EDIT.

Data Science & Business Analytics

Machine Learning Models

David Issá
davidribeiro.issa@gmail.com

1. Árvores de Decisão

1. Árvores de Decisão

- Algoritmos não paramétricos de supervised learning usados para classificação ou regressão.
- Uma das principais vantagens é o facto da solução do modelo representa um conjunto de regras, fáceis de interpretar.

1. Árvores de Decisão

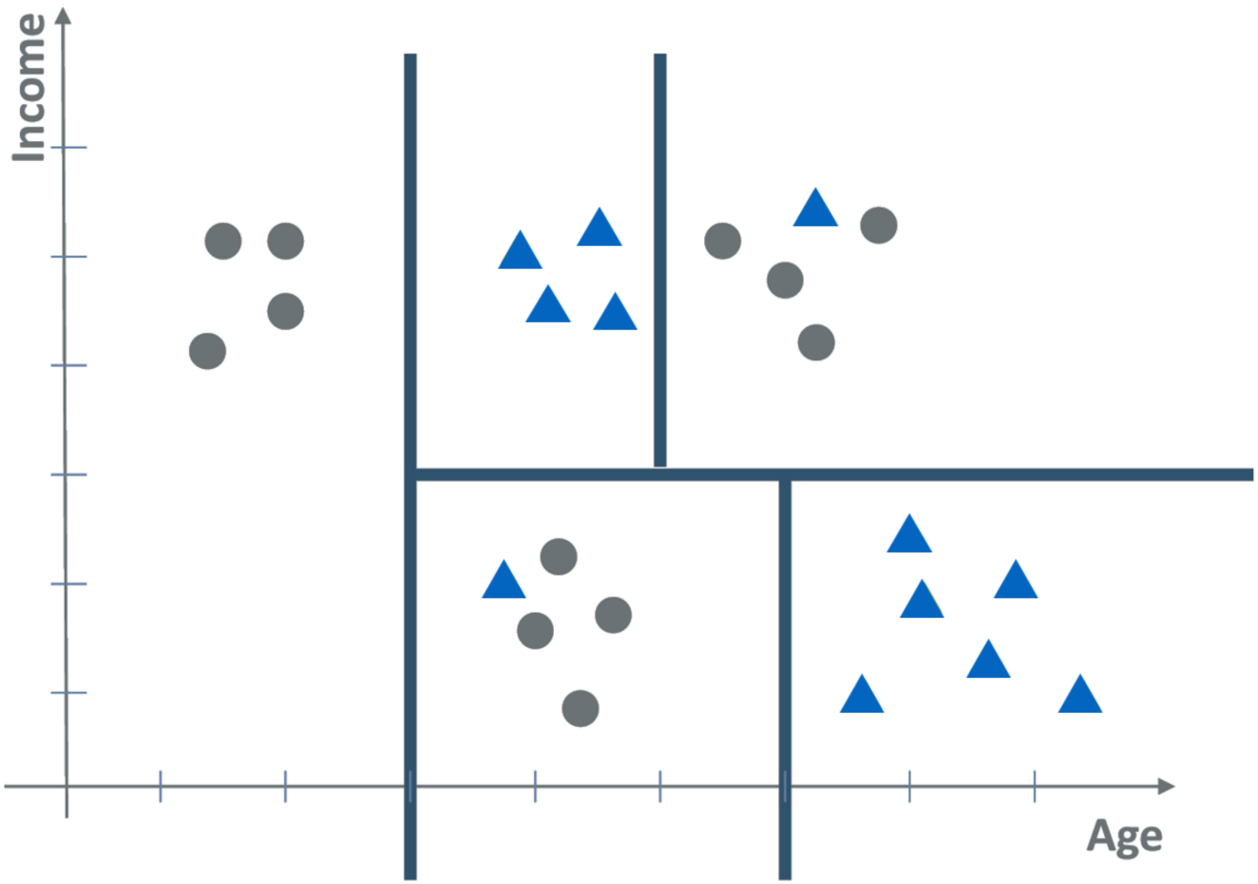
- Algoritmos não paramétricos de supervised learning usados para classificação ou regressão.

