Estimation, Detection and Learning II

General Fundamentals

Catarina Oliveira

DCT DEPARTAMENTO CIÊNCIA E TECNOLOGIA

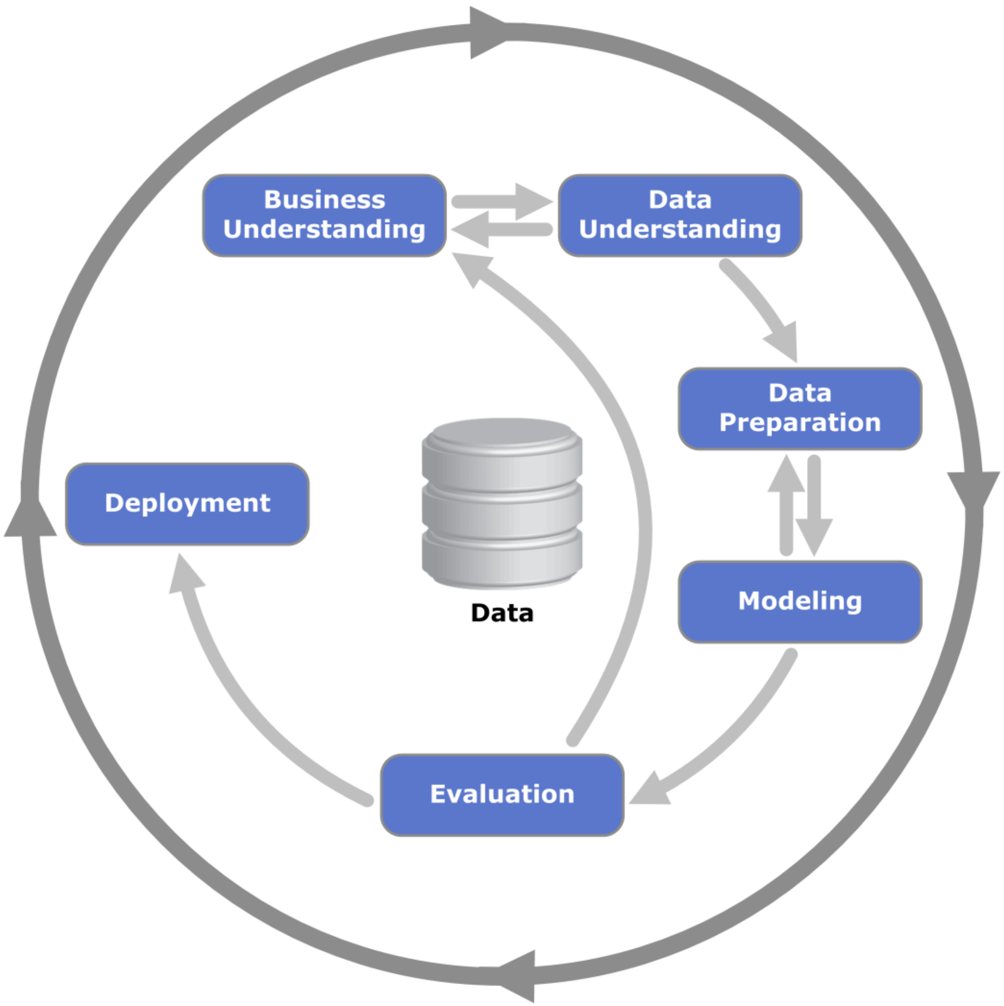
UPT UNIVERSIDADE PORTUCALENSE

IMP.GE.190.0

CONTENT

1. Introduction
2. descriptive *data mining*
    1. Association Rules
3. predictive *data mining*
    1. Classification
    2. Regression

CRISP-DM Model



Business Understanding :

- Understanding of project objectives and requirements
- Conversion of this knowledge into:
  - *data mining*
  - preliminary plan

Data Understanding :

- Initial data collection
- Familiarization with the data
  - identify data quality issues
  - uncover first insights into the data
  - detect interesting subsets

Data Preparation :

- Build the final dataset from the initial raw data.

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

UPT DCT DEPARTAMENTO CIÊNCIA E TECNOLOGIA

CRISP-DM Model

Modeling :

- Select and apply modeling techniques
    - Some techniques have specific requirements regarding the shape of the data
        - Possible need to return to preparation

Rating :

- Testing the models created
    - Based on performance measures to evaluate
        - Generalization
        - response to the goal
- Choice of the "winning" model

Deployment :

- Publication of the chosen model so that the model can be applied to new data

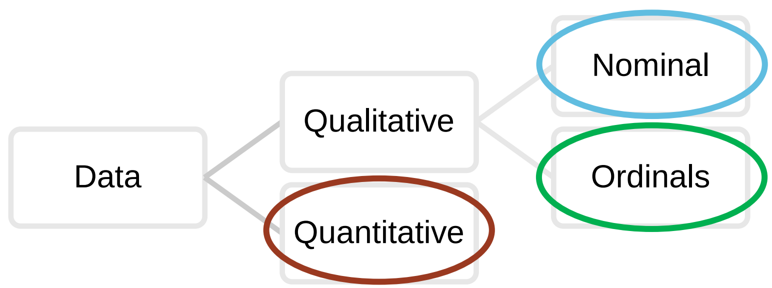
https://www.datasciencecentral.com/profiles/blogs/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome

Data Types

CLIENTE	IDADE	ALTURA	PESO	GENERO	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

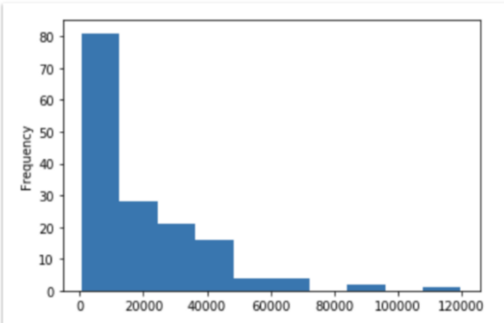
Independent variables

dependent variable
objective variable
target

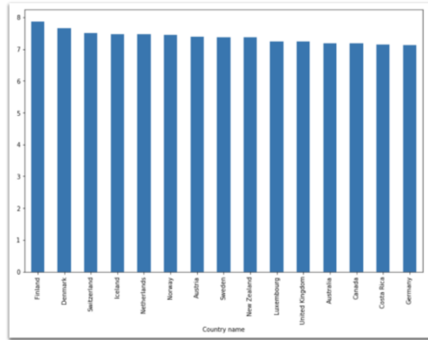


DEPARTAMENTO CIÊNCIA E TECNOLOGIA

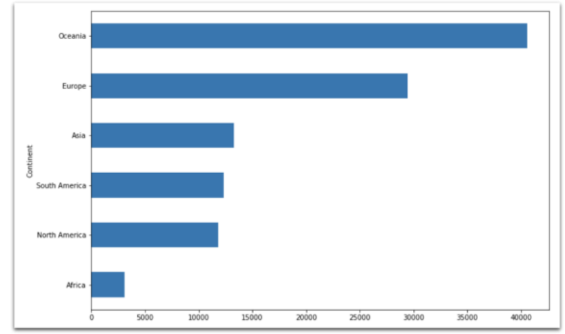
data visualization



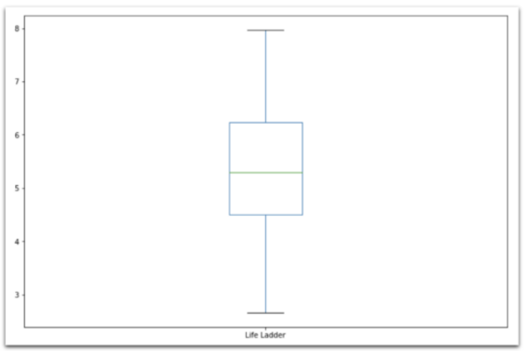
histogram



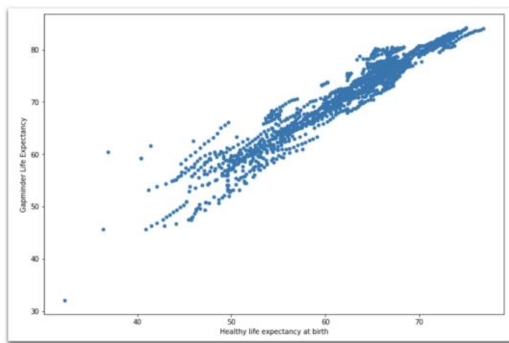
vertical bars



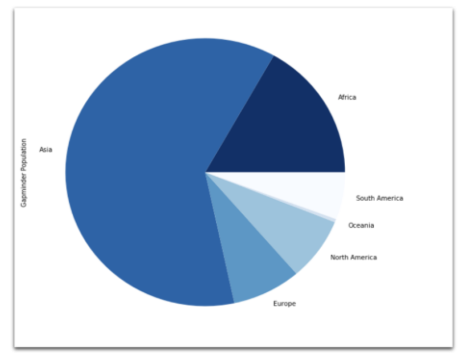
horizontal bars



boxplot _



scatter plot



pie chart

Plots in Python : https://towardsdatascience.com/plotting-with-python-c2561b8c0f1f

Data Exploration in Python : https://towardsdatascience.com/exploring-univariate-data-e7e2dc8fde80

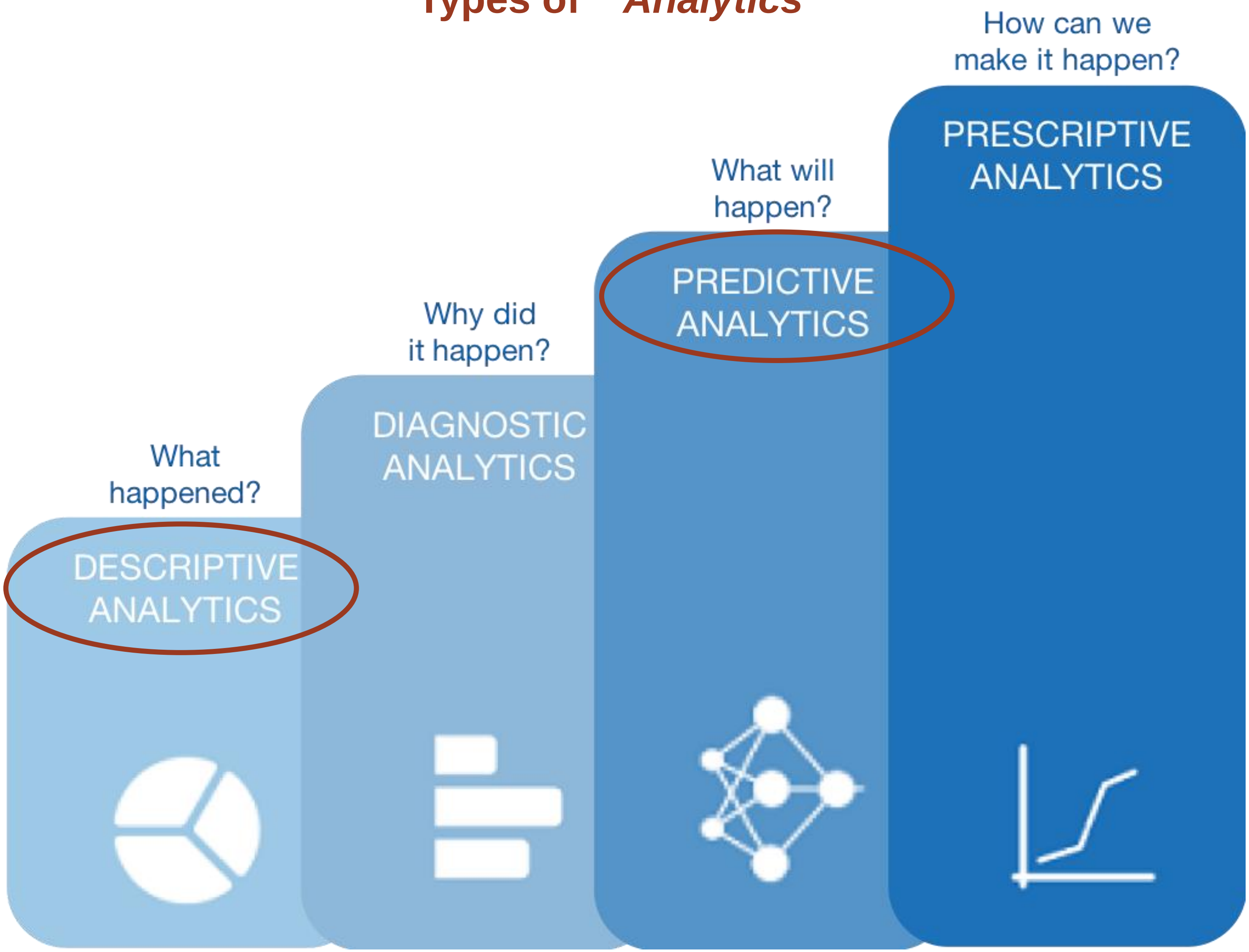
(Possible) Problems with Data / Solutions

Problems	Examples	Solutions:
unfilled fields		• Represent other information (ex : ND, NA) • Semiautomatic completion (preset value, expected [mean, mode, median], predicted)
errors/noise	locality=" Prto " age=-1	• replace values • Represent other information (ex : ND, NA, Error)
systematic errors	age=99 (because the system forces you to enter age even when it is unknown)	• Represent other information (ex : ND, NA, Error) • Semi-automatic filling (preset value, "expected" [mean, mode, median], predicted)
inconsistencies	location="VN de Gaia" and "Gaia"	• replace values
Outliers	expenses on cars with bank cards worth €100,000	• Replace with extreme value of distribution

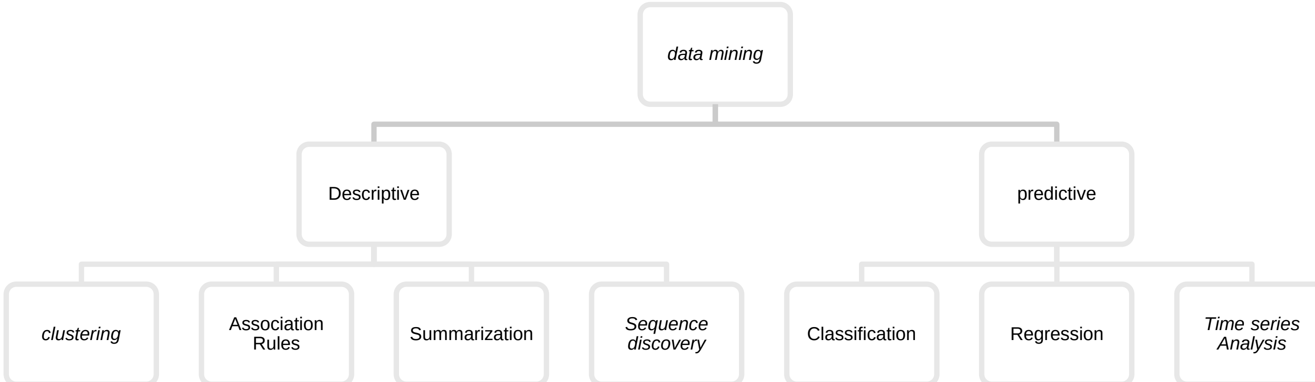
Limit solution : delete rows/columns

- impact depends on the number of rows affected by the issue

Types of " *Analytics* "

