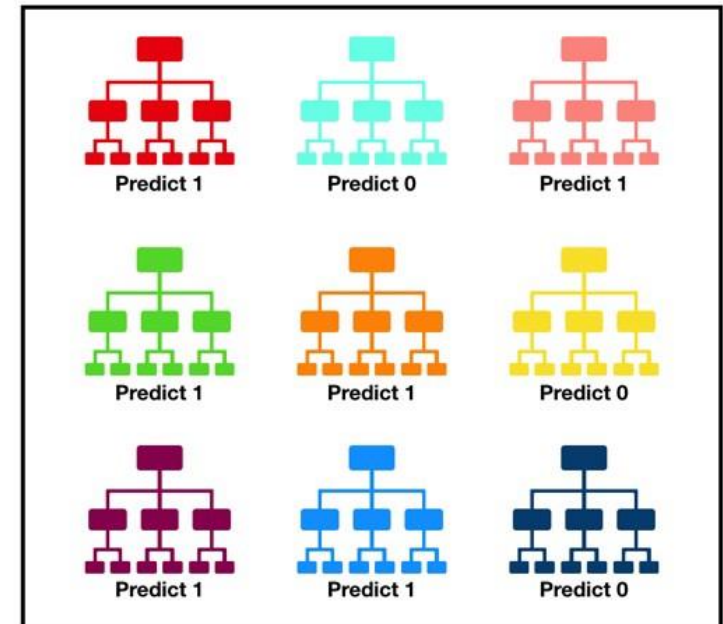
Constrói uma representação de uma tabela de decisão sob a forma de árvore



<https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>

Random forest classifier

Constrói um conjunto de árvores de decisão (decision trees)



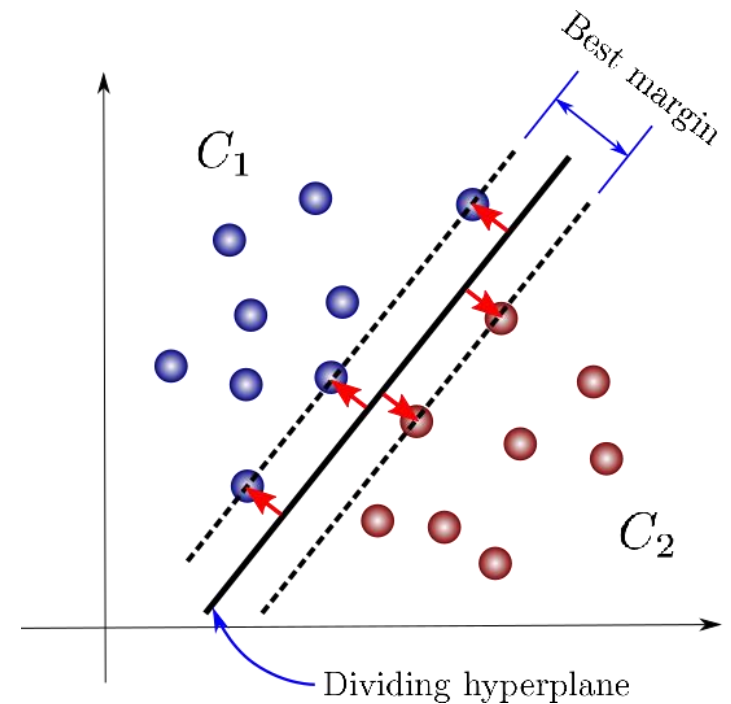
Tally: Six 1s and Three 0s

Prediction: 1

<https://towardsdatascience.com/understanding-random-forest-58381e0602d2>

Support Vector Machine classifier

Constrói uma representação dos exemplos como pontos no espaço, mapeados de maneira que os exemplos de cada categoria estejam tão longe quanto possível

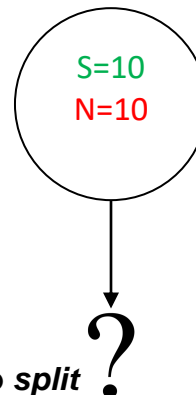


<https://towardsdatascience.com/support-vector-machines-for-classification-fc7c1565e3>

Algoritmo de construção de árvores de decisão

x1	x2	x3	classe
b	q	1	S
a	q	3	N
b	q	5	S
b	q	8	S
a	p	10	S
a	p	13	S
a	p	16	S
b	q	18	N
b	q	19	N
a	p	20	N
a	q	22	N
b	p	24	S
a	q	26	N
b	q	26	N
b	q	26	N
b	p	28	S
b	q	30	N
b	q	32	N
b	p	34	S
b	p	38	S

Raiz da árvore:



Por onde vamos fazer o *split* ?

É escolhido o *split* com menor custo (algoritmo *greedy*)

Recursivamente (grupos formados podem voltar a subdividir-se)

Medidas de avaliação de árvores de decisão

Gini Index: mede o grau ou probabilidade de uma determinada variável ser classificada incorretamente quando é escolhida aleatoriamente

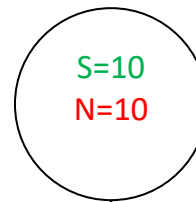
$$G = 1 - \sum_{i=1}^n (p_i)^2$$

p_i é a probabilidade de uma observação ser classificada numa classe em particular

Algoritmo de construção de árvores de decisão (Gini Index)

x1	x2	x3	classe
b	q	1	S
a	q	3	N
b	q	5	S
b	q	8	S
a	p	10	S
a	p	13	S
a	p	16	S
b	q	18	N
b	q	19	N
a	p	20	N
a	q	22	N
b	p	24	S
a	q	26	N
b	q	26	N
b	q	26	N
b	p	28	S
b	q	30	N
b	q	32	N
b	p	34	S
b	p	38	S

Raiz da árvore:



?

Por onde vamos fazer o *split* ?

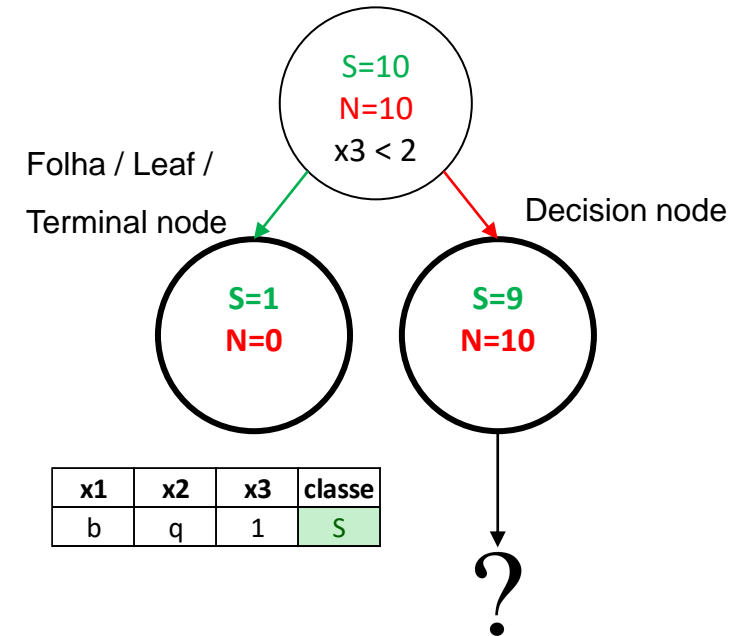
$$G = 1 - \sum_{i=1}^n (p_i)^2 = 1 - \left(\left(\frac{10}{20} \right)^2 + \left(\frac{10}{20} \right)^2 \right) = 0,5$$

Algoritmo de construção de árvores de decisão (Gini Index)

x1	x2	x3	classe
a	q	3	N
a	p	20	N
a	p	10	S
a	p	13	S
a	p	16	S
a	q	22	N
a	q	26	N
b	q	1	S
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	p	24	S
b	q	8	S
b	p	34	S
b	q	5	S
b	p	28	S
b	q	30	N
b	q	26	N
b	p	38	S

x1	x2	x3	classe
a	p	20	N
b	p	24	S
a	p	10	S
a	p	13	S
a	p	16	S
b	p	34	S
b	p	28	S
b	p	38	S
b	q	1	S
a	q	3	N
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	q	8	S
b	q	5	S
b	q	30	N
b	q	26	N
a	q	22	N
a	q	26	N

x1	x2	x3	classe
b	q	1	S
a	q	3	N
b	q	5	S
b	q	8	S
a	p	10	S
a	p	13	S
a	p	16	S
b	q	18	N
b	q	19	N
a	p	20	N
a	q	22	N
b	p	24	S
b	q	26	N
b	q	26	N
a	q	26	N
b	p	28	S
b	q	30	N
b	q	32	N
b	p	34	S
b	p	38	S



x1 = a	V	S	$p=3/7$	$G = 0,49$
	$p=7/20$	N	$p=4/7$	
	F	S	$p=7/13$	$G = 0,497$
	$p=13/20$	N	$p=6/13$	

x2 = p	V	S	$p=7/8$	$G = 0,219$
	$p=8/20$	N	$p=1/8$	
	F	S	$p=3/12$	$G = 0,375$
	$p=12/20$	N	$p=9/12$	

x3 < 2	V	S	$p=1/1$	$G = 0$
	$p=1/20$	N	$p=0/1$	
	F	S	$p=9/19$	$G = 0,499$
	$p=19/20$	N	$p=10/19$	