- Uma das principais vantagens é o facto da solução do modelo representa um conjunto de regras, fáceis de interpretar.

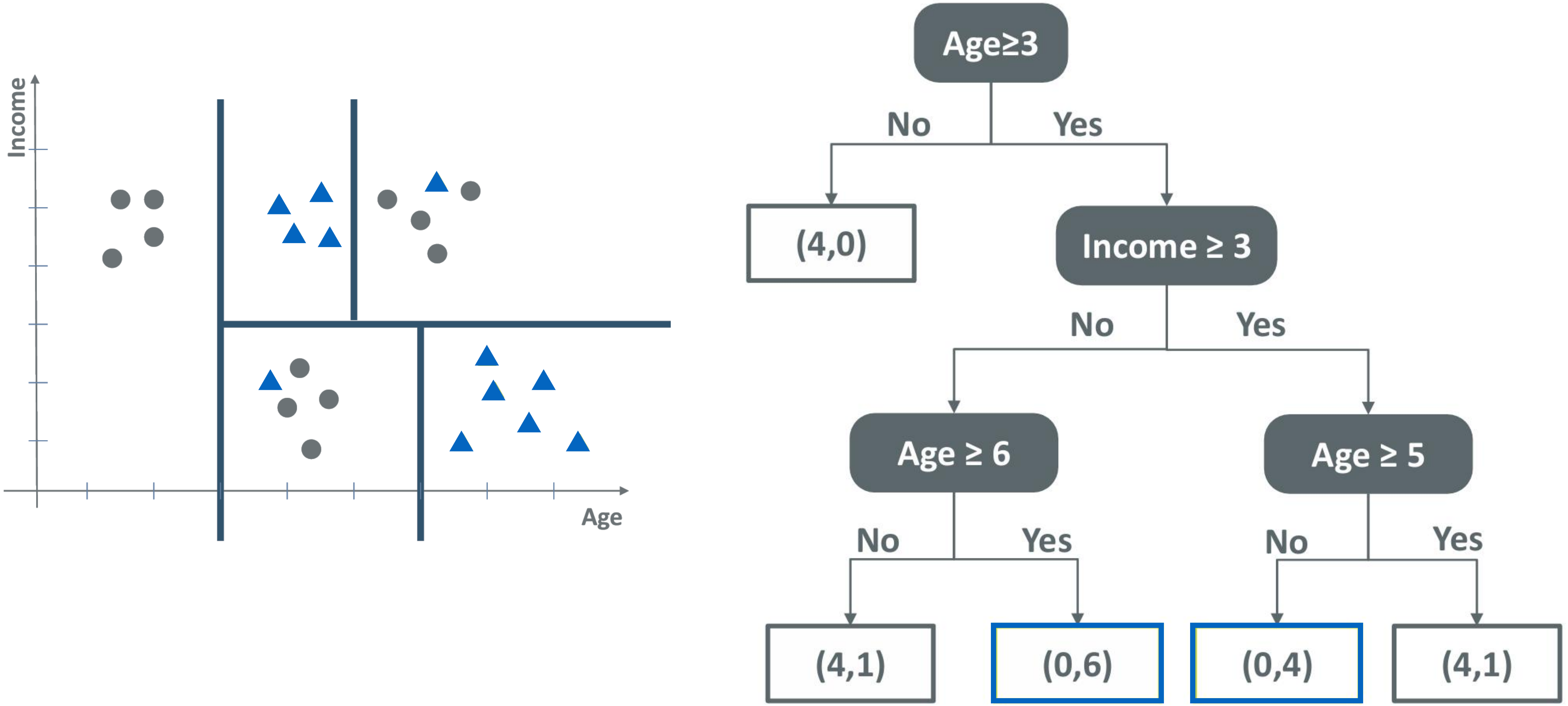
- Em alguns problemas, estamos apenas interessados em obter a melhor precisão possível. Noutros, estamos mais interessados em compreender os resultados e a forma como o modelo está a produzir as estimativas!