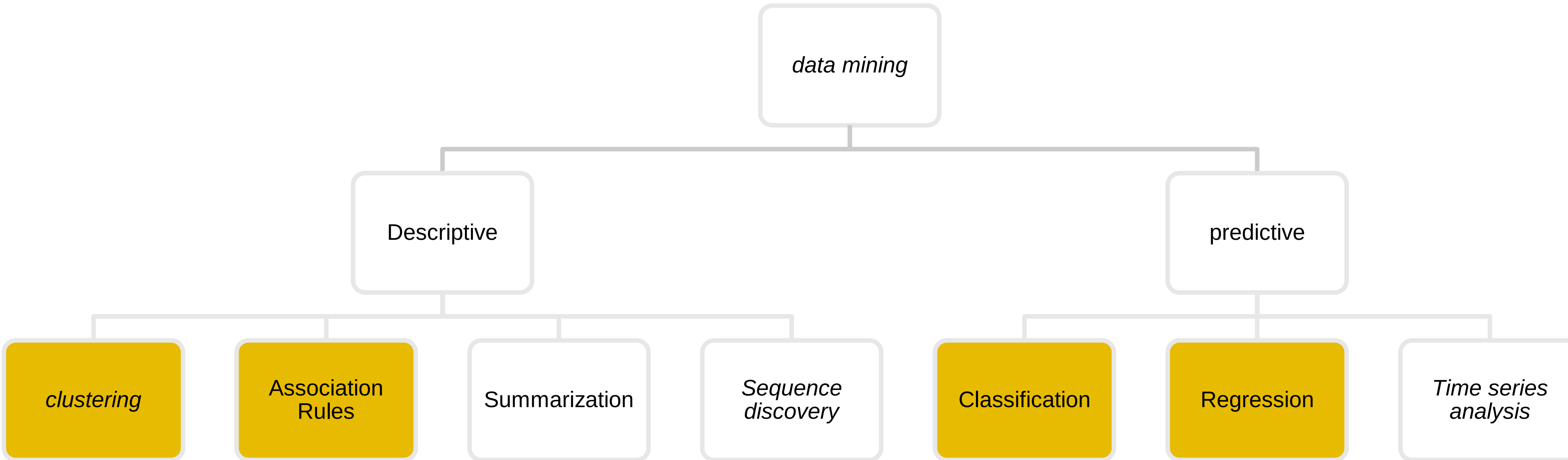


Descriptive and predictive

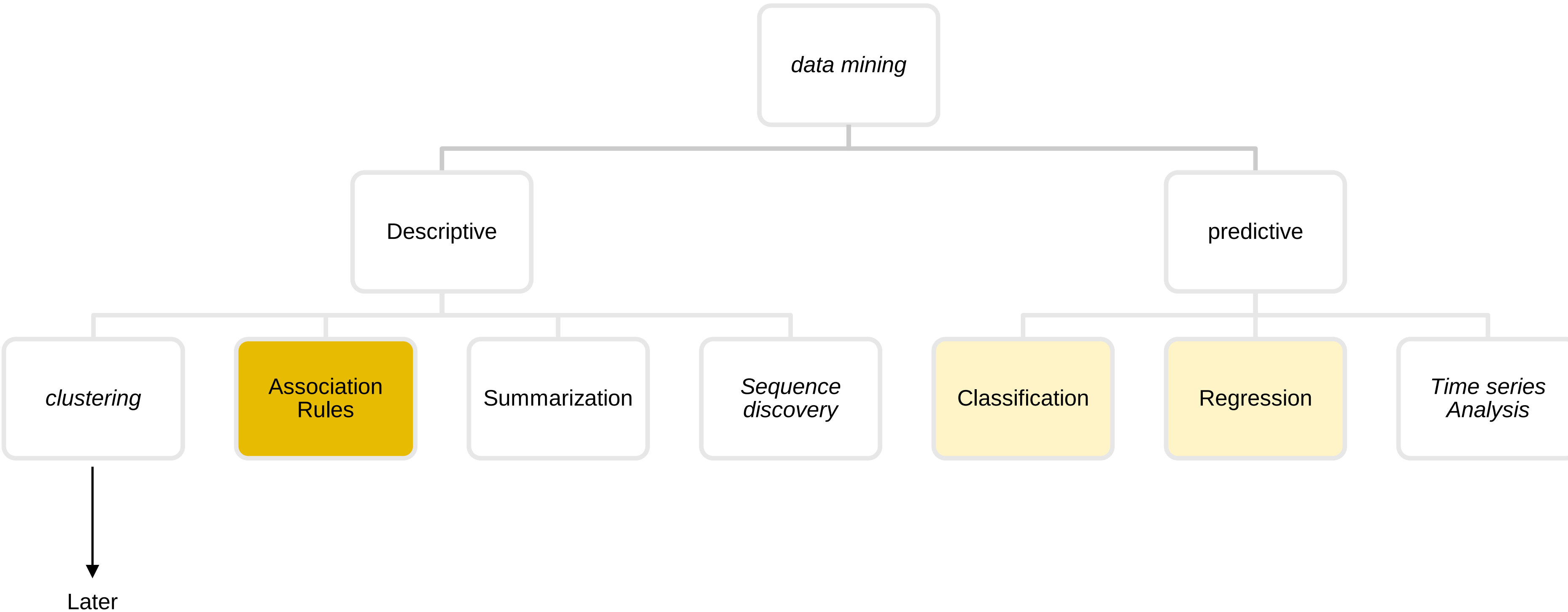


Determines	what happened in the past	What might happen in the future
Precision	accurate results	not guaranteed
Practical analysis methods	Standard reports, query	Predictive modeling, forecasting, simulation and alerting.
Requirements	Data aggregation and data mining	Statistical and forecasting methods
Approach	reactive	Proactive
Describe	Data characteristics	Induction on present and past data to make predictions
Questions	• What happened? • What is the problem? • How often does the problem happen?	• What will happen? • What is the outcome if the trend holds? • What actions are needed?

Estimation, Detection and Learning II

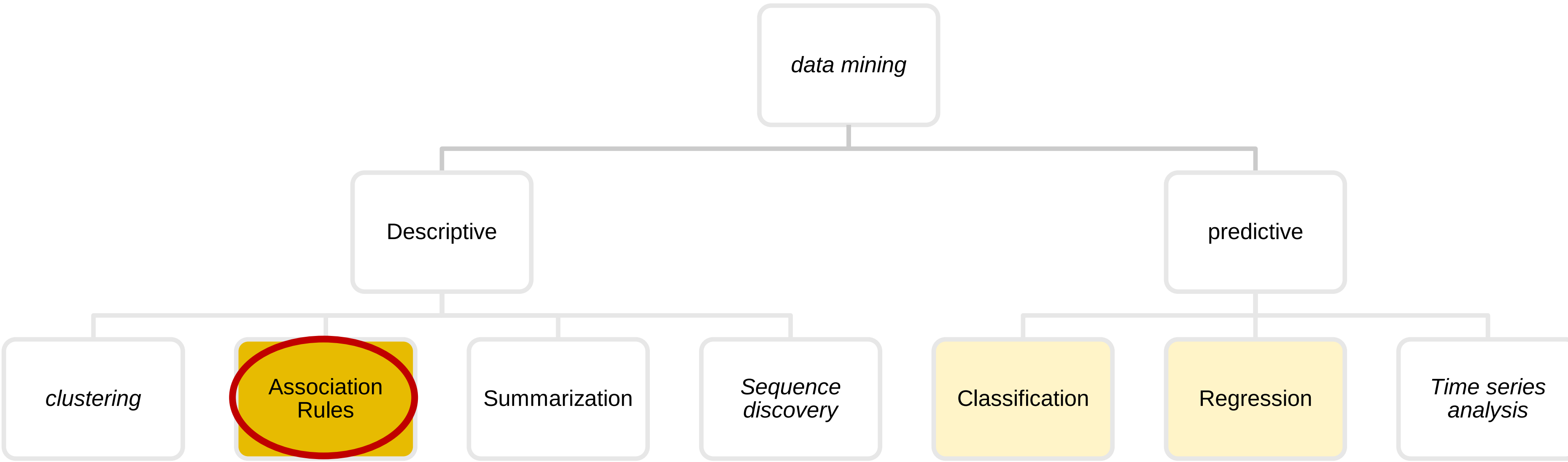


this chapter

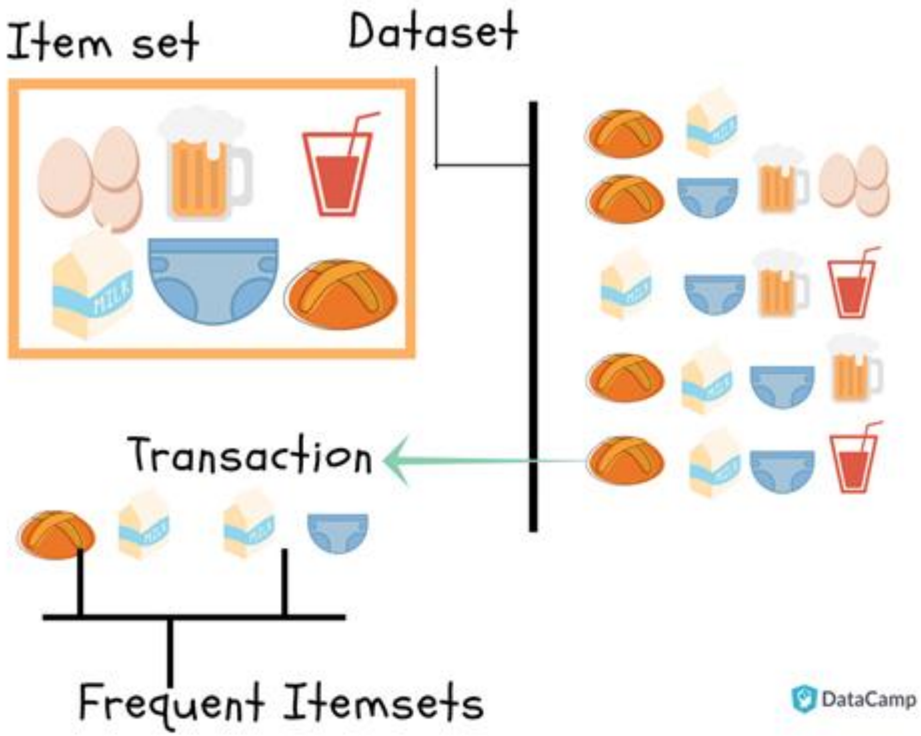


DEPARTAMENTO CIÊNCIA E TECNOLOGIA

this chapter



Association Rules



Given a set of transactions (*Transactions*)

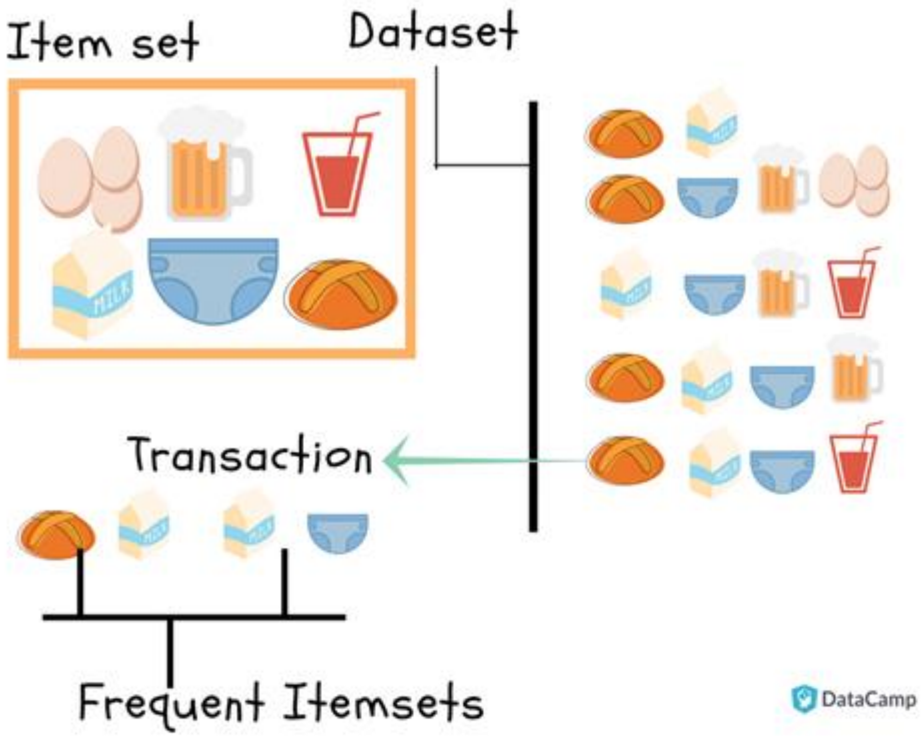
Identify:

- frequent co-occurrences (*Frequent itemsets*)
- *Itemsets* that originate other *itemsets* (Association Rules)

Examples:

- shopping baskets
- credit transactions
- clickstreams
- recommendation systems

itemset mining : Definition



Given away:

- A set of transactions $D = \{t_1, t_2, \ldots, t_n\}$

```
D = {{bread, milk},
{bread, diapers, beer, eggs},
{milk, diapers, beer, soft drinks},
{bread, milk, diapers, beer},
{bread, milk, diapers, soft drinks}}
```