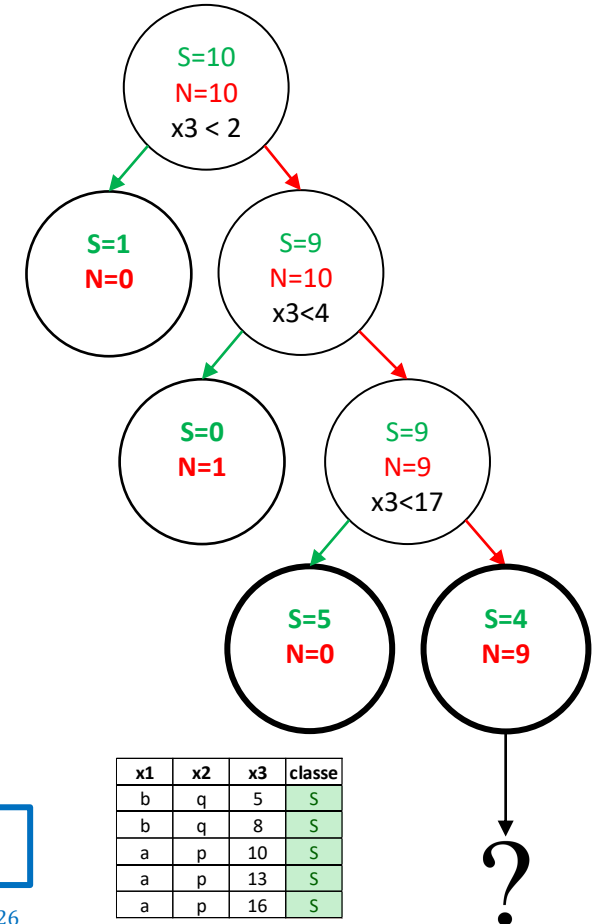
...

Algoritmo de construção de árvores de decisão (Gini Index)

x1	x2	x3	classe
a	p	20	N
a	p	10	S
a	p	13	S
a	p	16	S
a	q	22	N
a	q	26	N
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	p	24	S
b	q	8	S
b	p	34	S
b	q	5	S
b	p	28	S
b	q	30	N
b	q	26	N
b	p	38	S

x1	x2	x3	classe
a	p	20	N
b	p	24	S
a	p	10	S
a	p	13	S
a	p	16	S
b	p	34	S
b	p	28	S
b	p	38	S
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	q	8	S
b	q	5	S
b	q	30	N
b	q	26	N
a	q	22	N
a	q	26	N

x1	x2	x3	classe
b	q	5	S
b	q	8	S
a	p	10	S
a	p	13	S
a	p	16	S
b	q	18	N
b	q	19	N
a	p	20	N
a	q	22	N
b	p	24	S
b	q	26	N
b	q	26	N
a	q	26	N
b	p	28	S
b	q	30	N
b	q	32	N
b	p	34	S
b	p	38	S



x1 = a	V	S	p=3/6	G = 0,5
	p=6/18	N	p=3/6	
	F	S	p=6/12	G = 0,5
	p=12/18	N	p=6/12	

x2 = 9	V	S	p=7/8	G = 0,498
	p=8/18	N	p=1/8	
	F	S	p=2/10	G = 0,278
	p=10/18	N	p=8/10	

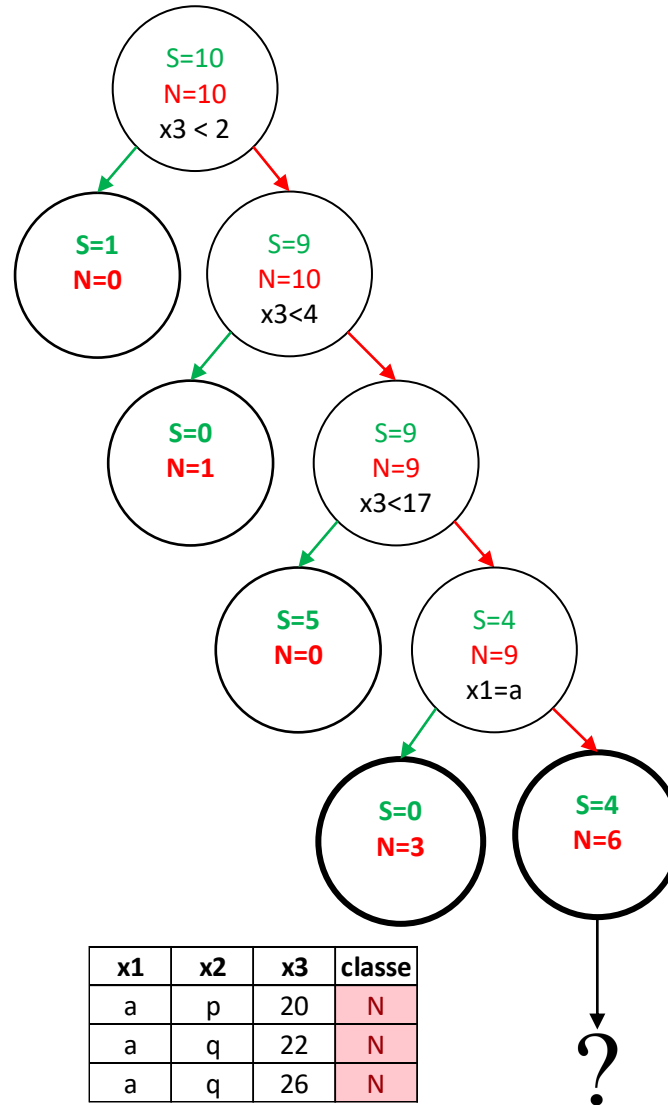
x3 < 17	V	S	p=5/5	G = 0
	p=5/18	N	p=0/5	
	F	S	p=4/13	G = 0,426
	p=13/18	N	p=9/13	

x1	x2	x3	classe
b	q	5	S
b	q	8	S
a	p	10	S
a	p	13	S
a	p	16	S

Algoritmo de construção de árvores de decisão (Gini Index)

x1	x2	x3	classe
a	p	20	N
a	q	22	N
a	q	26	N
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	p	24	S
b	p	34	S
b	p	28	S
b	q	30	N
b	q	26	N
b	p	38	S

x1 = a	V	S	p=0/3	G = 0
	p=3/13	N	p=3/3	
	F	S	p=4/10	
	p=10/13	N	p=6/10	

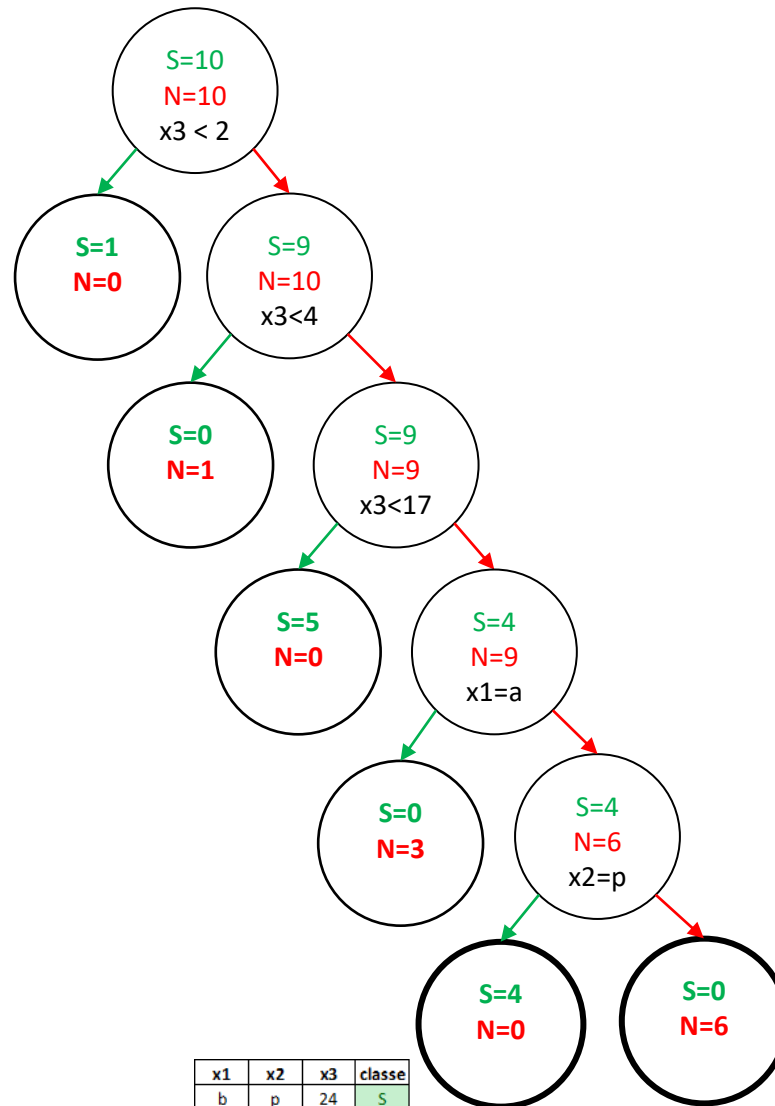


x1	x2	x3	classe
a	p	20	N
a	q	22	N
a	q	26	N

Algoritmo de construção de árvores de decisão (Gini Index)

x1	x2	x3	classe
b	p	24	S
b	p	34	S
b	p	28	S
b	p	38	S
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	q	30	N
b	q	26	N