Interpretabilidade

1. Árvores de Decisão – Classificação

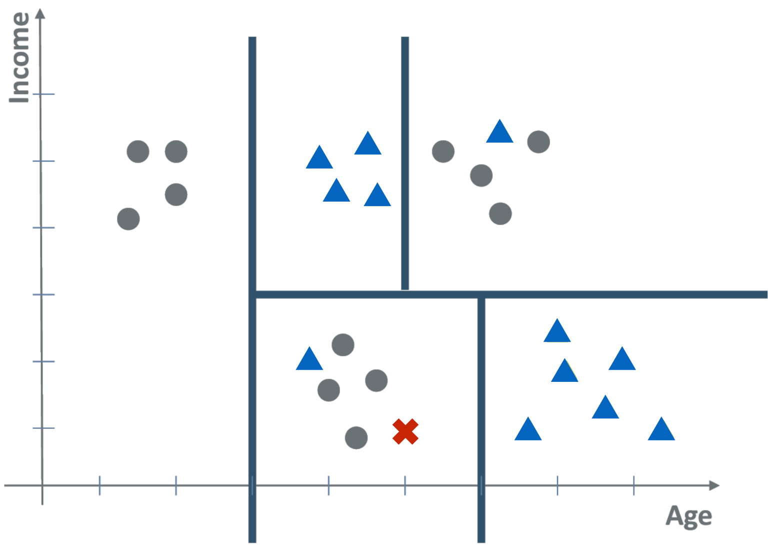


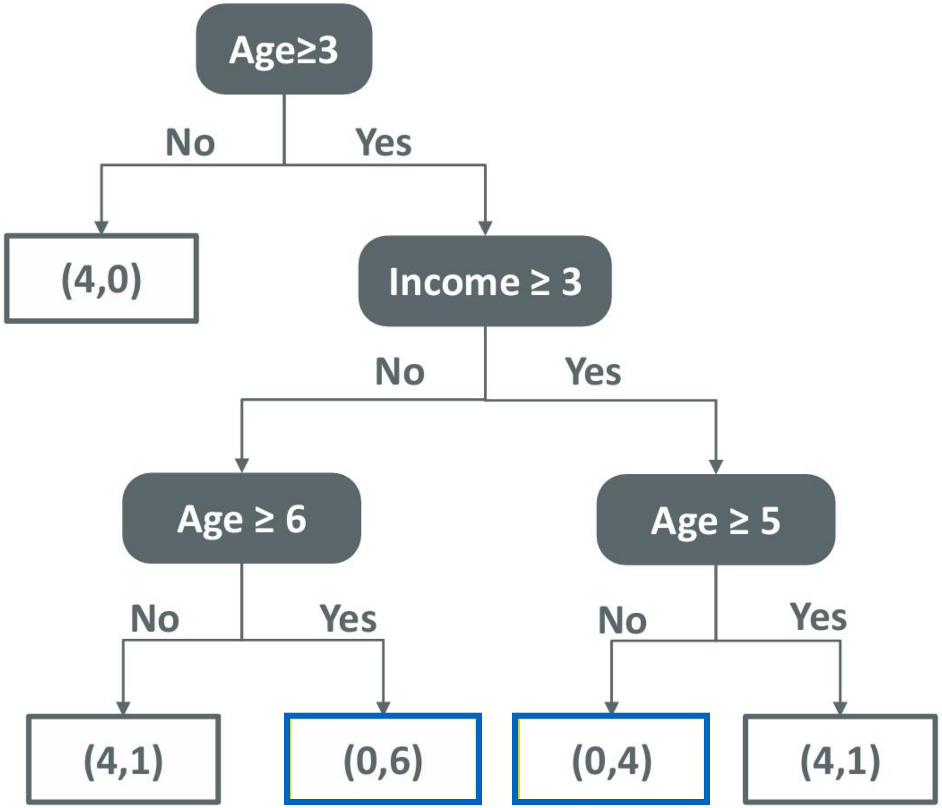
1. Árvores de Decisão – Classificação



1. Árvores de Decisão – Classificação

E se tivermos uma nova observação
onde Age = 5 e Income =1?