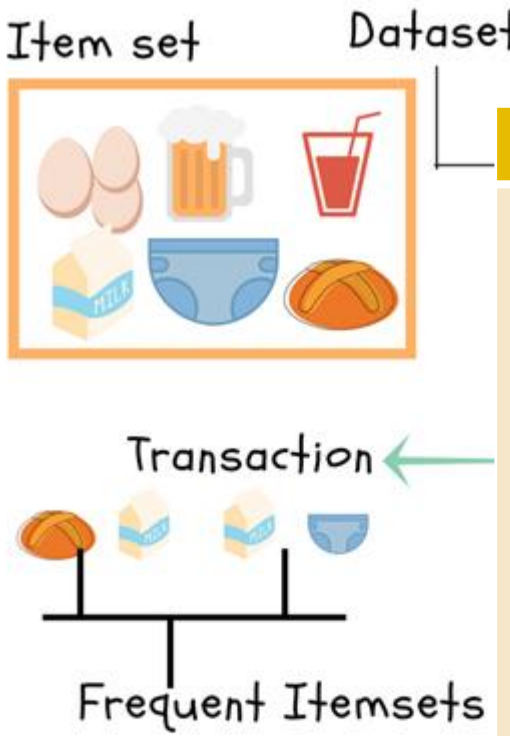
- a minimum support $sup_{min} \in [0,1]$

find the itemsets $X : Support(X) > sup_{min}$,

Support: Relative frequency (probability) of X in D

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

itemset mining : Example




```
Products = {eggs, beer, soda, milk, diapers, bread}
={O,C,R,L,F,P}
```

```
Possible 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}
{W}	{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 : Example

Frequency of Itemsets :

1	#	two	#	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
{W}	3	{O, R}	0	{O, C, L}	0	{O, C, R, F}	0	{O, C, R, L, P}	0		
{R}	two	{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}	two	{O, R, P}	0	{O, R, L, F}	0				
		{C, F}	3	{O, L, F}	0	{O, R, L, P}	0				
		{C, P}	two	{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}	two	{C, L, F, P}	1				
		{F, P}	3	{C, L, P}	1	{R, L, F, P}	1				
				{C, F, P}	two						
				{R, L, F}	1						
				{R, L, P}	1						
				{R, F, P}	1						
				{L, F, P}	two						

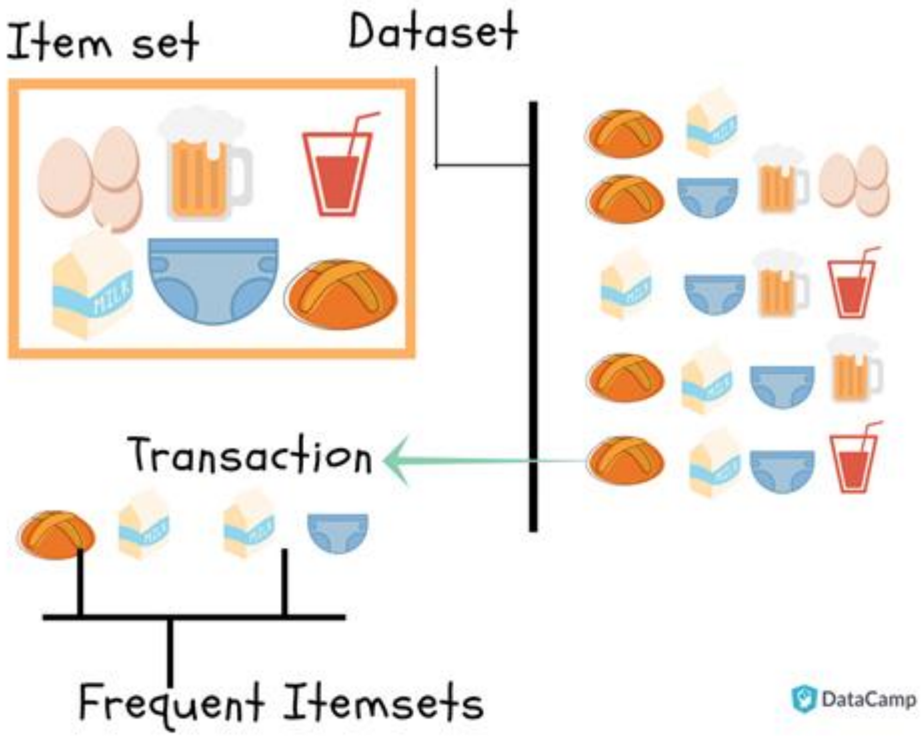
frequent Itemsets (Sup > 50%):

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

With $sup_{min} = 50\%$

(# > 2.5)

Association Rules: Concepts



Format: **Antecedent → Consequent**

"When the antecedent is observed, the consequent should also (probably) be observed"

Example:

{A, B} → {C, D}

"When items A and B are observed, items C and D should also (probably) be observed"

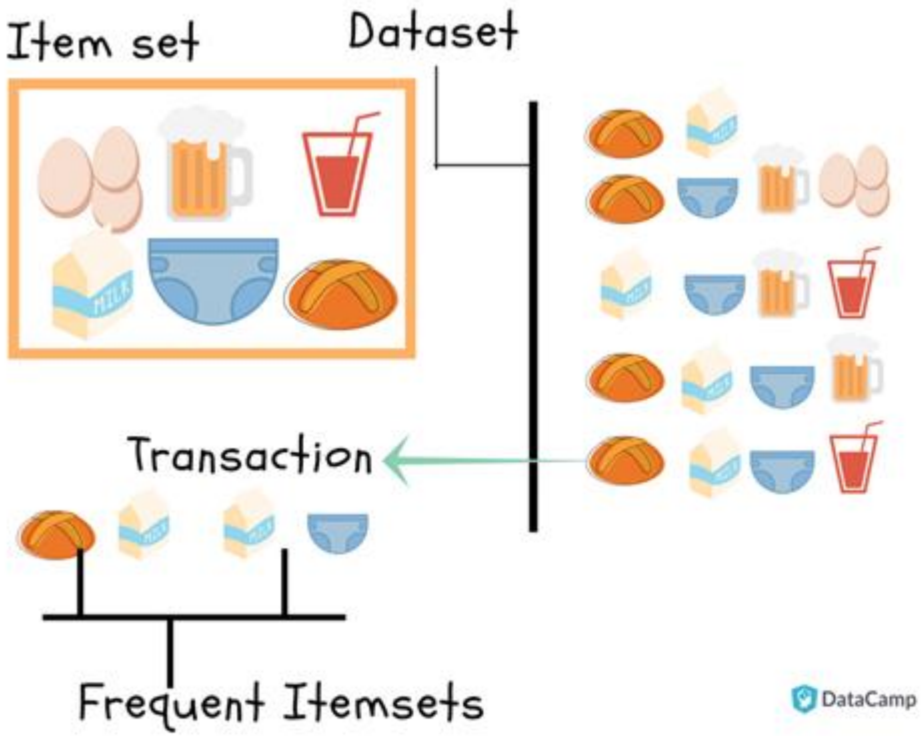
Support: percentage of transactions where co-occurrence is observed

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

Confidence: percentage of transactions in which the occurrence of the antecedent correctly predicts the occurrence of the consequent

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

Mining Association Rules : Definition



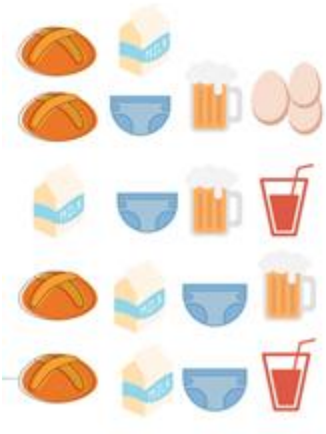
Given away:

- A set of transactions $D = \{t_1, t_2, \ldots, t_n\}$
- a minimum support $sup_{min} \in [0,1]$
- a minimal trust $conf_{min} \in [0,1]$

Find all rules A →C such that:

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

Mining Association Rules : Example



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}
```

Rule	Ant	cons	freq(Ant)	Sup(Ant)	freq(cons)	Sup(Cons)	freq(Ant→Cons)	Sup(Ant→Cons)	Conf(Ant→Cons)
C→F	W	F	3	60%	4	80%	3	60%	100%
F→C	F	W	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%

with $sup_{min} = 50\%$

And $conf_{min} = 90\%$

DEPARTAMENTO CIÊNCIA E TECNOLOGIA

Association Rules: Exercise

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%

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

In Python :