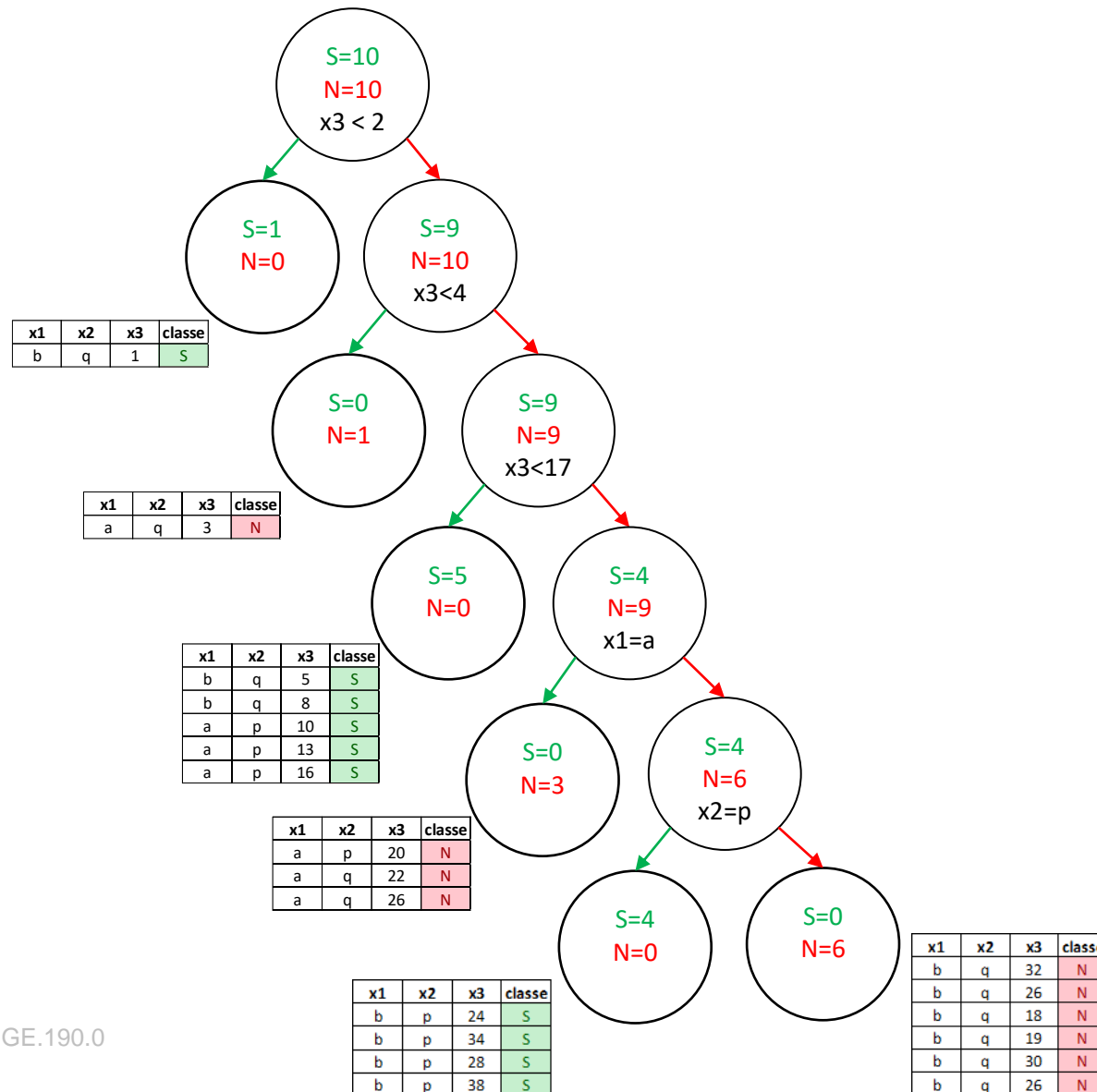
x2 = p	V	S	p=4/4	G = 0
	p=4/10	N	p=0/4	
	F	S	p=0/6	G = 0
	p=6/10	N	p=6/6	



x1	x2	x3	classe
b	p	24	S
b	p	34	S
b	p	28	S
b	p	38	S

x1	x2	x3	classe
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	q	30	N
b	q	26	N

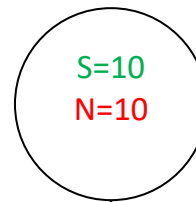
Algoritmo de construção de árvores de decisão (Gini Index)



Algoritmo de construção de árvores de decisão

x1	x2	x3	classe
b	q	1	S
a	q	3	N
b	q	5	S
b	q	8	S
a	p	10	S
a	p	13	S
a	p	16	S
b	q	18	N
b	q	19	N
a	p	20	N
a	q	22	N
b	p	24	S
a	q	26	N
b	q	26	N
b	q	26	N
b	p	28	S
b	q	30	N
b	q	32	N
b	p	34	S
b	p	38	S

Raiz da árvore:



?

Por onde vamos fazer o *split* ?

Medidas de avaliação de árvores de decisão

Entropy: é usada como uma forma de medir a aleatoriedade de uma coluna

$$E = - \sum_{i=1}^n p_i \times \log_2 p_i$$

p_i é a probabilidade de uma observação ser classificada numa classe em particular

Medidas de avaliação de árvores de decisão

Information Gain: mede a redução de entropia de um determinado *split*

$$IG(T, A) = E(T) - \sum \frac{|T_v|}{T} \times E(T_v)$$

T é o target (a classe)

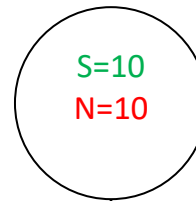
A é a variável

v é cada valor possível da variável

Algoritmo de construção de árvores de decisão (Information Gain)

x1	x2	x3	classe
b	q	1	S
a	q	3	N
b	q	5	S
b	q	8	S
a	p	10	S
a	p	13	S
a	p	16	S
b	q	18	N
b	q	19	N
a	p	20	N
a	q	22	N
b	p	24	S
a	q	26	N
b	q	26	N
b	q	26	N
b	p	28	S
b	q	30	N
b	q	32	N
b	p	34	S
b	p	38	S

Raiz da árvore:



Por onde vamos fazer o *split* ?

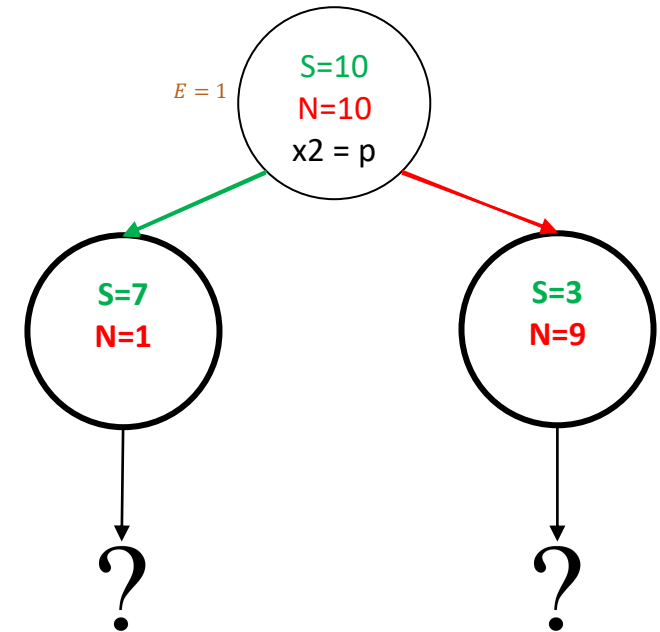
$$E = - \sum_{i=1}^n p_i \times \log_2 p_i = - \left(\frac{10}{20} \right) \times \log_2 \left(\frac{10}{20} \right) = 1$$