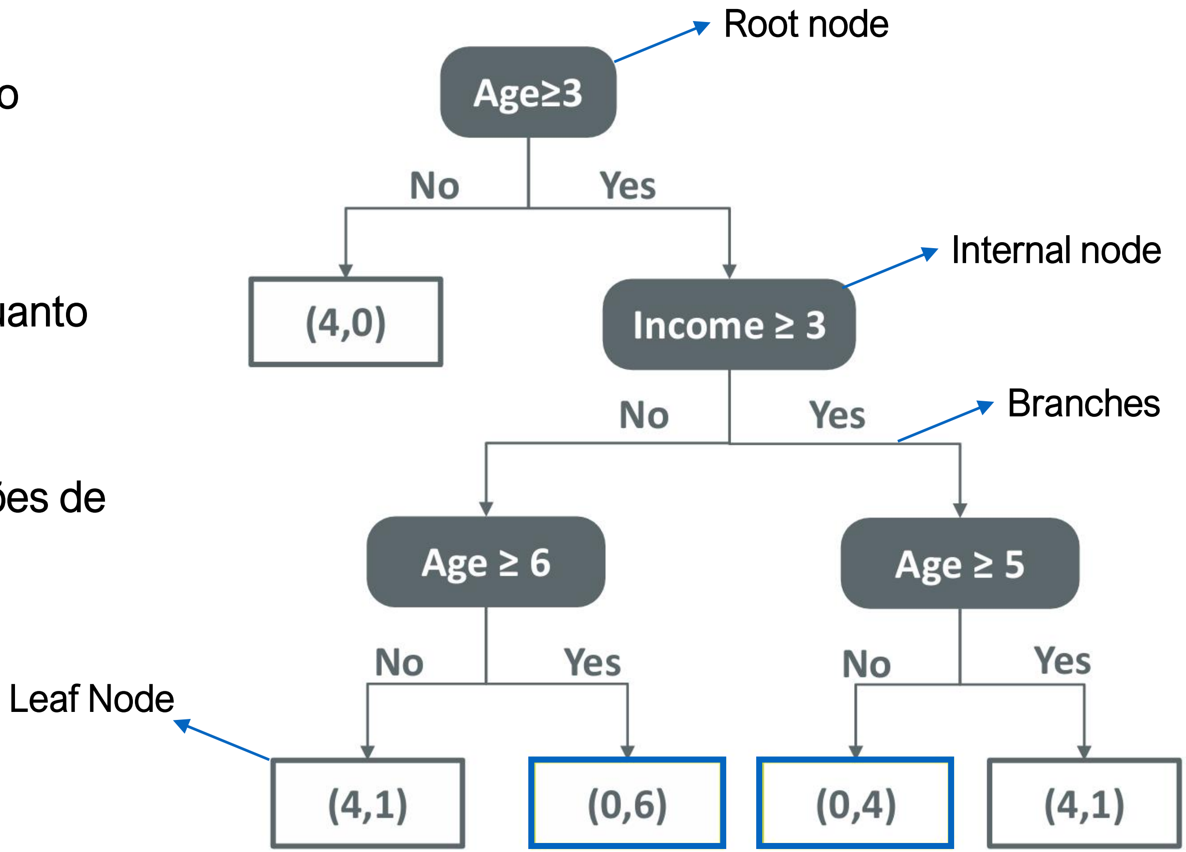




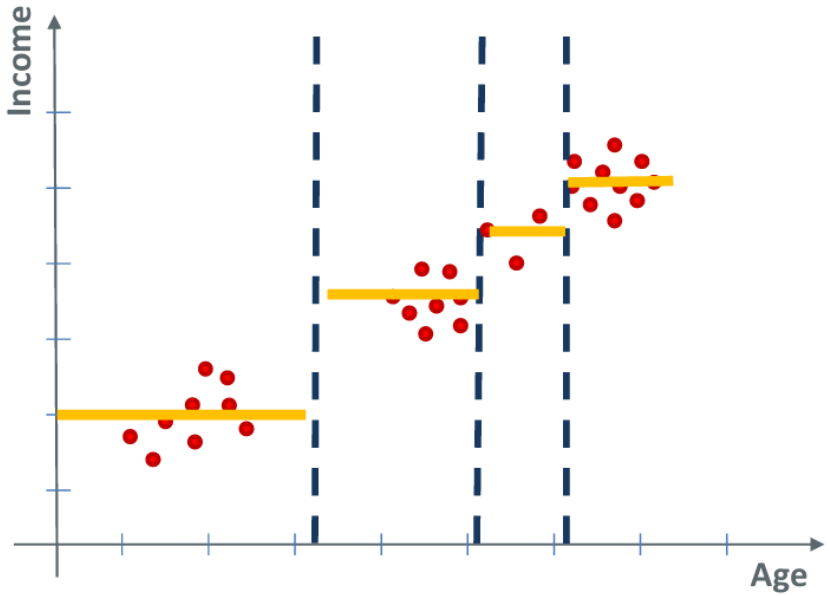
1. Árvores de Decisão – Classificação

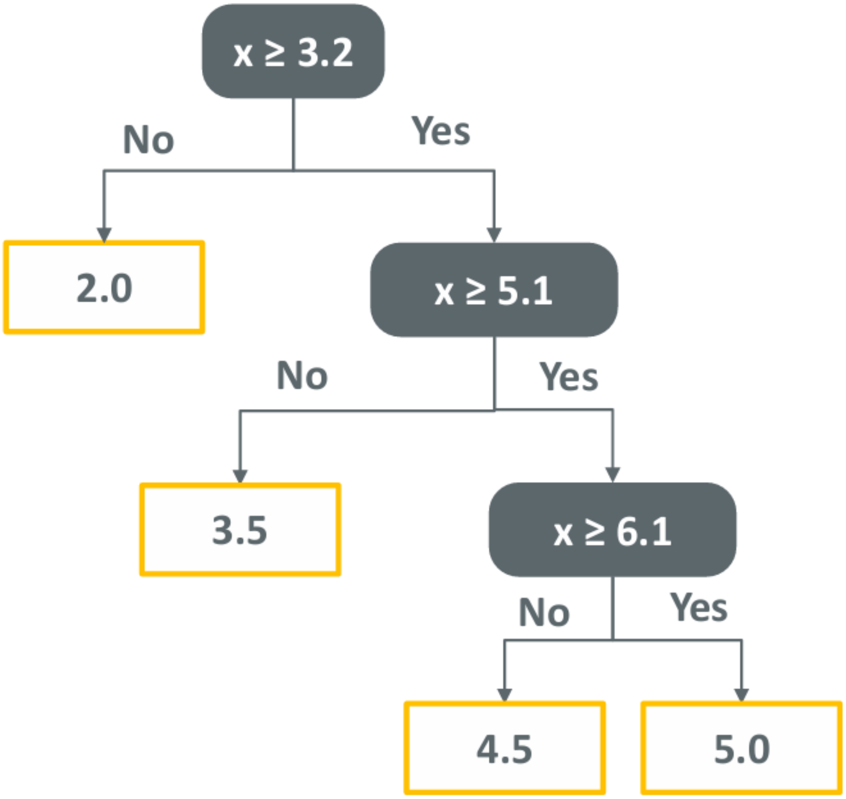
Num problema de classificação, o objetivo é:

- Discriminar entre classes;
- Obter leaf nodes tão puras quanto possível;
- Idealmente, cada leaf node represente apenas observações de uma determinada classe.