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

from apyori import a priori

rules = apriori (records, min_support =0.5, min_confidence
=0.9)
listrules = list (rules)
for item in listrules :
    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(" ===================================== ")
```

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


DEPARTAMENTO CIÊNCIA E TECNOLOGIA

Association Rules Evaluation: Interest

For an association rule to be interesting, it has to be:

- Unexpected (deviate from expected)
- Useful (with expected benefit)

Example: at a gas station, {newspaper} → {fuel} not unexpected or useful

Generally, an A C rule is interesting if A and C are **not** statistically independent.

- A and C are statistically independent if:
    - $Support(A \cup C) \approx Support(C) \times Support(C)$
    - $Confidence(A \rightarrow C) \approx Confidence(\emptyset \rightarrow C)$

Evaluation of Association Rules

Lift: measures the importance of a rule:

- Lift > 1: the antecedent and consequent appear together more often than expected
  - the occurrence of the antecedent has a positive effect on the occurrence of the consequent.
- Lift < 1: the antecedent and consequent appear less often together than expected
  - the occurrence of the antecedent has a negative effect on the occurrence of the consequent.
- lift ≈1 the antecedent and consequent appear almost as often together as expected
  - the occurrence of the antecedent has almost no effect on the occurrence of the consequent.

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

Conviction : measures “implication” (frequency at which the antecedent occurs without the consequent)

- Conviction ≈1 if the antecedent and consequent are unrelated

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

Leverage : measures the proportion of additional elements covered by the rule (versus what would be expected if they were independent)

- Leverage ≈0 if they are independent

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

Evaluation of Association Rules

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)	$\frac{P(A,B)P(\overline{A},\overline{B})}{P(A,\overline{B})P(\overline{A},B)}$
4	Yule's Q	$\frac{P(A,B)P(\overline{AB})-P(A,\overline{B})P(\overline{A},B)}{P(A,B)P(\overline{AB})+P(A,\overline{B})P(\overline{A},B)} = \frac{\alpha-1}{\alpha+1}$
5	Yule's Y	$\frac{\sqrt{P(A,B)P(\overline{AB})}-\sqrt{P(A,\overline{B})P(\overline{A},B)}}{\sqrt{P(A,B)P(\overline{AB})}+\sqrt{P(A,\overline{B})P(\overline{A},B)}} = \frac{\sqrt{\alpha}-1}{\sqrt{\alpha}+1}$
6	Kappa (κ)	$\frac{P(A,B)+P(\overline{A},\overline{B})-P(A)P(B)-P(\overline{A})P(\overline{B})}{1-P(A)P(B)-P(\overline{A})P(\overline{B})}$
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)	$\max\Big(P(A,B)\log(\frac{P(B\|A)}{P(B)}) + P(A\overline{B})\log(\frac{P(\overline{B}\|A)}{P(\overline{B})}),$ $P(A,B)\log(\frac{P(A\|B)}{P(A)}) + P(\overline{A}B)\log(\frac{P(\overline{A}\|B)}{P(\overline{A})})\Big)$
9	Gini index (G)	$\max\Big(P(A)[P(B\|A)^2+P(\overline{B}\|A)^2]+P(\overline{A})[P(B\|\overline{A})^2+P(\overline{B}\|\overline{A})^2]$ $-P(B)^2-P(\overline{B})^2,$ $P(B)[P(A\|B)^2+P(\overline{A}\|B)^2]+P(\overline{B})[P(A\|\overline{B})^2+P(\overline{A}\|\overline{B})^2]$ $-P(A)^2-P(\overline{A})^2\Big)$
10	Support (s)	$P(A,B)$
11	Confidence (c)	$\max(P(B\|A), P(A\|B))$
12	Laplace (L)	$\max\left(\frac{NP(A,B)+1}{NP(A)+2}, \frac{NP(A,B)+1}{NP(B)+2}\right)$
13	Conviction (V)	$\max\left(\frac{P(A)P(\overline{B})}{P(A\overline{B})}, \frac{P(B)P(\overline{A})}{P(B\overline{A})}\right)$
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)	$\max\left(\frac{P(B\|A)-P(B)}{1-P(B)}, \frac{P(A\|B)-P(A)}{1-P(A)}\right)$
18	Added Value (AV)	$\max(P(B\|A)-P(B), P(A\|B)-P(A))$
19	Collective strength (S)	$\frac{P(A,B)+P(\overline{AB})}{P(A)P(B)+P(\overline{A})P(\overline{B})}\times\frac{1-P(A)P(B)-P(\overline{A})P(\overline{B})}{1-P(A,B)-P(\overline{AB})}$
20	Jaccard (ζ)	$\frac{P(A,B)}{P(A)+P(B)-P(A,B)}$
21	Klosgen (K)	$\sqrt{P(A,B)}\max(P(B\|A)-P(B), P(A\|B)-P(A))$

Tan, PN, 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

APRIORI Algorithm

Apriori(T, ϵ)

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

$k \leftarrow 2$

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 $c \in D_t$

$count[c] \leftarrow count[c] + 1$

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

$k \leftarrow k + 1$