Algoritmo de construção de árvores de decisão (Information Gain)

x1	x2	x3	classe
a	q	3	N
a	p	20	N
a	p	10	S
a	p	13	S
a	p	16	S
a	q	22	N
a	q	26	N
b	q	1	S
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	p	24	S
b	q	8	S
b	p	34	S
b	q	5	S
b	p	28	S
b	q	30	N
b	q	26	N
b	p	38	S

x1	x2	x3	classe
a	p	20	N
b	p	24	S
a	p	10	S
a	p	13	S
a	p	16	S
b	p	34	S
b	p	28	S
b	p	38	S
b	q	1	S
a	q	3	N
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	q	8	S
b	q	5	S
b	q	30	N
b	q	26	N
a	q	22	N
a	q	26	N

x1	x2	x3	classe
b	q	1	S
a	q	3	N
b	q	5	S
b	q	8	S
a	p	10	S
a	p	13	S
a	p	16	S
b	q	18	N
b	q	19	N
a	p	20	N
a	q	22	N
b	p	24	S
b	q	26	N
b	q	26	N
a	q	26	N
b	p	28	S
b	q	30	N
b	q	32	N
b	p	34	S
b	p	38	S



x1 = a	V	S	p=3/7
	p=7/20	N	p=4/7 E = 0,985
	F	S	p=7/13
	p=13/20	N	p=6/13 E = 0,996

IMP.GE.190.0

$$IG = 1 - \left(\frac{7}{20} \times 0,985 + \frac{13}{20} \times 0,996 \right) = 0,00795$$

x2 = p	V	S	p=7/8
	p=8/20	N	p=1/8 E = 0,544
	F	S	p=3/12
	p=12/20	N	p=9/12 E = 0,811

IG = 0,295807

x3 < 2	V	S	p=1/1	G = 0
	p=1/20	N	p=0/1	E = 0
	F	S	p=9/19	
	p=19/20	N	p=10/19	E = 0,998

...

IG = 0,051899

IG(outros) << 0,295807

DEPARTAMENTO CIÊNCIA
E TECNOLOGIA

Algoritmo de construção de árvores de decisão (Information Gain)

x1	x2	x3	classe
a	p	20	N
a	p	10	S
a	p	13	S
a	p	16	S
b	p	24	S
b	p	34	S
b	p	28	S
b	p	38	S

x1	x2	x3	classe
a	p	10	S
a	p	13	S
a	p	16	S
a	p	20	N
b	p	24	S
b	p	28	S
b	p	34	S
b	p	38	S

x1 = a	V	S	p=3/4	E = 0,811
		N	p=1/4	
	F	S	p=4/4	
		N	p=0/4	

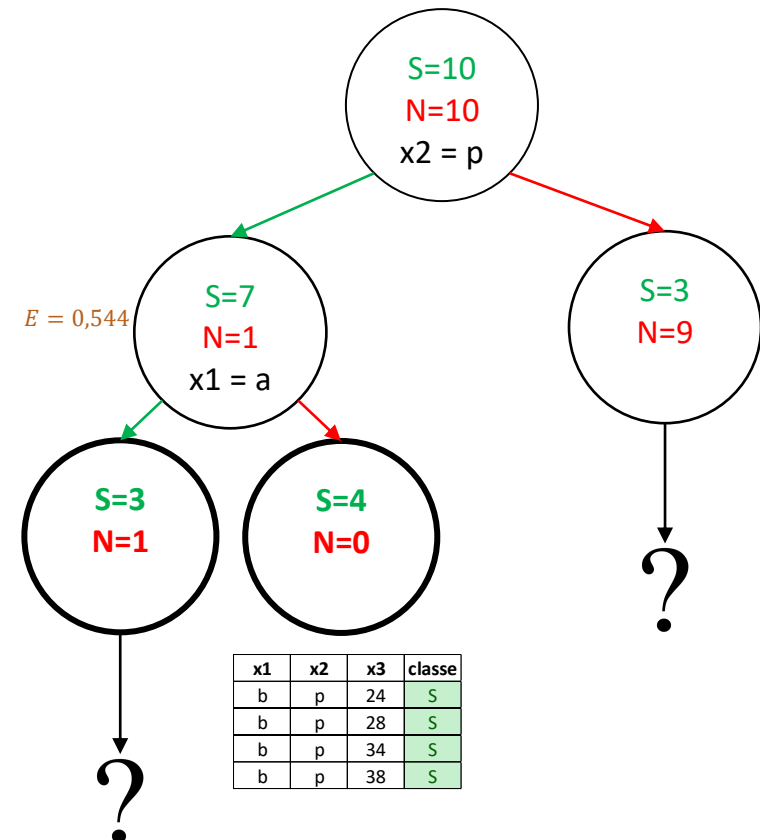
IG = 0,138361

x3 < 17	V	S	p=3/3	E = 0
		N	p=0/3	
	F	S	p=4/5	
		N	p=1/5	

IG = 0,092795

x3 < 23	V	S	p=3/4	E = 0,811
		N	p=1/4	
	F	S	p=4/4	
		N	p=0/4	

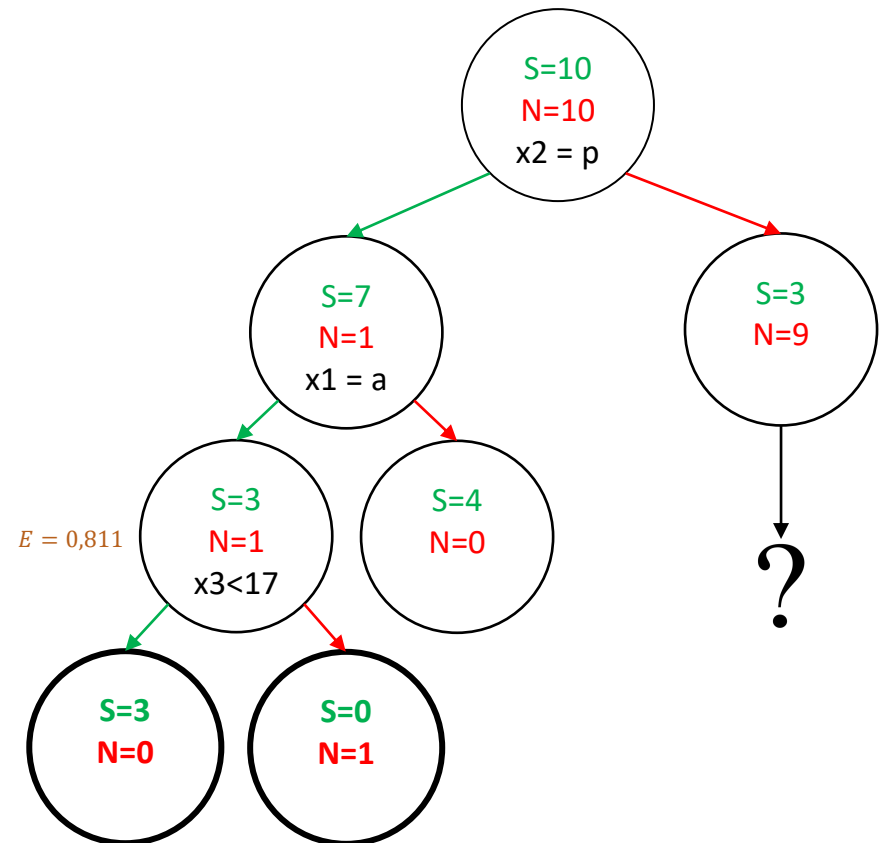
IG = 0,138361



Algoritmo de construção de árvores de decisão (Information Gain)

x1	x2	x3	classe
a	p	10	S
a	p	13	S
a	p	16	S
a	p	20	N

x3 < 17	V	S	p=3/3	E = 0
	p=3/4	N	p=0/3	
	F	S	p=0/1	
	p=1/4	N	p=1/1	
IG = 0,811				



x1	x2	x3	classe
a	p	10	S
a	p	13	S
a	p	16	S

x1	x2	x3	classe
a	p	20	N

Algoritmo de construção de árvores de decisão (Information Gain)