1. Árvores de Decisão – Regressão

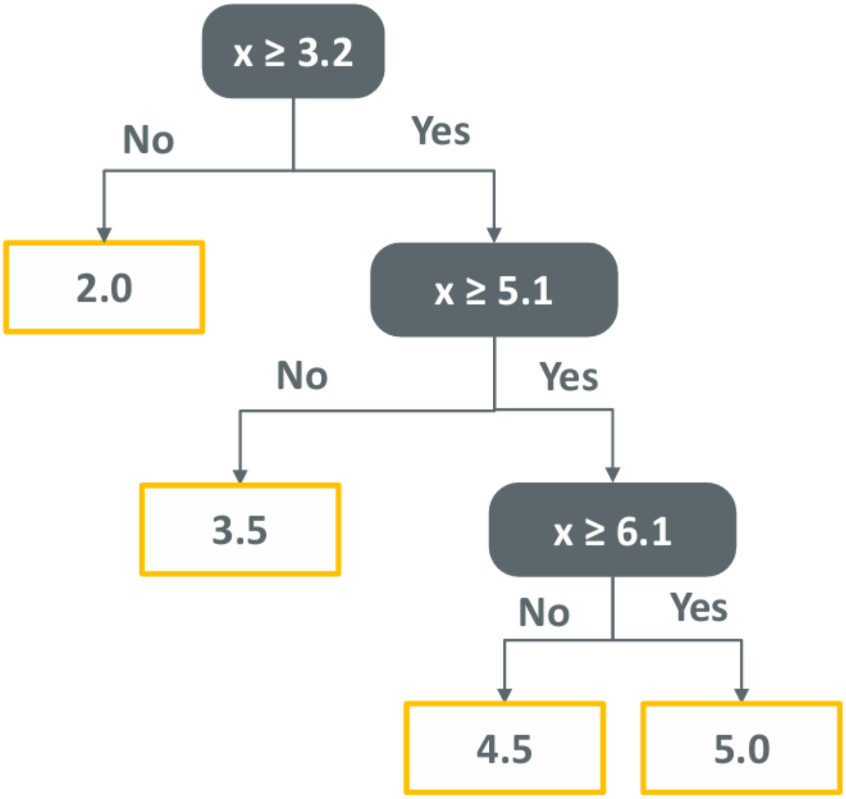




1. Árvores de Decisão – Regressão

Num problema de regressão, o objetivo é:

- Segmentar as variáveis independentes de modo a que as observações sejam o mais parecidas possíveis relativamente à variável a prever.
- O valor das leaf nodes devem ser o mais próximos possíveis das observações que a compõem.



1. Árvores de Decisão – Regras extraídas

- Cada branch adiciona uma conjunção (∧);
- Cada leaf adiciona uma disjunção (∨)