return $\bigcup_k L_k$

Min. support: 2 transactions

Database D

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

Scan D

C_1

itemset	sup.
{1}	2
{2}	3
{3}	3
{4}	1
{5}	3

L_1

itemset	sup.
{1}	2
{2}	3
{3}	3
{5}	3

C_2

itemset
{1 2}
{1 3}
{1 5}
{2 3}
{2 5}
{3 5}

Scan D

C_2

itemset	sup
{1 2}	1
{1 3}	2
{1 5}	1
{2 3}	2
{2 5}	3
{3 5}	2

L_2

itemset	sup
{1 3}	2
{2 3}	2
{2 5}	3
{3 5}	2

C_3

itemset
{2 3 5}

Scan D

L_3

itemset	sup
{2 3 5}	2

T: Set of transactions

ϵ: Minimum trust

L: List of itemsets

k: Size of itemsets

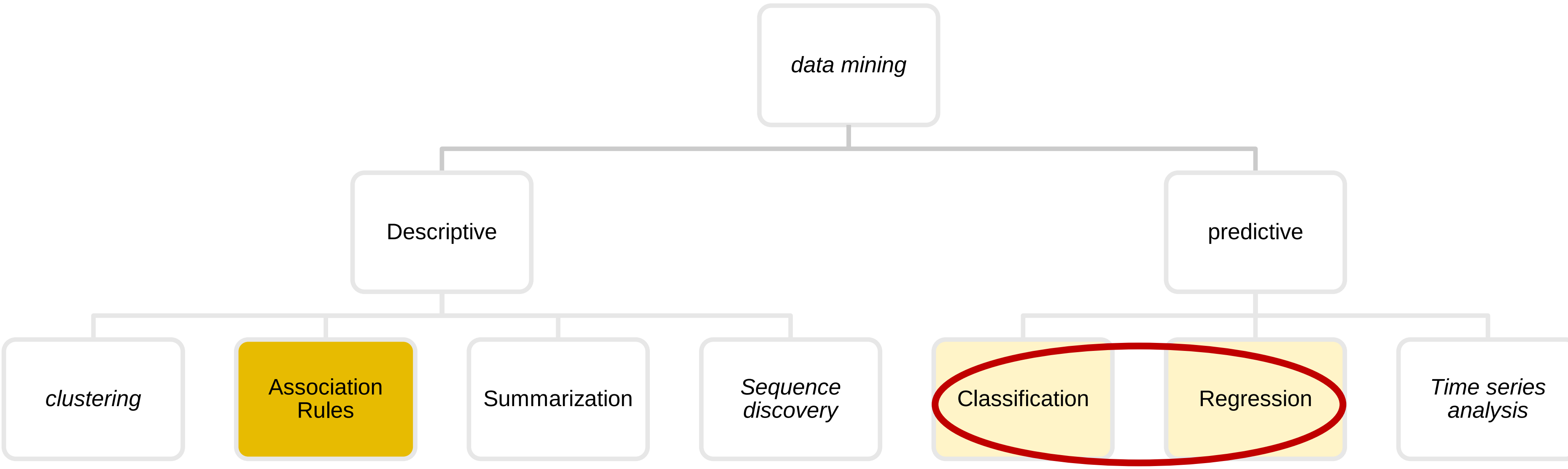
C: Candidate Itemsets

a, b: sets of items

D: itemsets contained in T

DEPARTAMENTO CIÊNCIA E TECNOLOGIA

this chapter

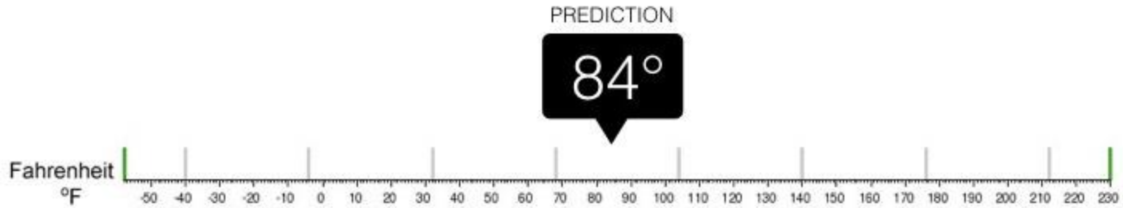


DEPARTAMENTO CIÊNCIA E TECNOLOGIA

Classification or regression?

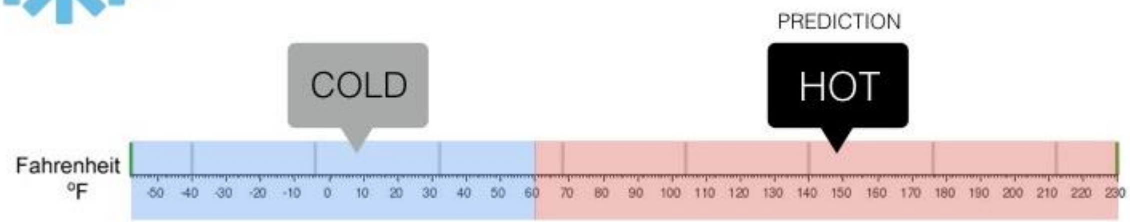
Regression

What is the temperature going to be tomorrow?

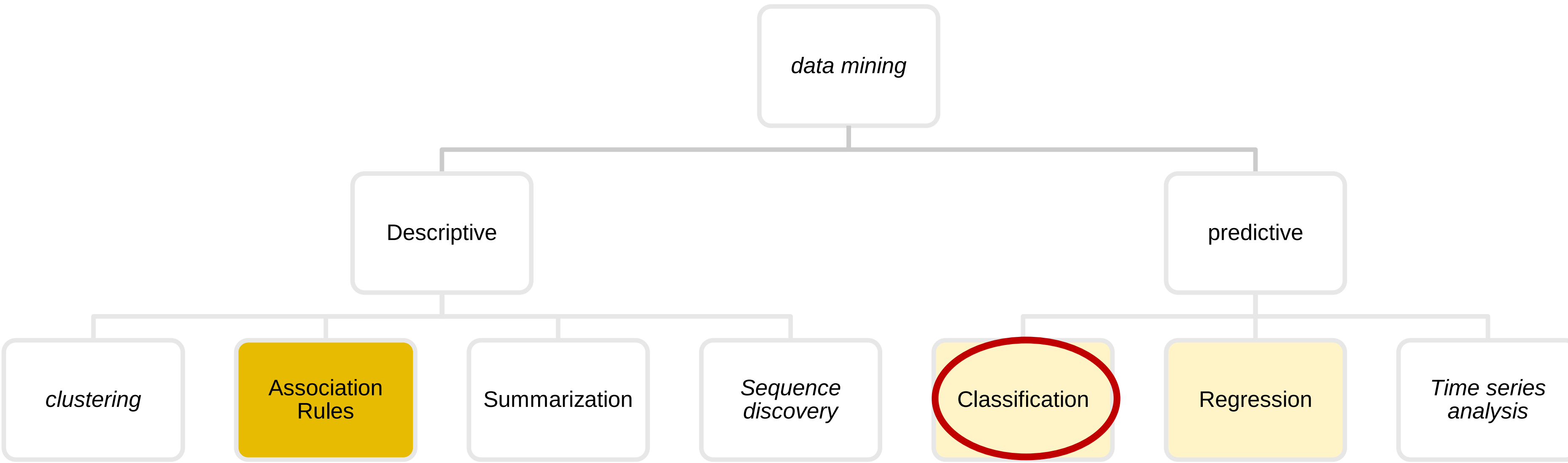


Classification

Will it be Cold or Hot tomorrow?



this chapter



Classification

Given the results of the last campaign, which we know (historical data)

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	4	0	0

dependent variable
objective variable

Independent variables

We want to predict potential adherents (Comprou=sim) (Bought=yes)

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

Idade: age, Rendimento:salary, Af.fam=Nr persons in house, Vendas anteriores: previsou sales, Ultima venda: last sale

Classification Algorithms

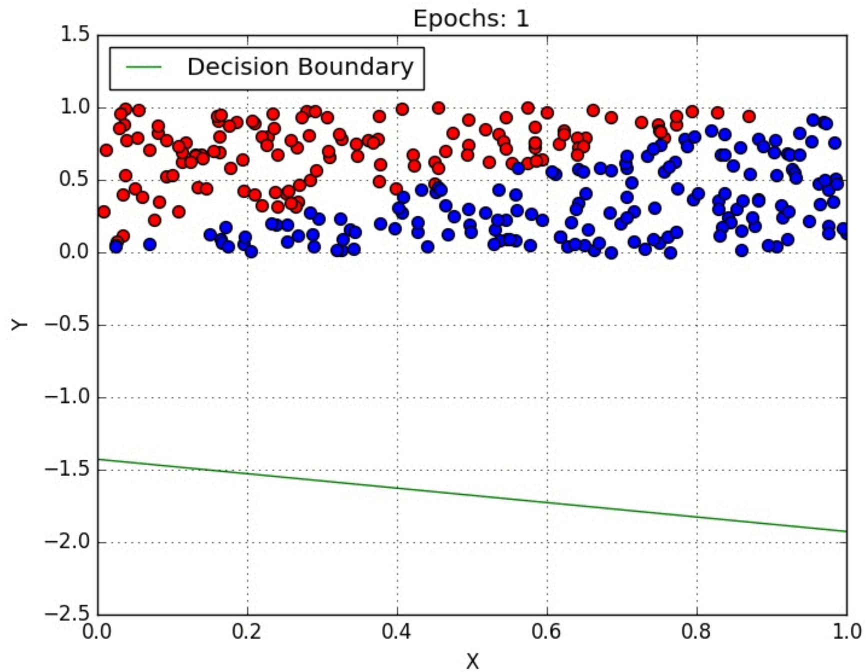
- Logistic Regression
- Naive Bayes Classifier
- K-Nearest Neighbors
- decision tree
    - random forest
- Support Vector Machines
- …

DEPARTAMENTO CIÊNCIA E TECNOLOGIA

logistic regression classifier

Calculates: $P(Y=1|X)$ or $P(Y=0|X)$

The probability that the dependent variable (Y) has a given value, given the values of the independent variables (X)

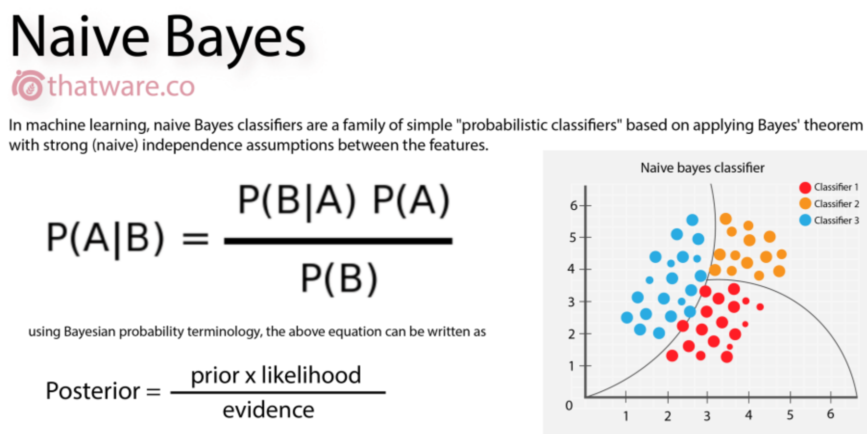


https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148

Naive bayes classifier

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

The probability that the dependent variable (Y) has a given value, given the values of the independent variables (X)

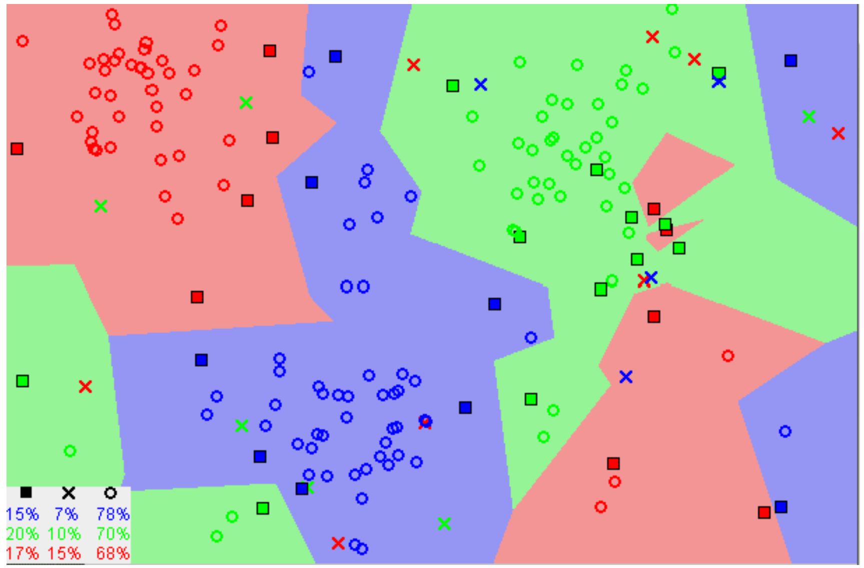


https://towardsdatascience.com/introduction-to-na%C3%AFve-bayes-classifier-fa59e3e24aaf

k nearest neighbors classifier

Calculates the distance between elements.

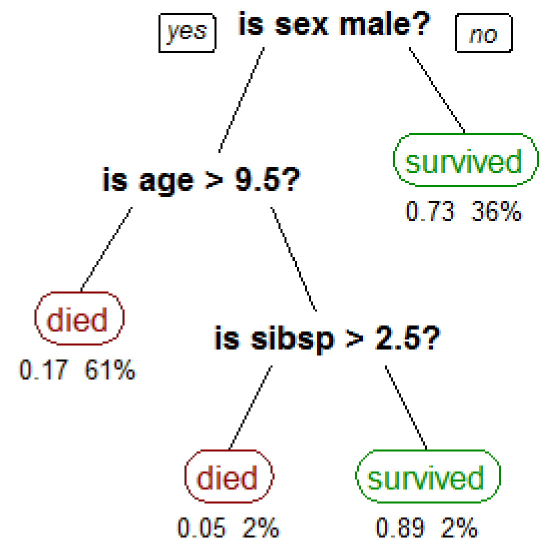
It assumes that similar elements are close to each other.



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

Decision tree classifier

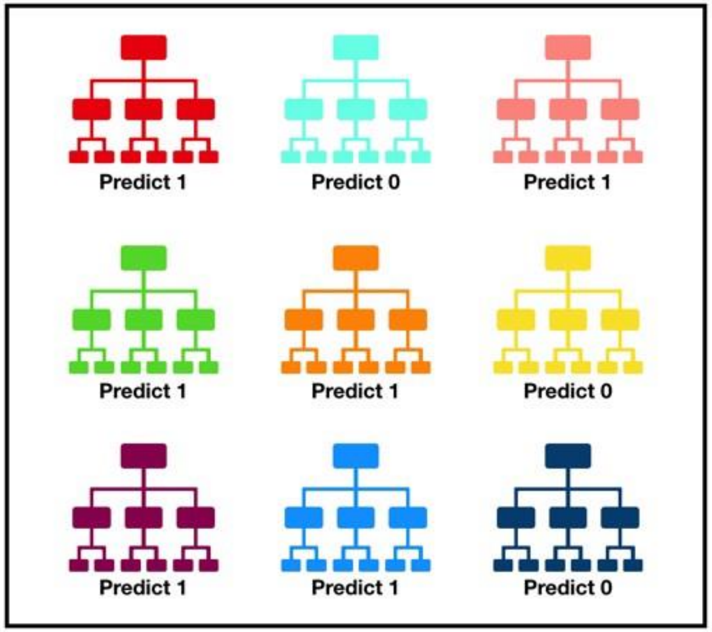
Constructs a representation of a decision table in the form of a tree



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

Random forest classifier

Constructs a set of decision trees (decision trees)



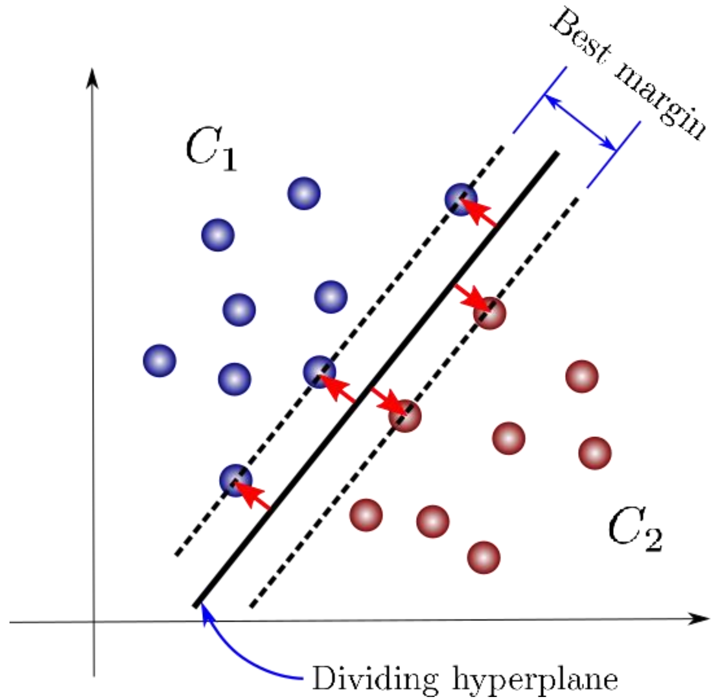
Tally: Six 1s and Three 0s
Prediction: 1

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

DEPARTAMENTO CIÊNCIA E TECNOLOGIA

Support vector Machine classifier

Constructs a representation of the examples as points in space, mapped so that examples from each category are as far apart as possible

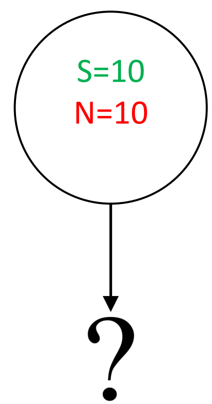


https://towardsdatascience.com/support-vector-machines-for-classification-fc7c1565e3

Decision tree building algorithm

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

Tree root:



Where are we going to do the *split*

split with the lowest cost is chosen (*greedy algorithm*)

Recursively (formed groups can be subdivided again)

DEPARTAMENTO CIÊNCIA E TECNOLOGIA

Decision tree evaluation measures

Gini Index: measures the degree or probability that a given variable is misclassified when it is chosen at random

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

p_i is the probability that an observation is classified in a particular class

DEPARTAMENTO CIÊNCIA E TECNOLOGIA

Algorithm for building decision trees (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

Tree root:

S=10
N=10

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

Where are we going to do the *split*?

Algorithm for building decision trees (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 = a
- **V** p=7/20: **S** p=3/7, **N** p=4/7 — $G = 0{,}49$
- **F** p=13/20: **S** p=7/13, **N** p=6/13 — $G = 0{,}497$

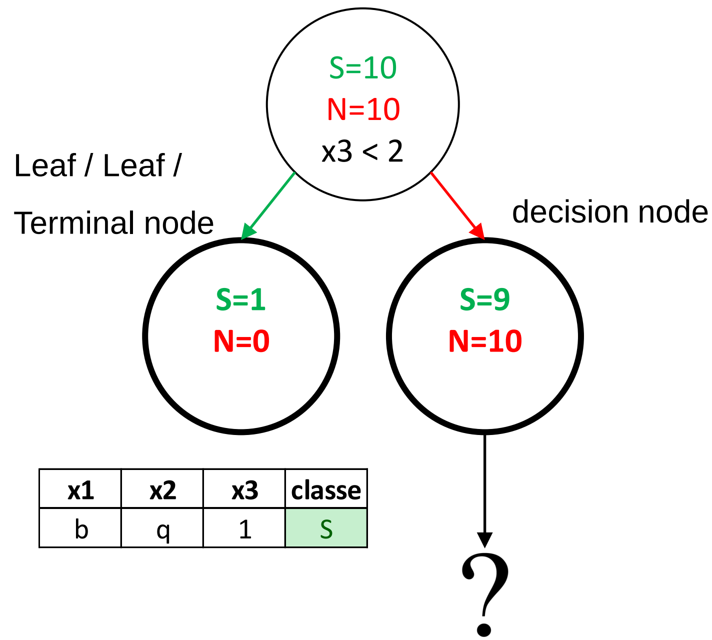
x1	x2	x3	classe
a	p	20	N
b	p	24	S
a	p	10	S