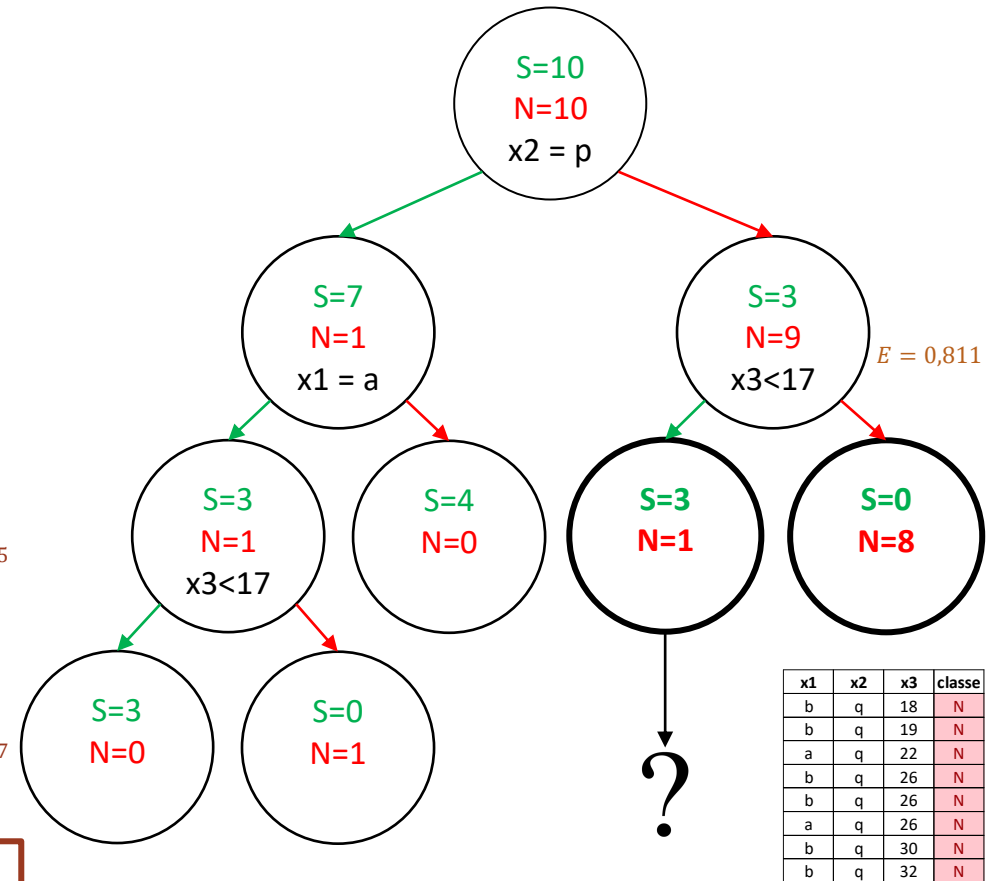
x1	x2	x3	classe
a	q	3	N
a	q	22	N
a	q	26	N
b	q	1	S
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	q	8	S
b	q	5	S
b	q	30	N
b	q	26	N

x1	x2	x3	classe
b	q	1	S
a	q	3	N
b	q	5	S
b	q	8	S
b	q	18	N
b	q	19	N
a	q	22	N
b	q	26	N
b	q	26	N
a	q	26	N
b	q	30	N
b	q	32	N

$$\begin{array}{l|l|l}
 \mathbf{x1 = a} & \begin{array}{l} \mathbf{V} \\ p=3/12 \end{array} & \begin{array}{l} \mathbf{S} \quad p=0/3 \\ \mathbf{N} \quad p=3/3 \end{array} & E = 0 \\
 & \mathbf{F} \\ p=9/12 & \begin{array}{l} \mathbf{S} \quad p=3/9 \\ \mathbf{N} \quad p=6/9 \end{array} & E = 0,918 \\
 & & IG = 0,122278
 \end{array}$$

$$\begin{array}{l|l|l}
 \mathbf{x3 < 2} & \begin{array}{l} \mathbf{V} \\ p=1/12 \end{array} & \begin{array}{l} \mathbf{S} \quad p=1/1 \\ \mathbf{N} \quad p=0/1 \end{array} & E = 0 \\
 & \mathbf{F} \\ p=11/12 & \begin{array}{l} \mathbf{S} \quad p=2/11 \\ \mathbf{N} \quad p=9/11 \end{array} & E = 0,684 \\
 & & IG = 0,183965 \\
 \\
 \mathbf{x3 < 4} & \begin{array}{l} \mathbf{V} \\ p=2/12 \end{array} & \begin{array}{l} \mathbf{S} \quad p=1/2 \\ \mathbf{N} \quad p=1/2 \end{array} & E = 1 \\
 & \mathbf{F} \\ p=10/12 & \begin{array}{l} \mathbf{S} \quad p=2/10 \\ \mathbf{N} \quad p=8/10 \end{array} & E = 0,722 \\
 & & IG = 0,042727
 \end{array}$$

$$\begin{array}{l|l|l}
 \mathbf{x3 < 17} & \begin{array}{l} \mathbf{V} \\ p=4/12 \end{array} & \begin{array}{l} \mathbf{S} \quad p=3/4 \\ \mathbf{N} \quad p=1/4 \end{array} & E = 0,811 \\
 & \mathbf{F} \\ p=8/12 & \begin{array}{l} \mathbf{S} \quad p=0/8 \\ \mathbf{N} \quad p=8/8 \end{array} & E = 0 \\
 & & IG = 0,540574
 \end{array}$$



x1	x2	x3	classe
b	q	18	N
b	q	19	N
a	q	22	N
b	q	26	N
b	q	26	N
a	q	26	N
b	q	30	N
b	q	32	N

Algoritmo de construção de árvores de decisão (Information Gain)

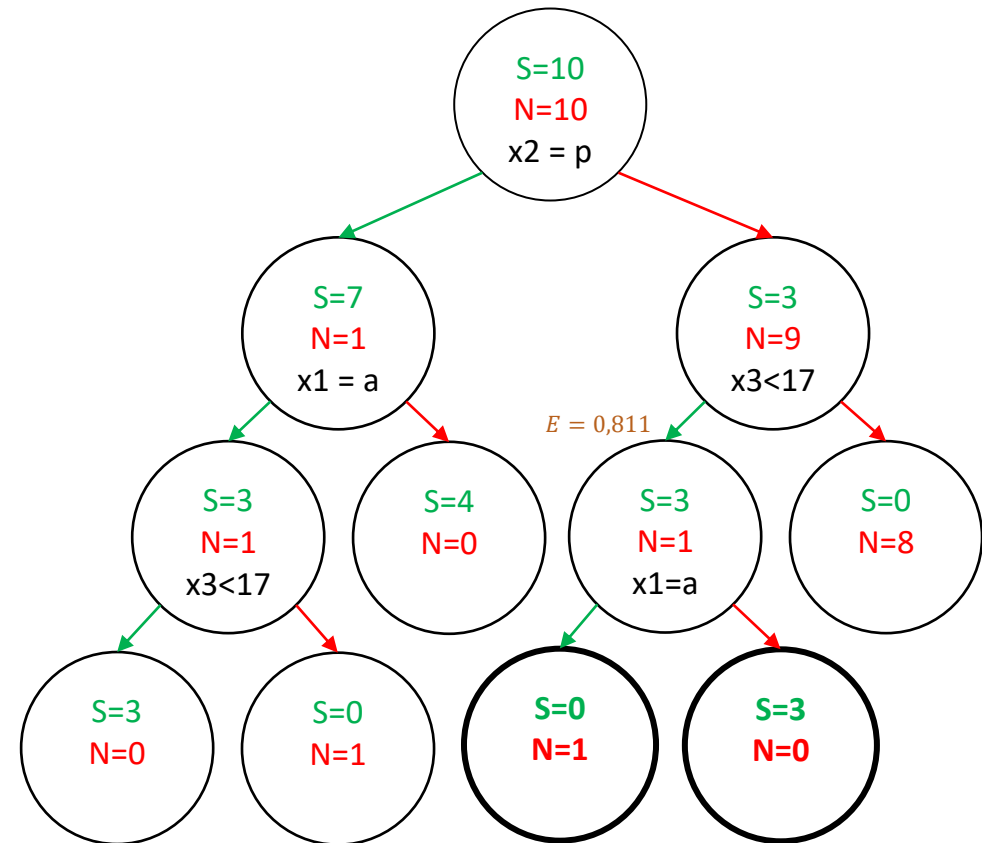
x1	x2	x3	classe
a	q	3	N
b	q	1	S
b	q	8	S
b	q	5	S

x1	x2	x3	classe
b	q	1	S
a	q	3	N
b	q	5	S
b	q	8	S

x1 = a	V	S	$p=0/1$
	$p=1/4$	N	$p=1/0$ $E = 0$
	F	S	$p=3/3$
	$p=3/4$	N	$p=0/3$ $E = 0$
$IG = 0,811$			

x3 < 2	V	S	$p=1/1$
	$p=1/4$	N	$p=0/1$ $E = 0$
	F	S	$p=2/3$ $IG = 0,122278$
	$p=3/4$	N	$p=1/3$ $E = 0,918$