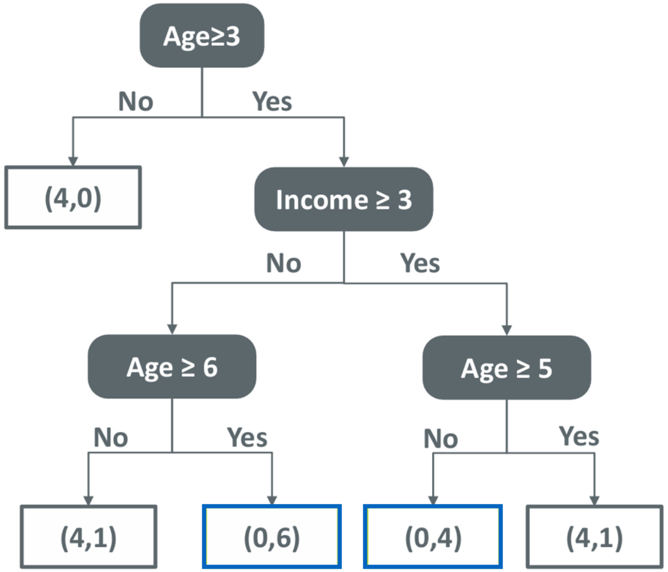
● $\leftrightarrow (age < 3)$

▲ $\leftrightarrow (age \geq 3) \wedge (income < 3) \wedge (age \geq 6)$

▲ $\leftrightarrow (age \geq 3) \wedge (income \geq 3) \wedge (age < 5)$

● $\leftrightarrow (age \geq 3) \wedge (income < 3) \wedge (age < 6)$

● $\leftrightarrow (age \geq 3) \wedge (income \geq 3) \wedge (age \geq 5)$



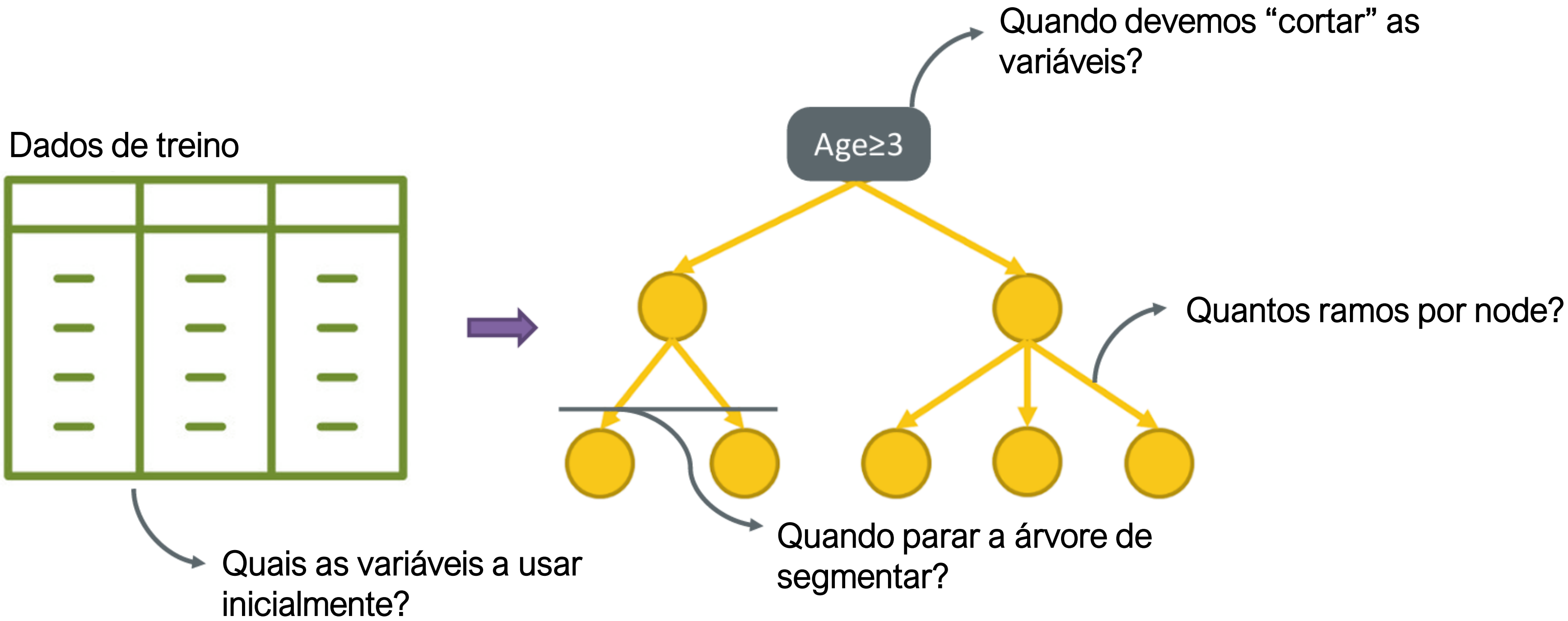
● $\leftrightarrow (age < 3) \vee ((age \geq 3) \wedge (income < 3) \wedge (age < 6)) \vee ((age \geq 3) \wedge (income \geq 3) \wedge (age \geq 5))$