a	p	13	S
a	p	16	S
b	p	34	S
b	p	28	S
b	p	38	S
b	q	1	S
a	q	3	N
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	q	8	S
b	q	5	S
b	q	30	N
b	q	26	N
a	q	22	N
a	q	26	N

x2 = p
- **V** p=8/20: **S** p=7/8, **N** p=1/8 — $G = 0{,}219$
- **F** p=12/20: **S** p=3/12, **N** p=9/12 — $G = 0{,}375$

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

x3 < 2
- **V** p=1/20: **S** p=1/1, **N** p=0/1 — $G = 0$
- **F** p=19/20: **S** p=9/19, **N** p=10/19 — $G = 0{,}499$

...



x1	x2	x3	classe
b	q	1	S

Algorithm for building decision trees (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	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 = a | V $p=7/19$ | S $p=3/7$, N $p=4/7$ | $G = 0{,}49$
| F $p=12/19$ | S $p=6/12$, N $p=6/12$ | $G = 0{,}5$

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
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

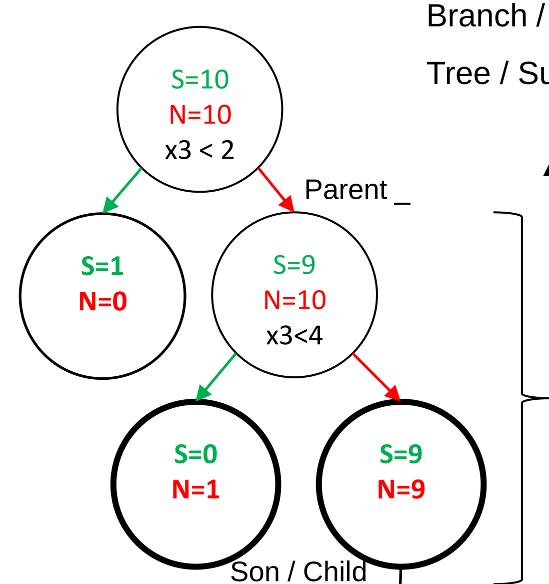
x2 = p | V $p=8/19$ | S $p=7/8$, N $p=1/8$ | $G = 0{,}219$
| F $p=11/19$ | S $p=2/11$, N $p=9/11$ | $G = 0{,}298$

x1	x2	x3	classe
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

$\boxed{\text{x3 < 4 | V } p=1/19 \text{ | S } p=0/1, \text{ N } p=1/1 \text{ | } G = 0}$
| F $p=18/19$ | S $p=9/18$, N $p=9/18$

...

Branch / Sub-
Tree / Sub - Tree



x1	x2	x3	classe
a	q	3	N

?

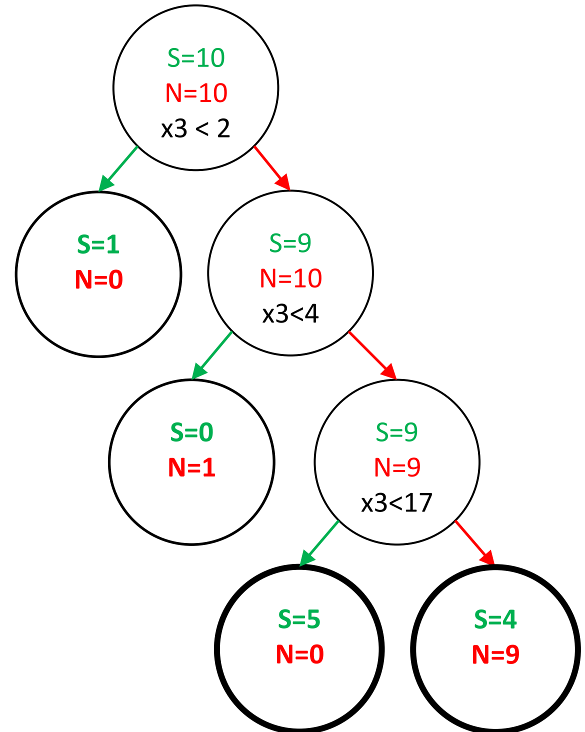
UPT DCT DEPARTAMENTO CIÊNCIA E TECNOLOGIA

Algorithm for building decision trees (Gini Index)

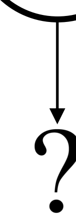
x1	x2	x3	classe
a	p	20	N
a	p	10	S
a	p	13	S
a	p	16	S
a	q	22	N
a	q	26	N
b	q	32	N
b	q	26	N
b	q	18	N
b	q	19	N
b	p	24	S
b	q	8	S
b	p	34	S
b	q	5	S
b	p	28	S
b	q	30	N
b	q	26	N
b	p	38	S

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



x1 = a
- **V** *p=6/18*: **S** *p=3/6*, **N** *p=3/6* — $G = 0{,}5$
- **F** *p=12/18*: **S** *p=6/12*, **N** *p=6/12* — $G = 0{,}5$

x2 = 9
- **V** *p=8/18*: **S** *p=7/8*, **N** *p=1/8* — $G = 0{,}498$
- **F** *p=10/18*: **S** *p=2/10*, **N** *p=8/10* — $G = 0{,}278$

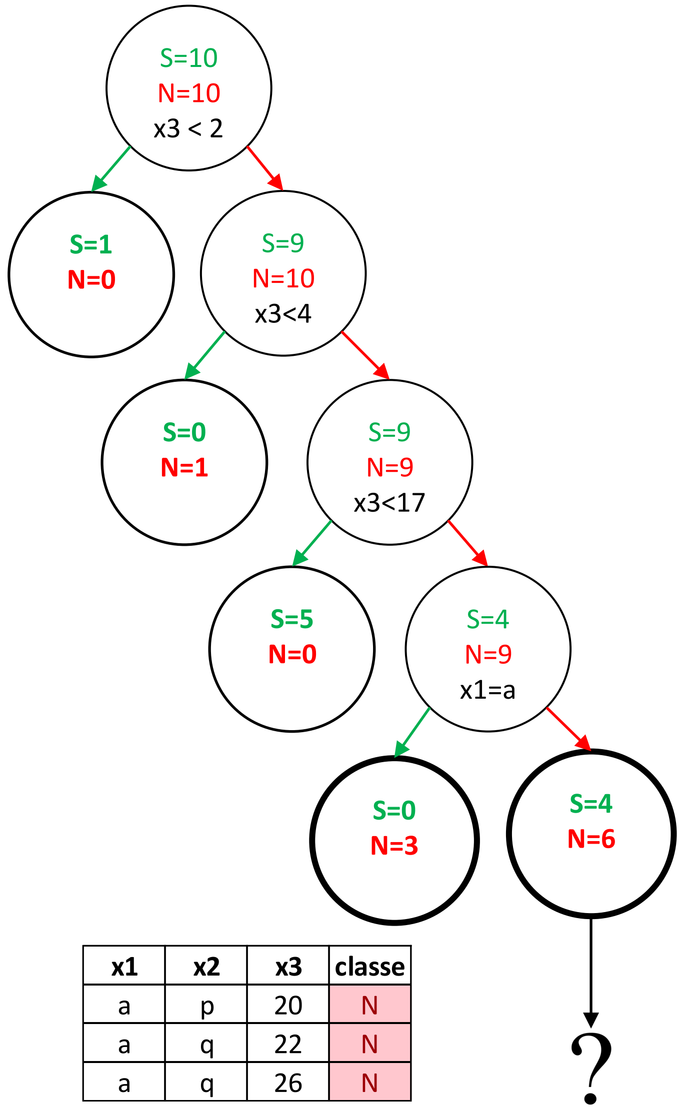
x3 < 17
- **V** *p=5/18*: **S** *p=5/5*, **N** *p=0/5* — $G = 0$
- **F** *p=13/18*: **S** *p=4/13*, **N** *p=9/13* — $G = 0{,}426$

Algorithm for building decision trees (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** p=3/13 | **S** p=0/3, **N** p=3/3 | $G = 0$

F p=10/13 | **S** p=4/10, **N** p=6/10

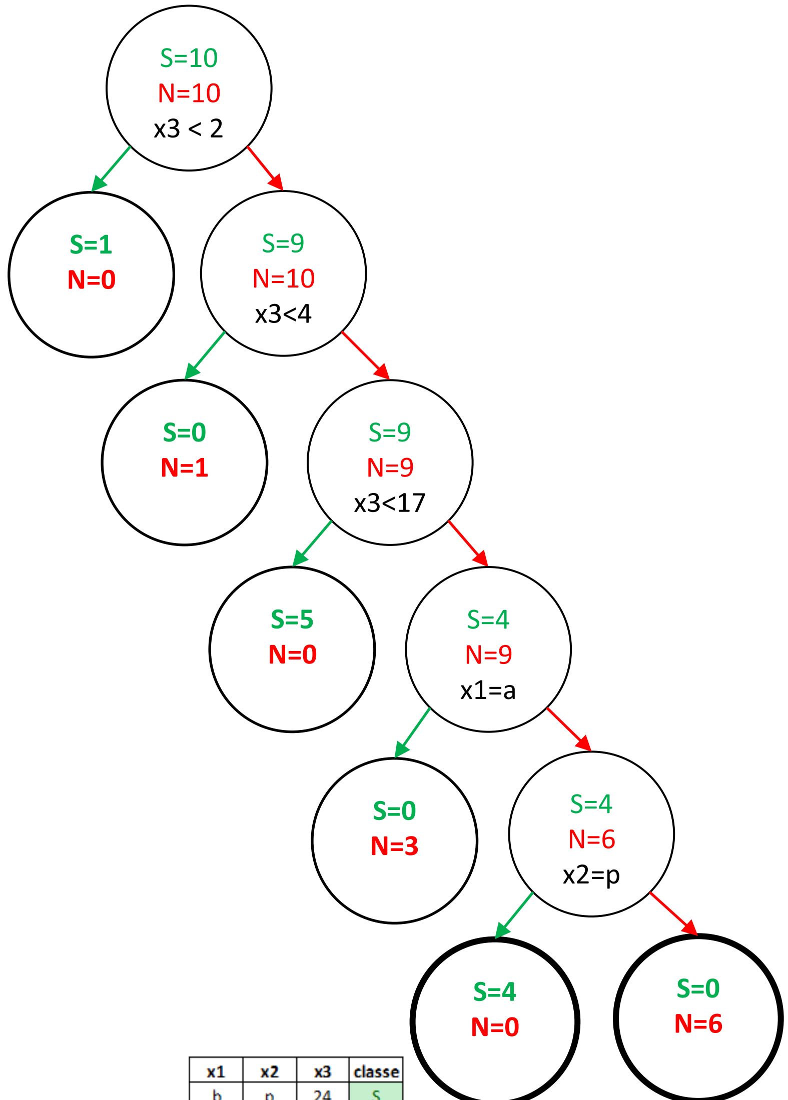


x1	x2	x3	classe
a	p	20	N
a	q	22	N
a	q	26	N

Algorithm for building decision trees (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

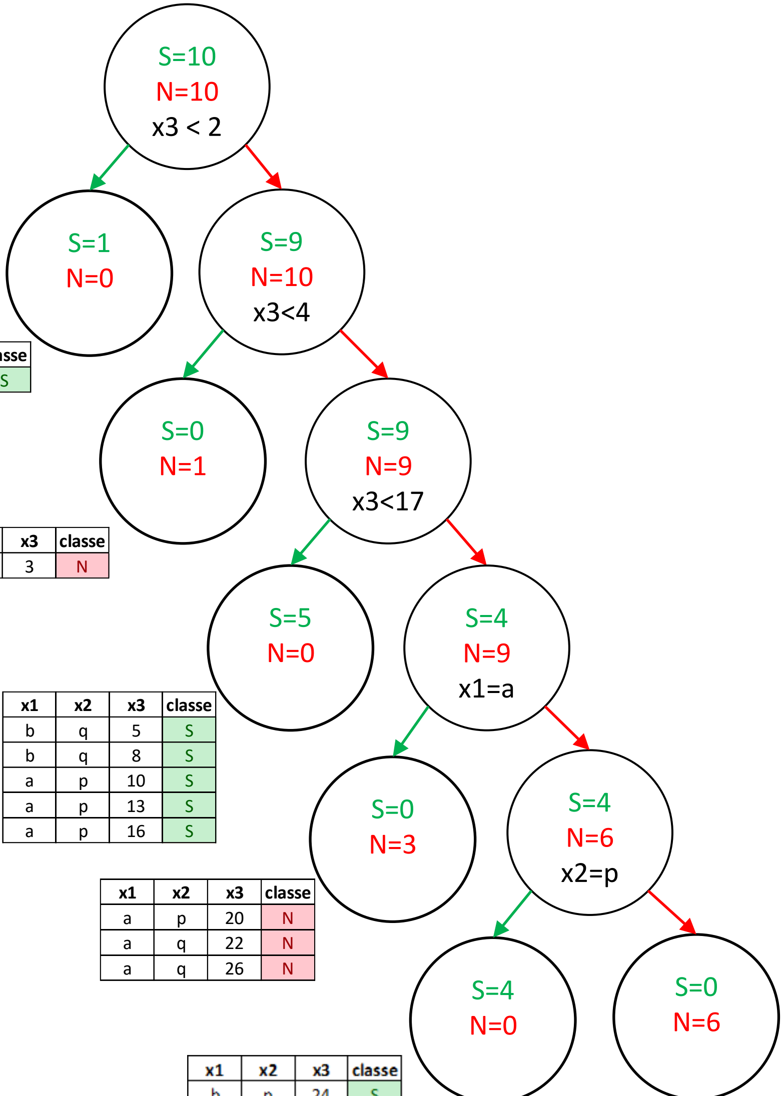
x2 = p	**V** *p=4/10*	**S**	*p=4/4*	$G = 0$
		N	*p=0/4*	
	F *p=6/10*	**S**	*p=0/6*	$G = 0$
		N	*p=6/6*	



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

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

Algorithm for building decision trees (Gini Index)



x1	x2	x3	classe
b	q	1	S

x1	x2	x3	classe
a	q	3	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

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

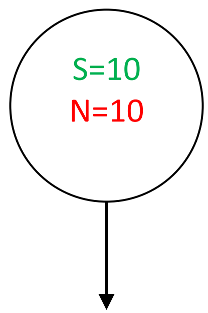
x1	x2	x3	classe
b	p	24	S
b	p	34	S
b	p	28	S
b	p	38	S

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

Decision tree building algorithm

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

Tree root:



Where are we going to do the *split*?

Decision tree evaluation measures

Entropy : is used as a way to measure the randomness of a column

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

p_i is the probability that an observation is classified in a particular class

Decision tree evaluation measures

Information Gain : measures the entropy reduction of a given *split*

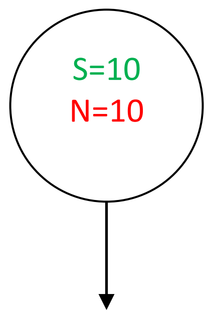
$$IG(T,A) = E(T) - \sum \frac{|T_v|}{T} \times E(T_v)$$

T is the target (the class)
A is the variable
v is every possible value of the variable

Algorithm for building decision trees (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

Tree root:



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

Where are we going to do the *split*?

Algorithm for building decision trees (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 = a
- V, p=7/20: S p=3/7; N p=4/7; $E = 0{,}985$
- F, p=13/20: S p=7/13; N p=6/13; $E = 0{,}996$

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

x2 = p
- V, p=8/20: S p=7/8; N p=1/8; $E = 0{,}544$
- F, p=12/20: S p=3/12; N p=9/12; $E = 0{,}811$

$IG = 0{,}295807$

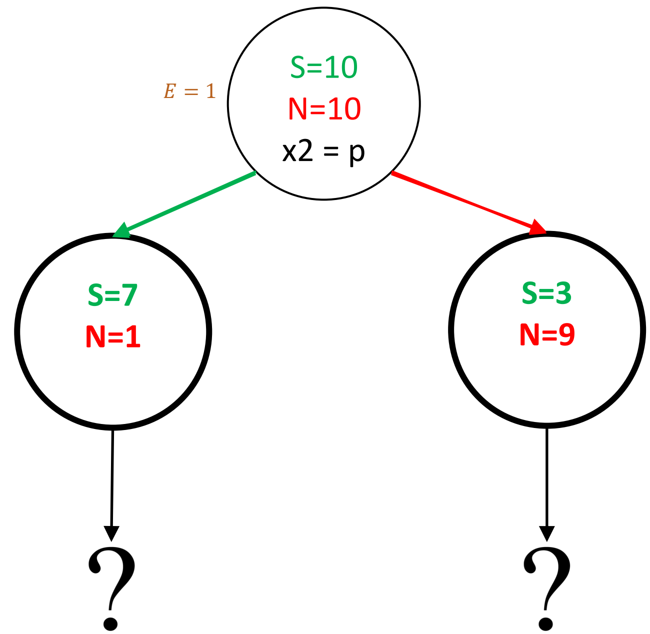
x1	x2	x3	classe