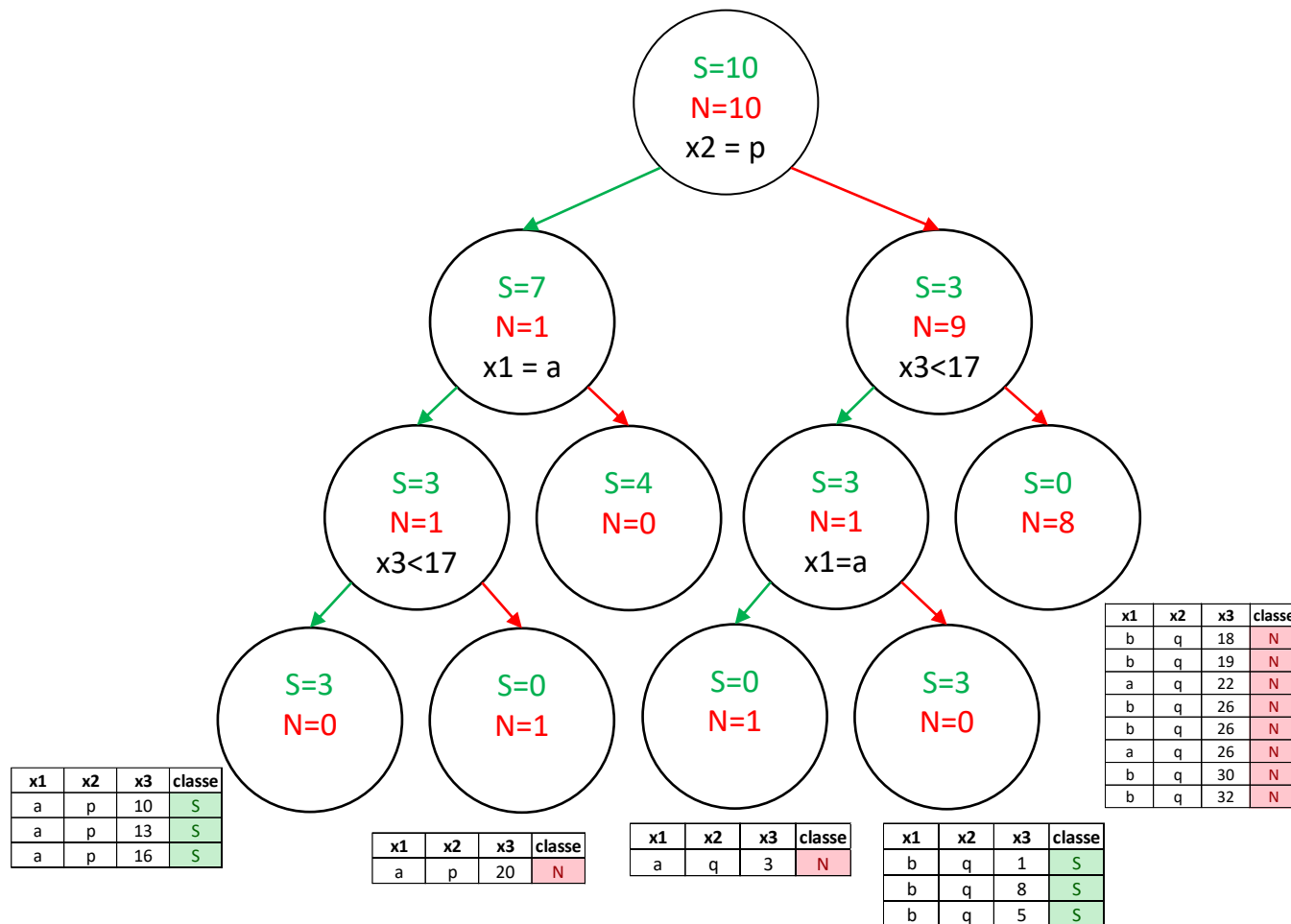
x3 < 4	V	S	$p=1/2$
	$p=2/4$	N	$p=1/2$ $E = 1$
	F	S	$p=2/2$ $IG = 0,311$
	$p=2/4$	N	$p=0/2$ $E = 0$



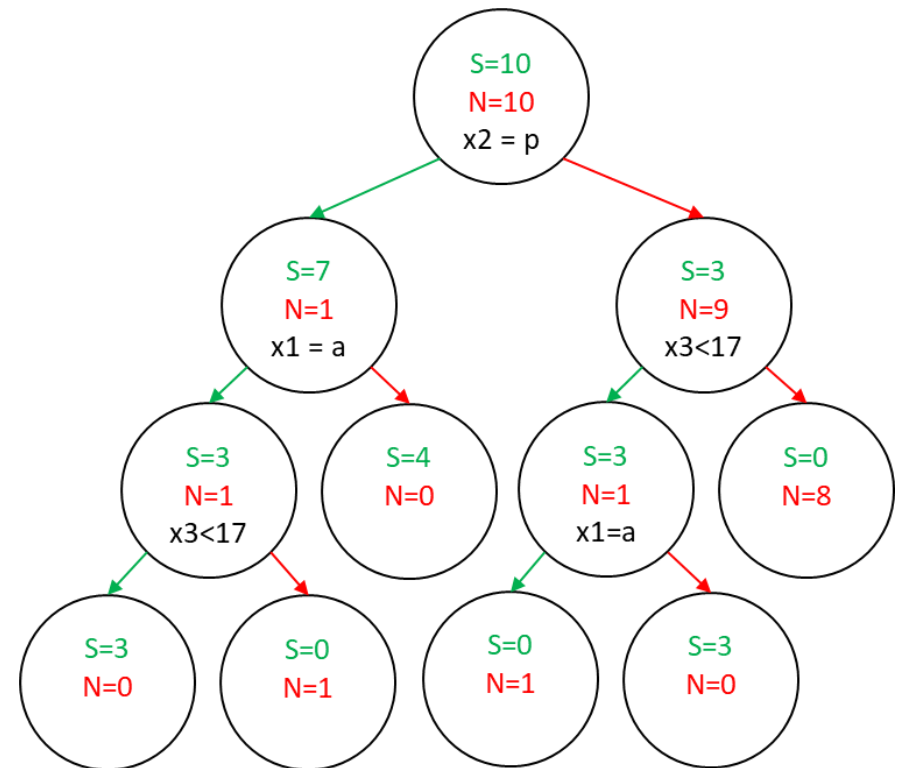
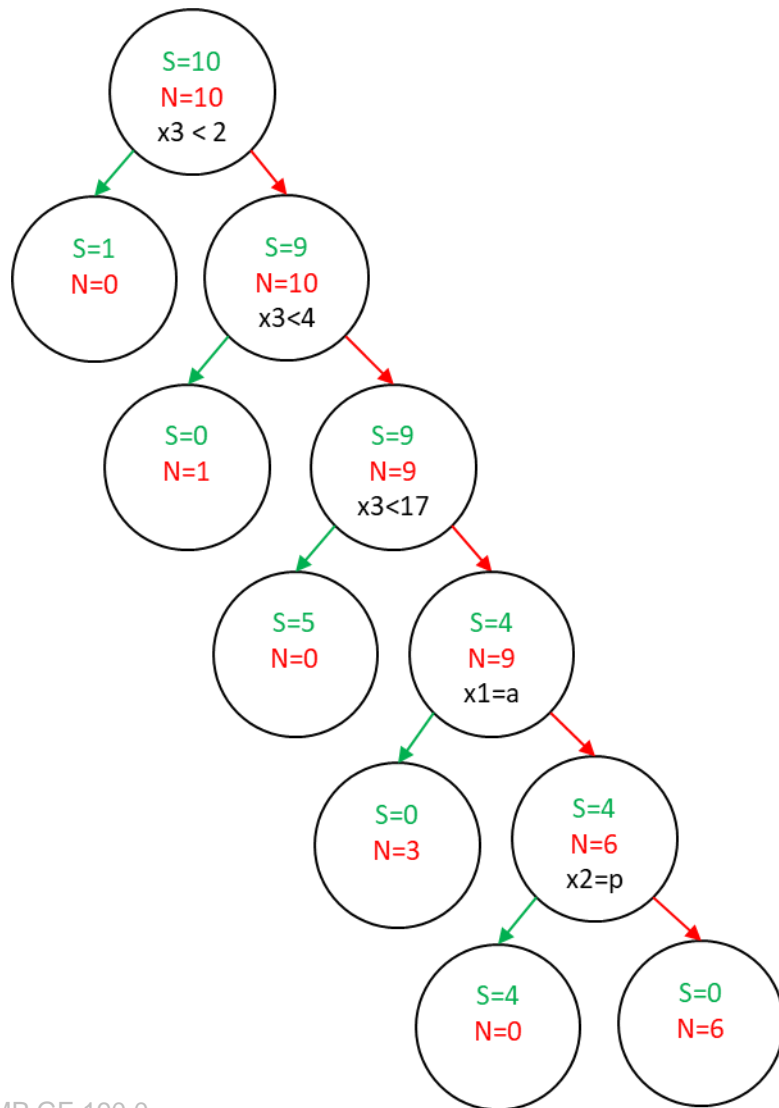
x1	x2	x3	classe
a	q	3	N

x1	x2	x3	classe
b	q	1	S
b	q	8	S
b	q	5	S

Algoritmo de construção de árvores de decisão (Information Gain)



Algoritmo de construção de árvores de decisão (Gini Index vs. Information Gain)



Algoritmo de árvores de decisão: quando parar?

No limite, podemos ter uma “folha” para cada exemplo (*overfitting*)

Como evitar? Saber quando parar:

- Estabelecendo um **número mínimo de exemplos por folha**
- Estabelecendo uma **profundidade máxima para a árvore**
- **Pruning**: remover ramos que utilizam variáveis de baixa importância
 - Reduced error pruning: começa numa folha e remove os nós com a classe mais popular dessa folha, se não piorar a métrica de avaliação. Repete para outras folhas.
 - Cost complexity pruning / weakest link pruning: é utilizado um parâmetro (α) para determinar se um determinado nó pode ser removido tendo como base o tamanho da sub-árvore.

Mais sobre *overfitting*:

Hawkins, D. M. (2004). The problem of overfitting. Journal of chemical information and computer sciences, 44(1), 1-12.