b	q	1	S
a	q	3	N
b	q	5	S
b	q	8	S
a	p	10	S
a	p	13	S
a	p	16	S
b	q	18	N
b	q	19	N
a	p	20	N
a	q	22	N
b	p	24	S
b	q	26	N
b	q	26	N
a	q	26	N
b	p	28	S
b	q	30	N
b	q	32	N
b	p	34	S
b	p	38	S

x3 < 2
- V, p=1/20: S p=1/1; N p=0/1; $G = 0$; $E = 0$
- F, p=19/20: S p=9/19; N p=10/19; $E = 0{,}998$

…

$IG = 0{,}051899$

$IG(outros) \ll 0{,}295807$



IMP.GE.190.0

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

DEPARTAMENTO CIÊNCIA E TECNOLOGIA

Algorithm for building decision trees (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** $p=4/8$
  - **S** $p=3/4$
  - **N** $p=1/4$ $E = 0{,}811$
- **F** $p=4/8$
  - **S** $p=4/4$
  - **N** $p=0/4$ $E = 0$

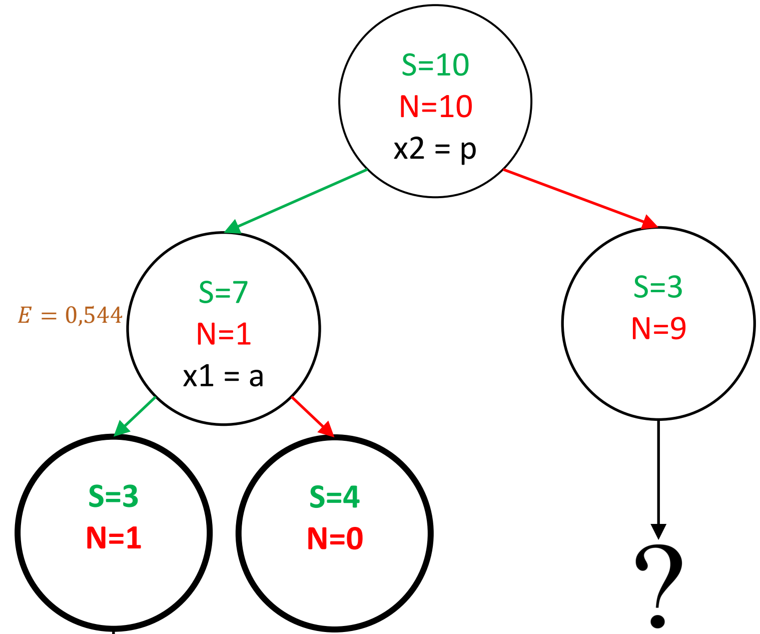
$IG = 0{,}138361$

x3 < 17
- **V** $p=3/8$
  - **S** $p=3/3$
  - **N** $p=0/3$ $E = 0$
- **F** $p=5/8$
  - **S** $p=4/5$
  - **N** $p=1/5$ $E = 0.722$

$IG = 0{,}092795$

x3 < 23
- **V** $p=4/8$
  - **S** $p=3/4$
  - **N** $p=1/4$ $E = 0{,}811$
- **F** $p=4/8$
  - **S** $p=4/4$
  - **N** $p=0/4$ $E = 0$

$IG = 0{,}138361$



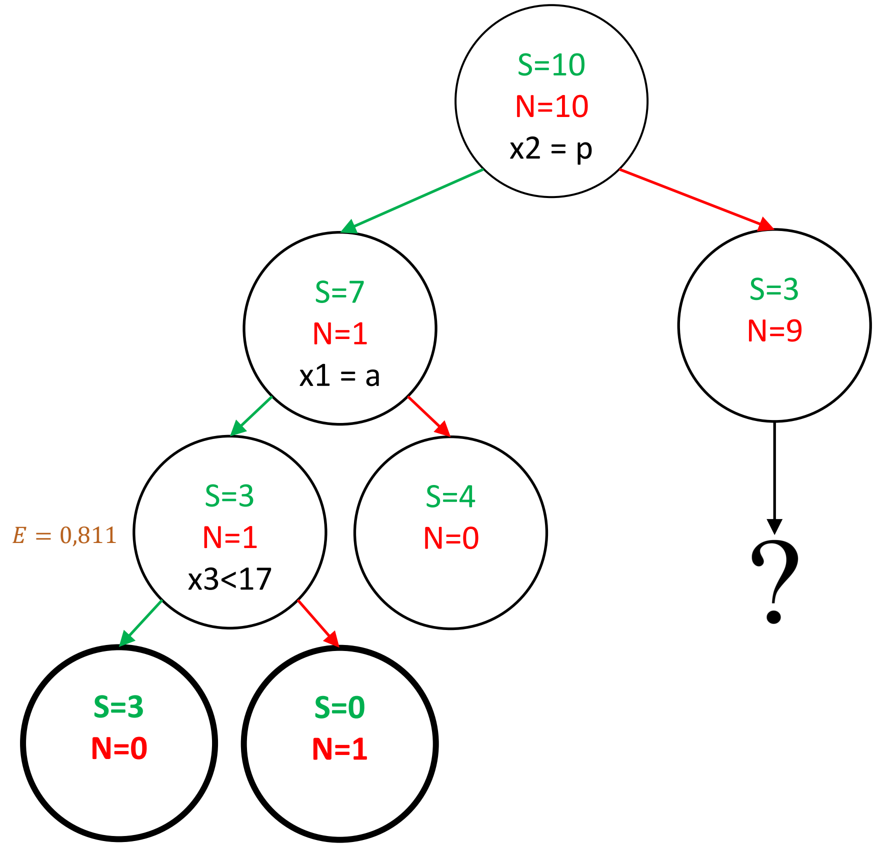
x1	x2	x3	classe
b	p	24	S
b	p	28	S
b	p	34	S
b	p	38	S

DEPARTAMENTO CIÊNCIA E TECNOLOGIA

Algorithm for building decision trees (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** *p=3/4*	**S**	*p=3/3*	
		N	*p=0/3*	$E = 0$
	F *p=1/4*	**S**	*p=0/1*	
		N	*p=1/1*	$E = 0$
				$IG = 0{,}811$



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

x1	x2	x3	classe
a	p	20	N

Algorithm for building decision trees (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 = a
- **V** $p=3/12$: **S** $p=0/3$, **N** $p=3/3$ — $E = 0$
- **F** $p=9/12$: **S** $p=3/9$, **N** $p=6/9$ — $E = 0{,}918$

$IG = 0{,}122278$

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

x3 < 2
- **V** $p=1/12$: **S** $p=1/1$, **N** $p=0/1$ — $E = 0$
- **F** $p=11/12$: **S** $p=2/11$, **N** $p=9/11$ — $E = 0{,}684$

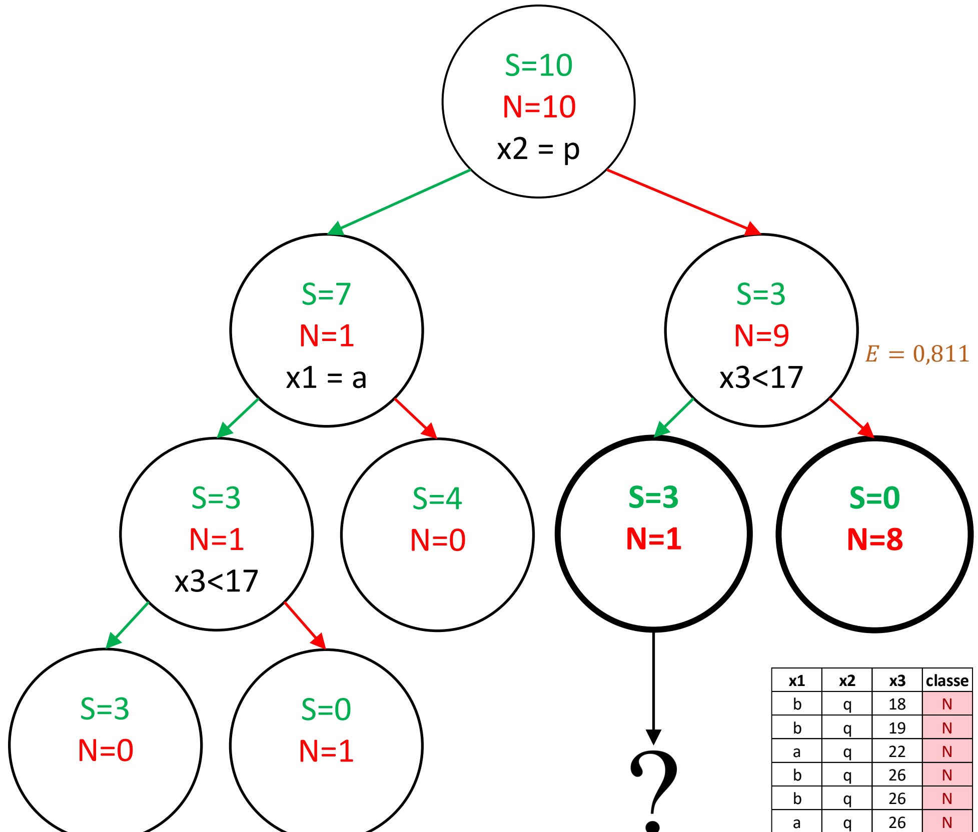
$IG = 0{,}183965$

x3 < 4
- **V** $p=2/12$: **S** $p=1/2$, **N** $p=1/2$ — $E = 1$
- **F** $p=10/12$: **S** $p=2/10$, **N** $p=8/10$ — $E = 0{,}722$

$IG = 0{,}042727$

x3 < 17
- **V** $p=4/12$: **S** $p=3/4$, **N** $p=1/4$ — $E = 0{,}811$
- **F** $p=8/12$: **S** $p=0/8$, **N** $p=8/8$ — $E = 0$

$IG = 0{,}540574$



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

UPT DCT DEPARTAMENTO CIÊNCIA E TECNOLOGIA

Algorithm for building decision trees (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** $p=1/4$: **S** $p=0/1$, **N** $p=1/0$, $E = 0$
- **F** $p=3/4$: **S** $p=3/3$, **N** $p=0/3$, $E = 0$

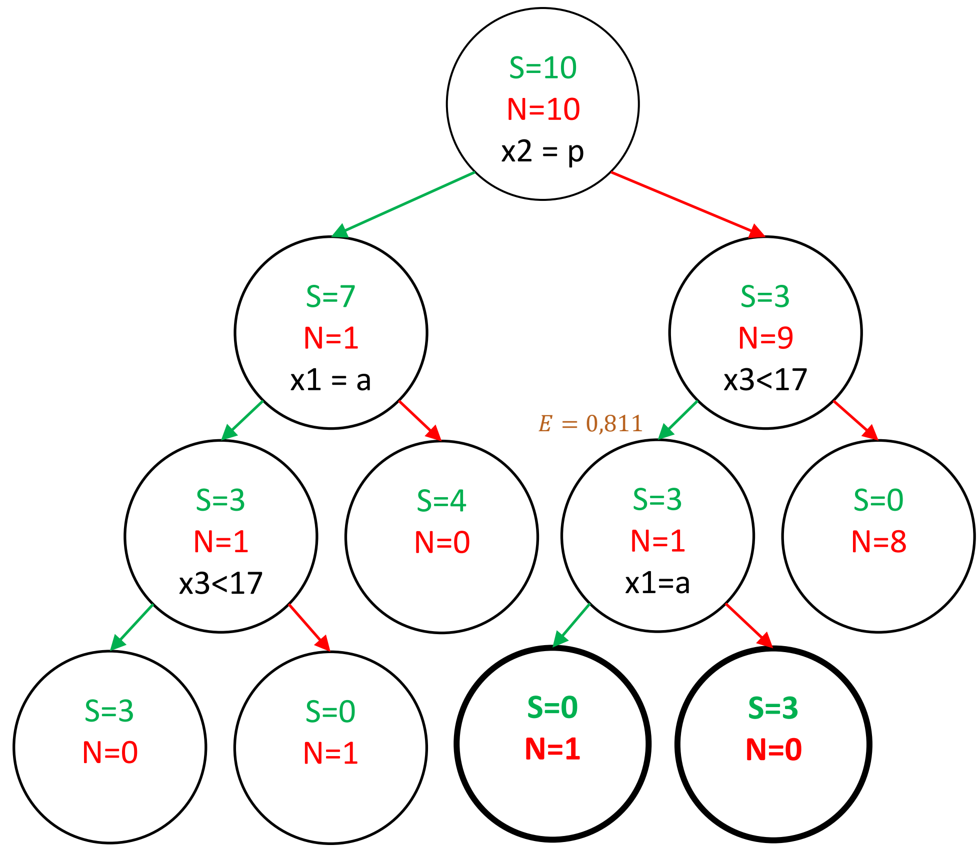
$IG = 0{,}811$

x3 < 2
- **V** $p=1/4$: **S** $p=1/1$, **N** $p=0/1$, $E = 0$
- **F** $p=3/4$: **S** $p=2/3$, **N** $p=1/3$, $E = 0{,}918$

$IG = 0{,}122278$

x3 < 4
- **V** $p=2/4$: **S** $p=1/2$, **N** $p=1/2$, $E = 1$
- **F** $p=2/4$: **S** $p=2/2$, **N** $p=0/2$, $E = 0$

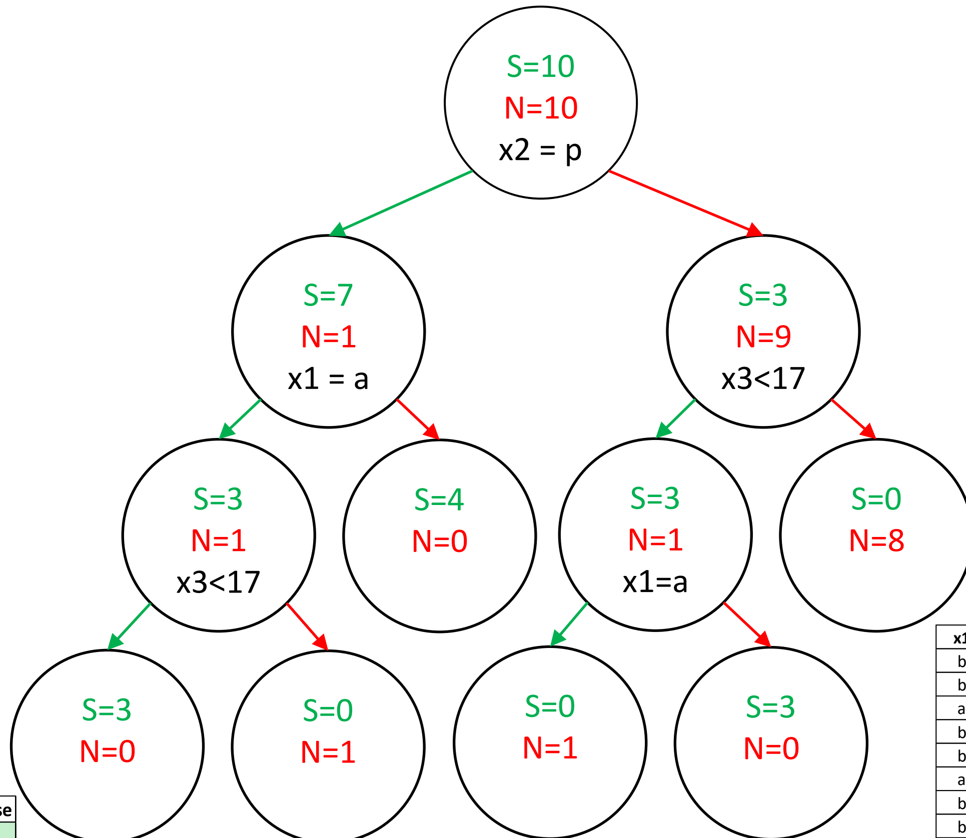
$IG = 0{,}311$



x1	x2	x3	classe
a	q	3	N

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

Algorithm for building decision trees (Information gain)



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

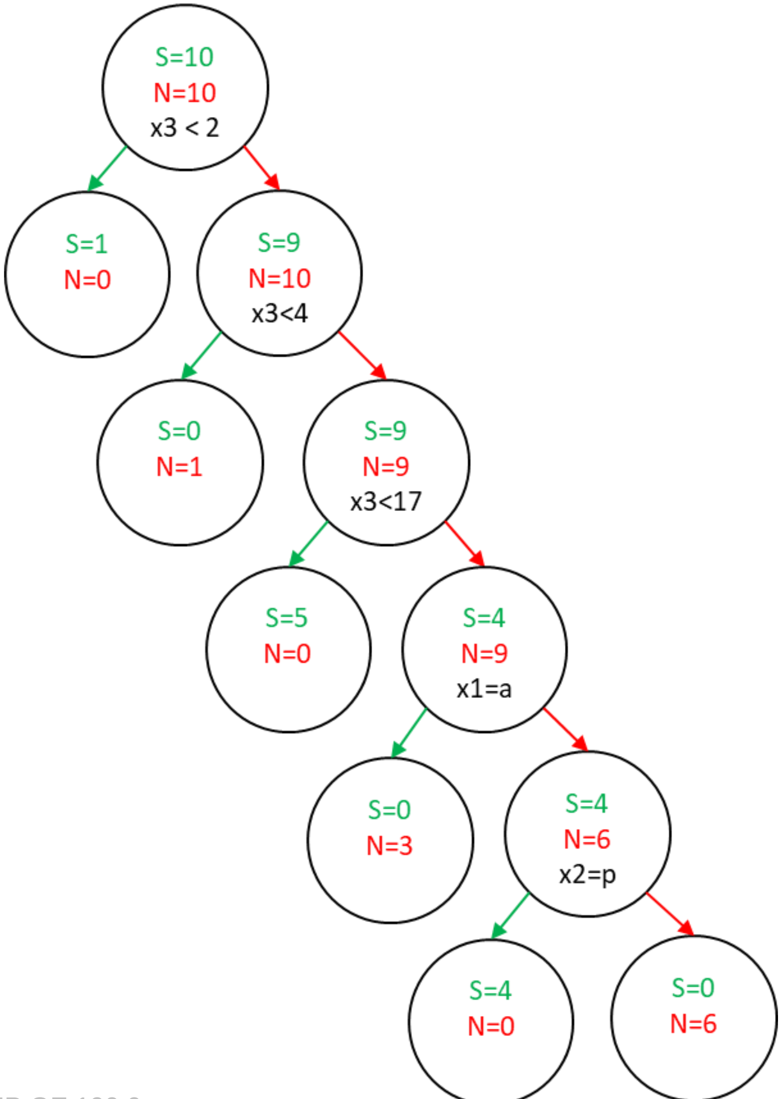
x1	x2	x3	classe
a	p	20	N

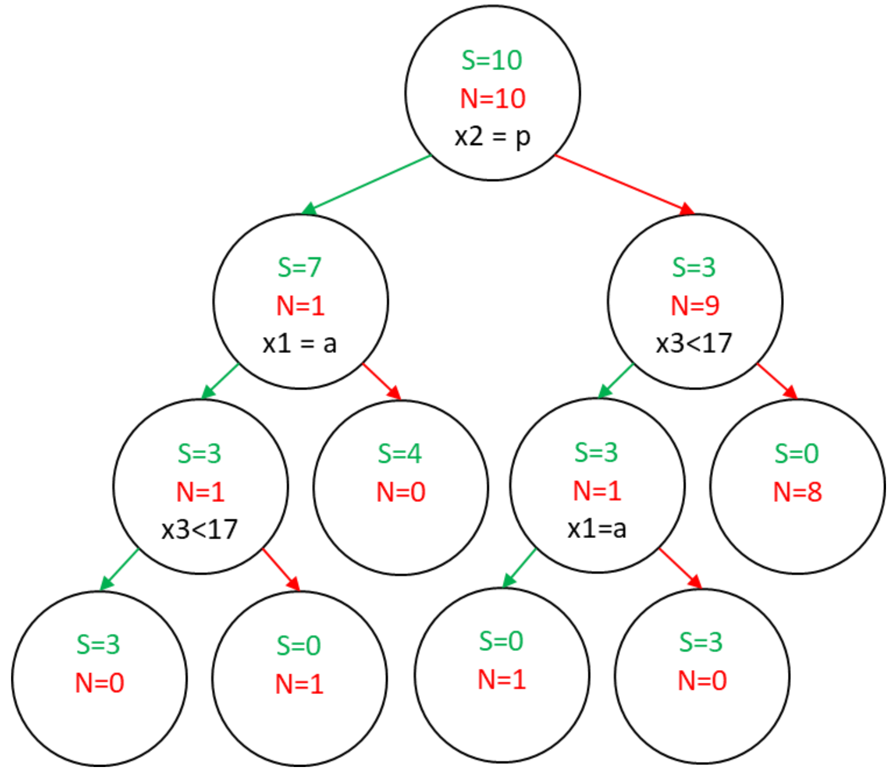
x1	x2	x3	classe
a	q	3	N

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

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

Algorithm for building decision trees (Gini Index vs. Information gain)





Decision tree algorithm: when to stop?

In the worst case, we can have a “leaf” for each example (*overfitting*)

How to avoid? Know when to stop:

- Setting a **minimum number of examples per sheet**
- Setting a **maximum depth for the tree**
- **Pruning** : remove branches that use low importance variables
    - Reduced error pruning : starts at a leaf and removes the nodes with the most popular class from that leaf, if it doesn't make the evaluation metric worse. Repeat for other sheets.
    - cost complexity pruning / weakest link pruning : a parameter (α) is used to determine whether a given node can be removed based on the size of the subtree .

More on *overfitting* :

Hawkins, MD (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

Evaluation of the Classification model

		Observação	
		P	N
Previsão	P	**TP**	**FP**
	N	**FN**	**TN**

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

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

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

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

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

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

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

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

Advantages and disadvantages of decision trees

Benefits	Disadvantages

• Easy to understand, interpret and visualize
• Execute *feature selection* implicitly
• Allow to use numeric and categorical data
• Can handle multi- *target data*
• They do not require much effort in the *data preparation process*
• Non-linear relationships do not affect tree performance
	• Can create overly complex trees that do not generalize (*overfitting*)
	• *Variance* : Slight variations in the data can cause a completely different tree to be created
	    • *Variance* can be reduced with methods like ***bagging*** (1) and ***boosting*** (1)
	• *Greedy* algorithms do not guarantee the creation of the optimal decision tree
	    • To overcome this, several trees can be created, in which *features* and samples are randomly selected: ***Random forest*** (1)
	• If there is a dominant class, the model can create a *biased* tree
	    • It is recommended to balance (2) the data before applying decision tree models

(1) Methods ***bagging*** , ***boosting*** and ***random forest*** will be seen further ahead during the semester

(2) Techniques for dealing with *imbalanced data* : https://www.analyticsvidhya.com/blog/2017/03/imbalanced-data-classification/

Decision trees for regression

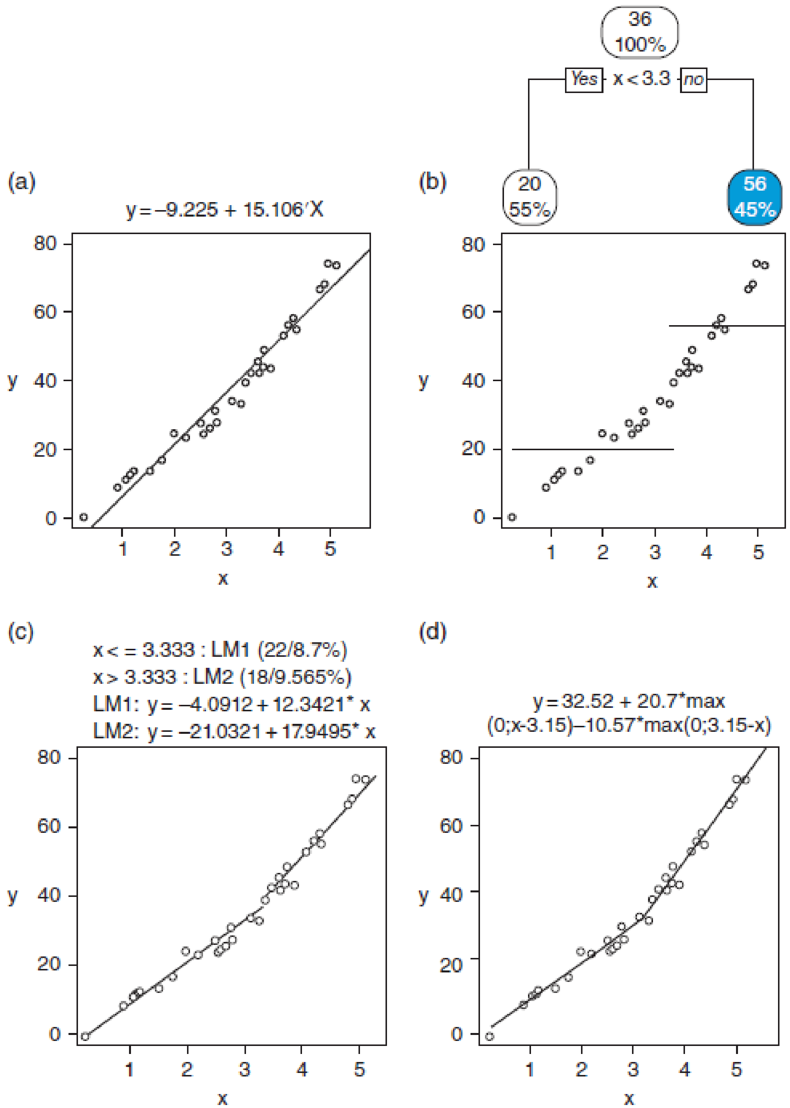
The principle and algorithm are the same

What changes:

- Metrics. For example:
  - Variance reduction
  - Reduction of Standard Deviation / Standard Deviation Reduction (see http://blog.saedsayad.com/decision_tree_reg.htm)
- How the sheets predict the values:
  - In classification, the majority is chosen, in regression, the average is usually chosen (CART)
  - model Trees and Multivariate Adaptive Regression Splines (MARS) use multivariate linear regression instead of mean

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

Decision trees for regression: comparing models



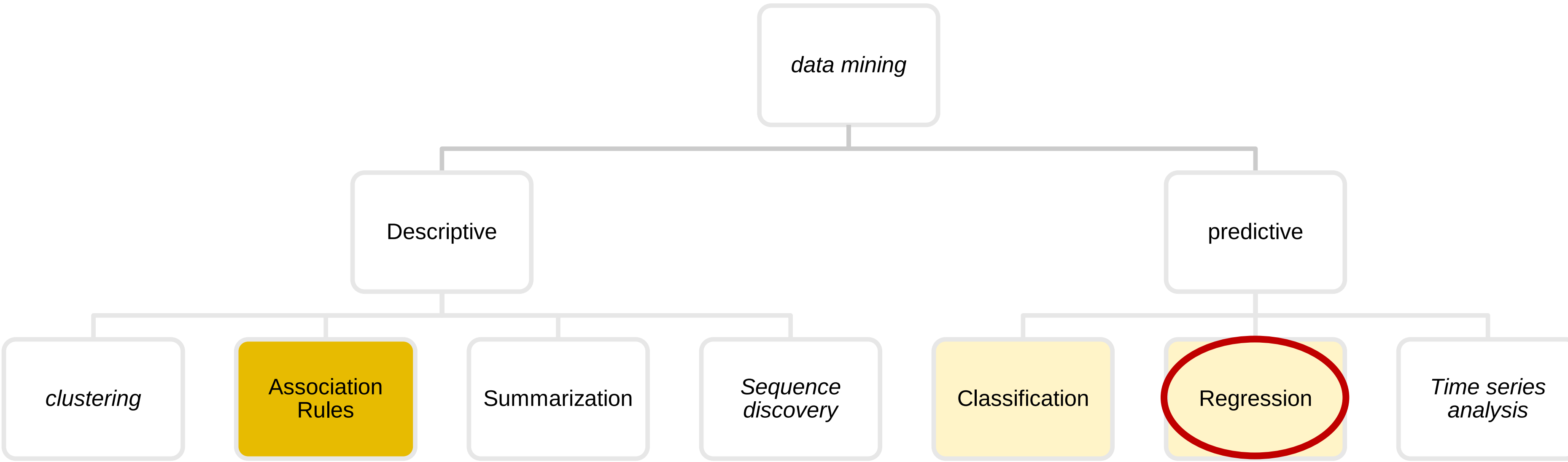
(a) MLR

(b) CART

(c) model trees

(d) MARS

this chapter



DEPARTAMENTO CIÊNCIA E TECNOLOGIA

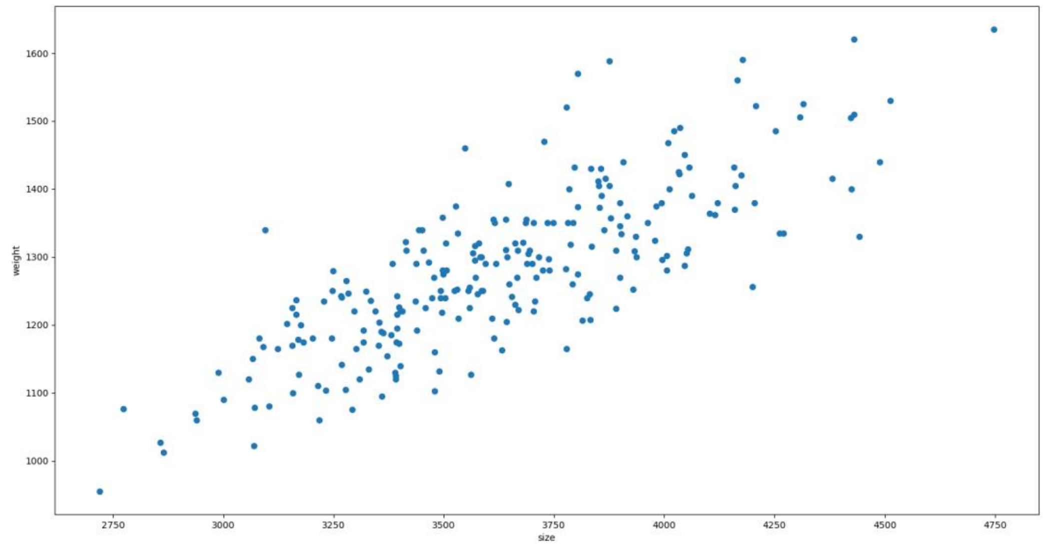
Notation

$\bar{x}$ Variable mean x

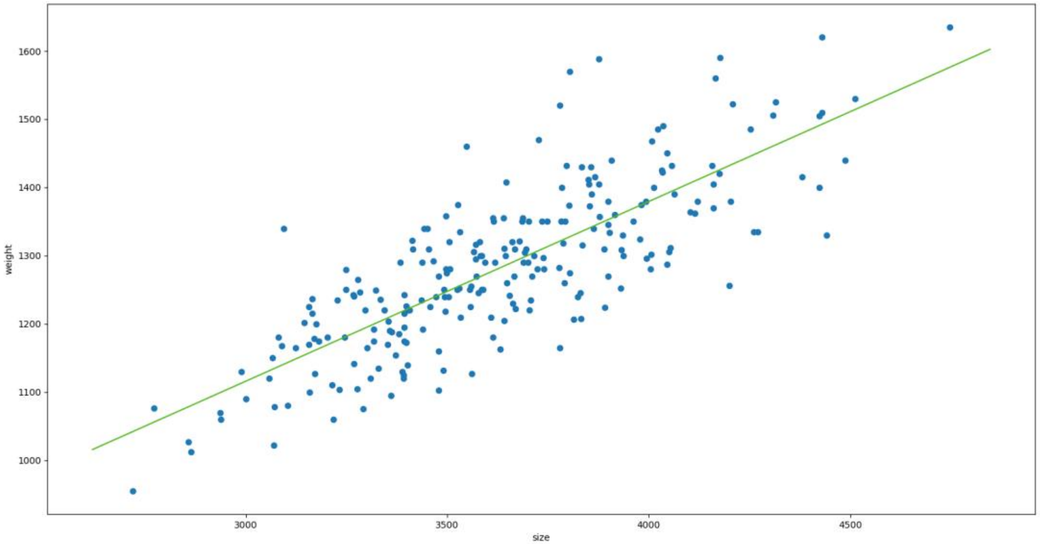
$\hat{x}$ Variable forecast x

Simple Linear Regression

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



Simple Linear Regression



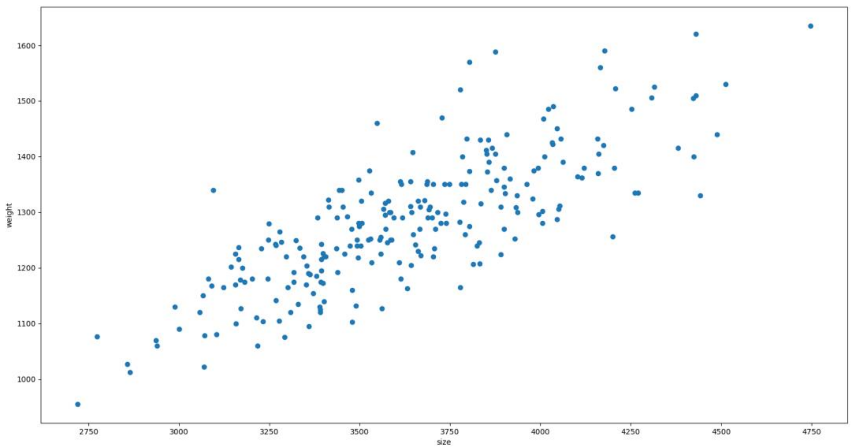
$$y = mx + b$$

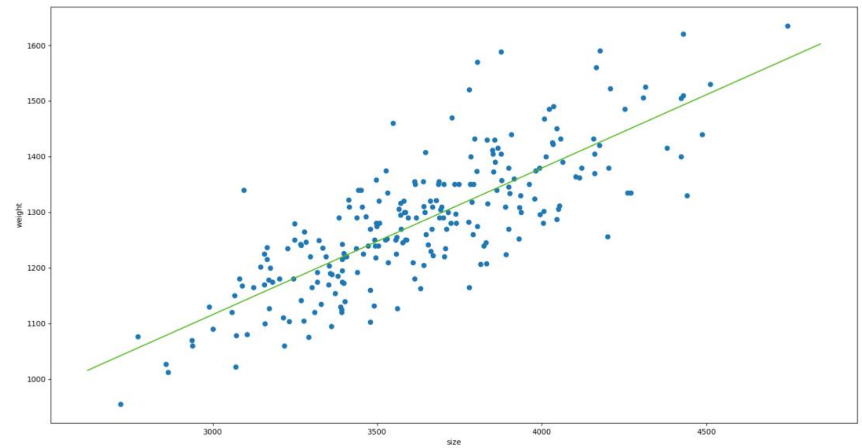
$$y = \beta_0 + \beta_1 x_1$$

$$weight = \underset{?}{\beta_0} \times size + \underset{?}{\beta_1}$$

DEPARTAMENTO **CIÊNCIA E TECNOLOGIA**

Determining the slope (m) and the ordinate at the origin (b)





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

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

$$\beta_1 = 0{,}26342933948939945$$

$$\beta_0 = 325{,}5734210494422 3$$

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

Evaluation of the Regression Model

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

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

It has the same unit of measurement as *y*

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

It has the same unit of measure as the square of *y.* Emphasizes the biggest mistakes more

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

It has the same unit of measurement as *y*

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

Compares predictive ability with that of trivial (average) prediction.

Possible values:

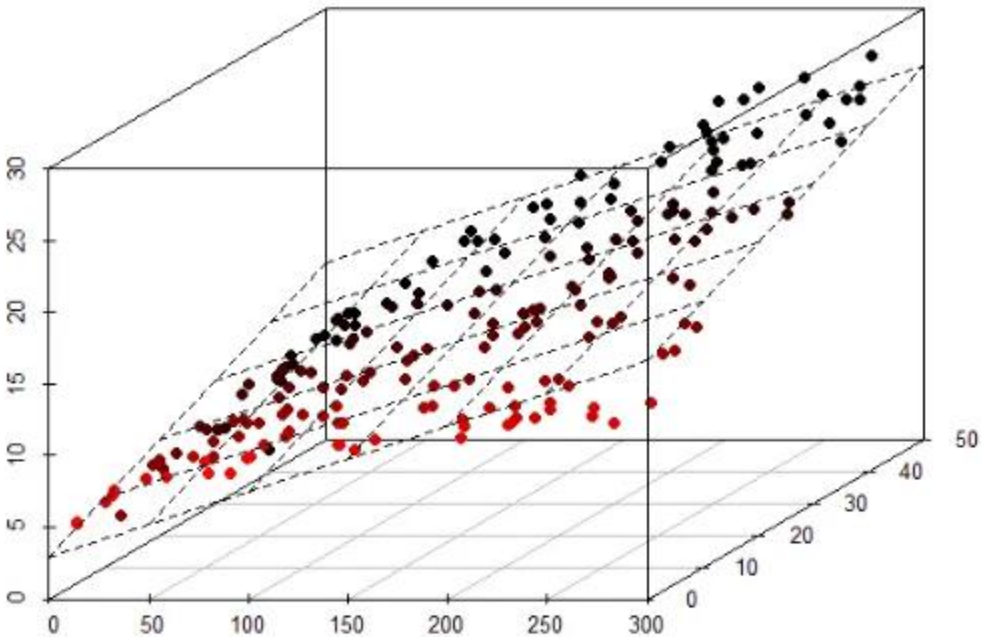
- **0:** perfect model
- **]0,1[:** useful template
- **1:** model as useful as predicting the mean
- **>1:** useless model (worse than predicting the mean)

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

Multiple Linear Regression

With two independent variables (x_1 x_2) one dependent (y):

Instead of a straight line, we will have a plane



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

Multiple Linear Regression

With *n* independent variables (x_1, x_2, …, x_n) one dependent (y):

- Difficult to visualise

- But we can generalize:

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

UNIVERSIDADE
PORTUCALENSE

Do conhecimento à prática.