<https://pubs.acs.org/doi/pdf/10.1021/ci0342472>

Avaliação do modelo de Classificação

		Observação	
		P	N
Previsão	P	TP	FP
	N	FN	TN

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision, Positive Predictive Value: PPV = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Negative Predictive Value: NPV = \frac{TN}{TN + FN}$$

$$False Positive Rate: FPR = \frac{FP}{FP + TN}$$

$$False Negative Rate: FNR = \frac{FN}{TP + FN}$$

Vantagens e desvantagens de árvores de decisão

Vantagens	Desvantagens
<ul style="list-style-type: none"> • Fáceis de compreender, interpretar e visualizar • Executam <i>feature selection</i> implicitamente • Permitem usar dados numéricos e categóricos • Conseguem lidar com dados com vários <i>targets</i> • Não obrigam a muito esforço no processo de <i>data preparation</i> • Relações não-lineares não afetam o desempenho da árvore 	<ul style="list-style-type: none"> • Podem criar árvores demasiado complexas que não generalizam (<i>overfitting</i>) • <i>Variance</i>: Pequenas variações nos dados podem fazer com que seja criada uma árvore completamente diferente <ul style="list-style-type: none"> • <i>Variance</i> pode ser reduzida com métodos como <i>bagging</i>⁽¹⁾ e <i>boosting</i>⁽¹⁾ • Algoritmos <i>greedy</i> não garantem a criação da árvore de decisão ótima <ul style="list-style-type: none"> • Para ultrapassar, podem criar-se várias árvores, em que as <i>features</i> e as amostras são selecionadas aleatoriamente: <i>Random Forest</i>⁽¹⁾ • Se houver uma classe dominante, o modelo poderá criar uma árvore enviesada (<i>biased</i>) <ul style="list-style-type: none"> • Recomenda-se balancear⁽²⁾ os dados antes da aplicação de modelos de árvores de decisão

⁽¹⁾ Métodos ***bagging***, ***boosting*** e ***Random Forest*** serão vistos mais à frente, durante o semestre

⁽²⁾ Técnicas para lidar com dados *imbalanced*: <https://www.analyticsvidhya.com/blog/2017/03/imbalanced-data-classification/>

Árvores de decisão para regressão

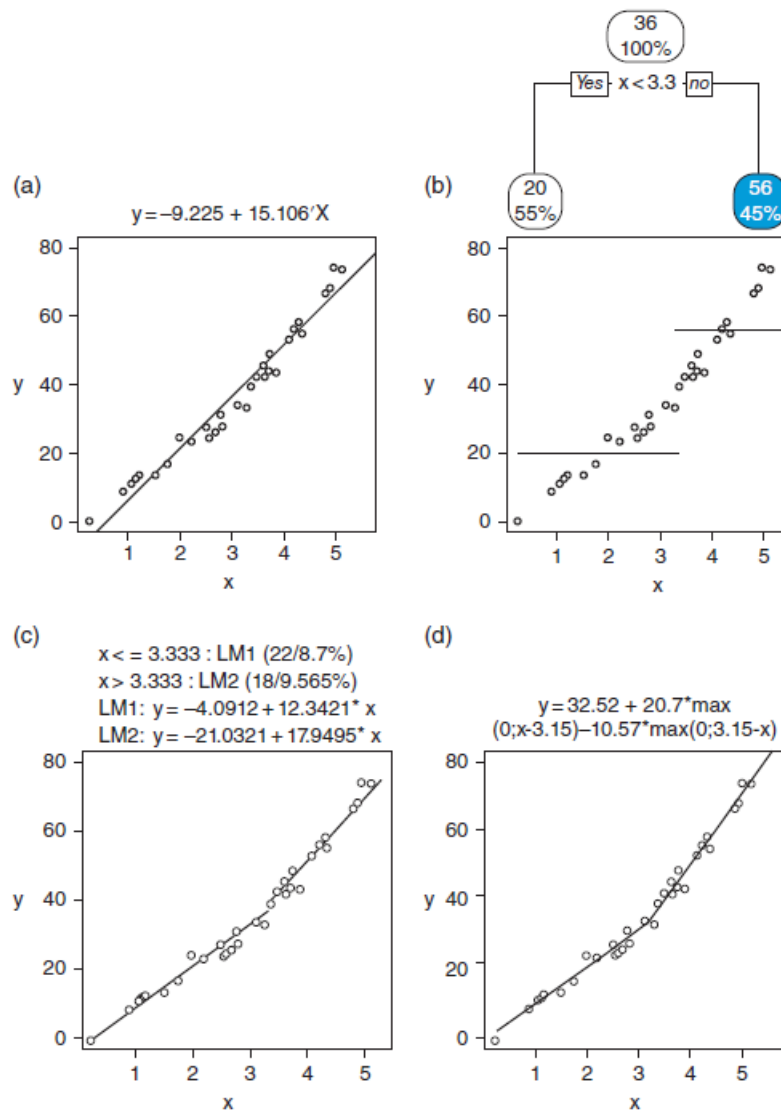
O princípio e o algoritmo são os mesmos

O que muda:

- Métricas. Por exemplo:
 - Redução da variância
 - Redução do desvio padrão / Standard Deviation Reduction (ver http://blog.saedsayad.com/decision_tree_reg.htm)
- Forma como as folhas preveem os valores:
 - Em classificação escolhe-se a maioria, em regressão geralmente escolhe-se a média (CART)
 - Model Trees e Multivariate Adaptive Regression Splines (MARS) usam multivariate linear regression em vez da média

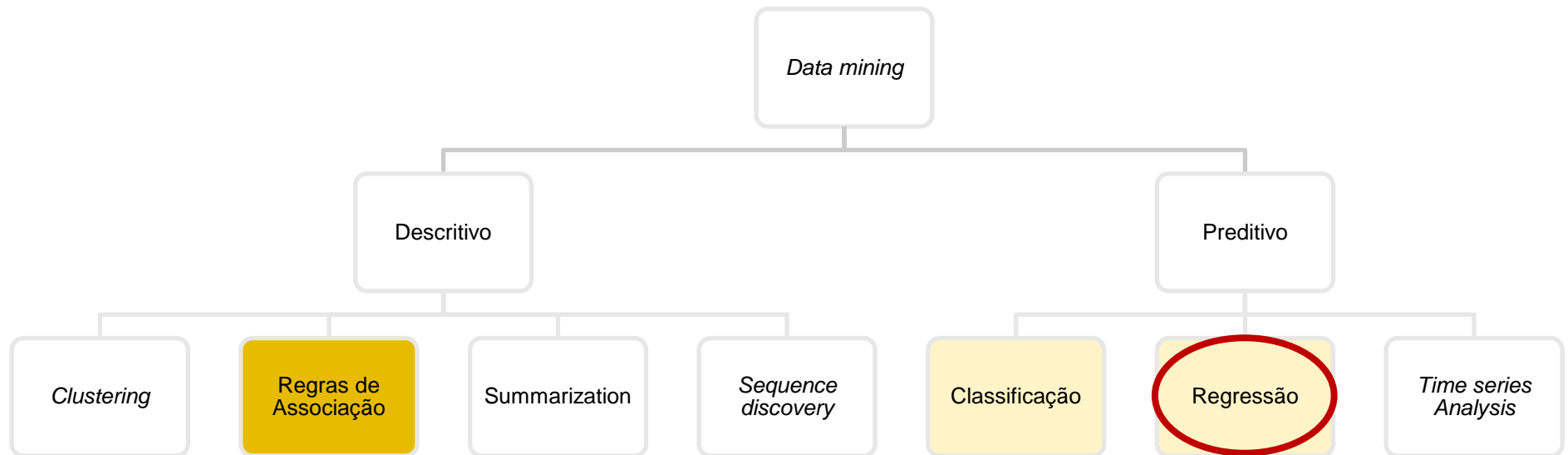
[1] Quinlan, J.R. (1992) Learning with continuous classes, in *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, World Scientific, pp. 343–348.

Árvores de decisão para regressão: comparação de modelos



- (a) MLR
- (b) CART
- (c) Model trees
- (d) MARS

Este capítulo



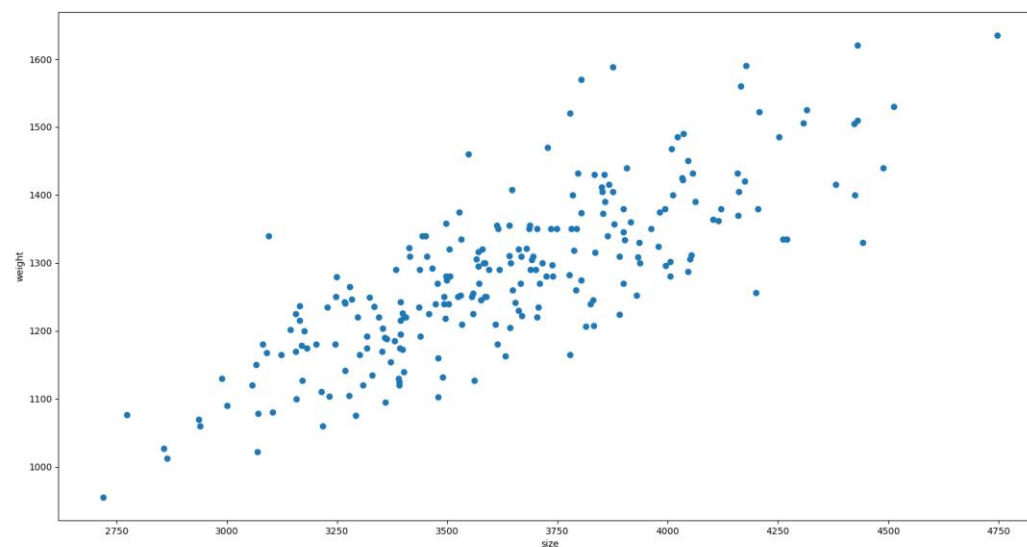
Notação

\bar{x} Média da variável x

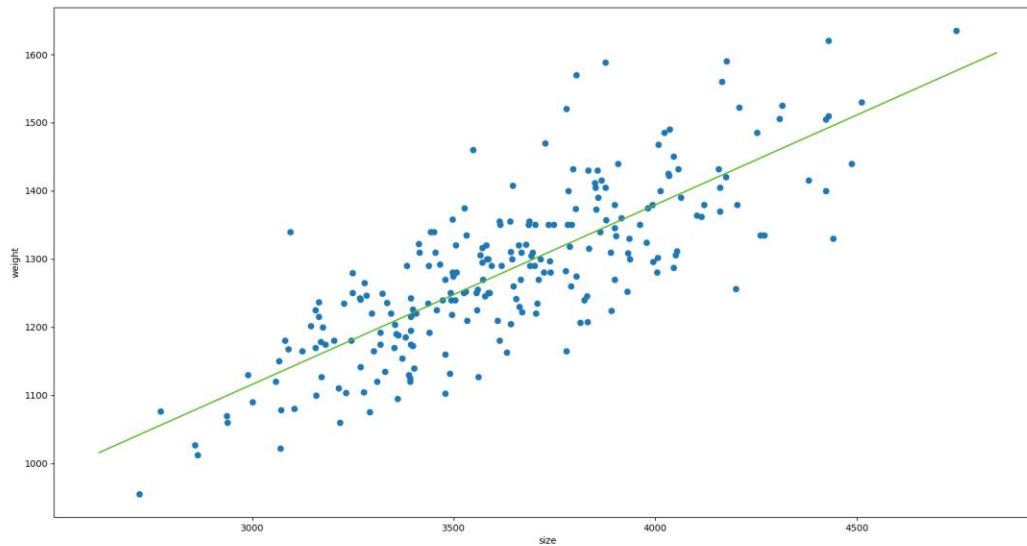
\hat{x} Previsão da variável x

Regressão Linear Simples

size	weight
4512	153
3738	1297
4261	1335
3777	1282
4177	159
3585	13
...	...



Regressão Linear Simples



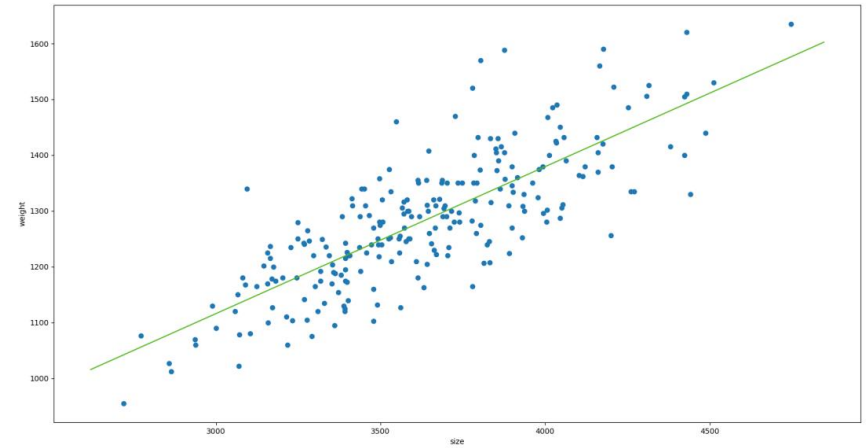
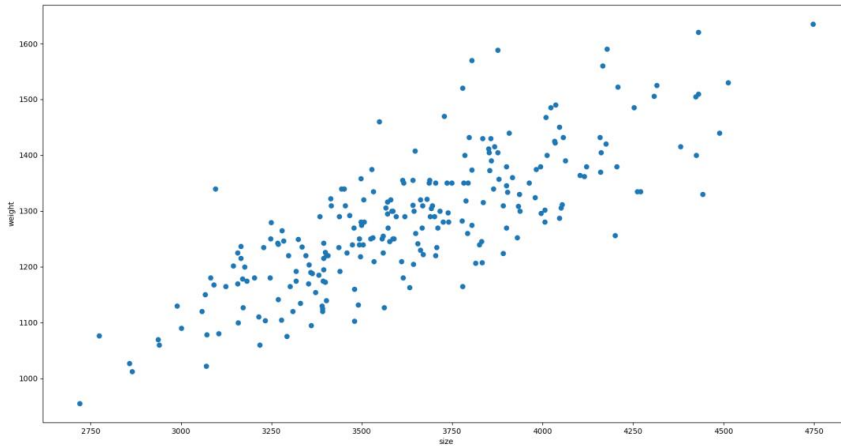
$$y = mx + b$$

$$y = \beta_0 + \beta_1 x_1$$

$$\text{weight} = \beta_0 \times \text{size} + \beta_1$$

? ?

Determinação do declive (m) e da ordenada na origem (b)



$$\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(x_i - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_1 = 0,26342933948939945$$

$$\beta_0 = 325,57342104944223$$

$$weight \approx 0,2634 \times size + 325,5734$$

Avaliação do modelo de Regressão

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

$$\text{Mean Absolute Error: } MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

Tem a mesma unidade de medida de y

$$\text{Mean Squared Error: } MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$$

Tem a mesma unidade de medida do quadrado de y . Enfatiza mais os erros maiores

$$\text{Root Mean Squared Error: } RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

Tem a mesma unidade de medida de y

$$\text{Relative Mean Squared Error: } RelMSW = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$$

Compara a capacidade preditiva com a da previsão trivial (média).

Valores possíveis:

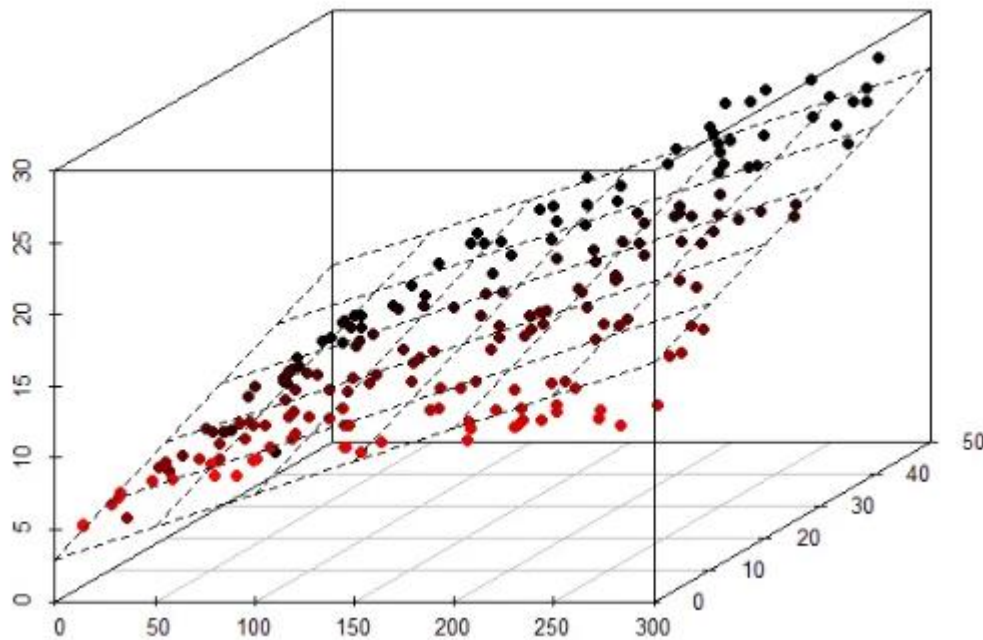
- **0**: modelo perfeito
- **]0,1[**: modelo útil
- **1**: modelo tão útil como prever a média
- **>1**: modelo inútil (pior do que prever a média)

$$\text{Relative Root Mean Squared Error: } RelRMSW = \sqrt{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}}$$

Regressão Linear Múltipla

Com duas variáveis independentes (x_1 e x_2) uma dependente (y):

Em vez de uma reta, teremos um plano



$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Regressão Linear Múltipla

Com n variáveis independentes (x_1, x_2, \dots, x_n) uma dependente (y):

Não é possível visualizar

Mas podemos generalizar:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$



UNIVERSIDADE
PORTUCALENSE

Do conhecimento à prática.