▲ $\leftrightarrow ((age \geq 3) \wedge (income < 3) \wedge (age \geq 6)) \vee ((age \geq 3) \wedge (income \geq 3) \wedge (age < 5))$

1. Árvores de Decisão – Vantagens

- Interpretação: Compreender facilmente a razão subjacente à decisão;
- Insensível a factores de escala: Podem ser utilizados diferentes tipos de medidas sem necessidade de normalização;
- Definição automática das variáveis mais relevantes em cada caso;
- As variáveis mais relevantes aparecem na parte superior da árvore;
- As árvores de decisão são consideradas um método não paramétrico: não há suposições sobre a distribuição das variáveis.

1. Árvores de Decisão – Problemas



2. Árvores de Classificação

2. Árvores de Classificação

Exemplo: Classificação de celas

# núcleos	# caudas	Cor	Membrana	Classe
1	1	Clara	Fina	X
2	1	Clara	Fina	X
1	1	Clara	Grossa	X
1	1	Escura	Fina	X
1	1	Escura	Grossa	X
2	2	Clara	Fina	Y
2	2	Escura	Fina	Y
2	2	Escura	Grossa	Y
2	1	Escura	Fina	Z
2	1	Escura	Grossa	Z
1	2	Clara	Fina	Z
1	2	Clara	Grossa	Z

2. Árvores de Classificação

Métrica discriminativa: mede o **poder discriminativo de uma determinada variável** em relação à variável alvo.

$$Poder\ discriminativo = \frac{1}{n}\sum_{i=1} C_i$$

Onde:
- n é o número de observações;
- C_i é o número de observações classificadas com a classe mais frequente da variável alvo, para cada categoria i de uma determinada variável independente.

2. Árvores de Classificação

# núcleos	# caudas	Cor	Membrana	Classe
1	1	Clara	Fina	X
2	1	Clara	Fina	X
1	1	Clara	Grossa	X
1	1	Escura	Fina	X
1	1	Escura	Grossa	X
2	2	Clara	Fina	Y
2	2	Escura	Fina	Y
2	2	Escura	Grossa	Y
2	1	Escura	Fina	Z
2	1	Escura	Grossa	Z
1	2	Clara	Fina	Z
1	2	Clara	Grossa	Z

Usando o # de núcleos:

# núcleos	1	2
X	4	1
Y	0	3
Z	2	2

$$Poder\ discriminativo = \frac{4+3}{12} = 0.58$$

2. Árvores de Classificação

# núcleos	# caudas	Cor	Membrana	Classe
1	1	Clara	Fina	X
2	1	Clara	Fina	X
1	1	Clara	Grossa	X
1	1	Escura	Fina	X
1	1	Escura	Grossa	X
2	2	Clara	Fina	Y
2	2	Escura	Fina	Y
2	2	Escura	Grossa	Y
2	1	Escura	Fina	Z
2	1	Escura	Grossa	Z
1	2	Clara	Fina	Z
1	2	Clara	Grossa	Z

Usando o # de caudas:

# caudas	1	2
X	5	0
Y	0	3
Z	2	2

$$Poder\ discriminativo = \frac{5+3}{12} = 0.67$$

2. Árvores de Classificação

# núcleos	# caudas	Cor	Membrana	Classe
1	1	Clara	Fina	X
2	1	Clara	Fina	X
1	1	Clara	Grossa	X
1	1	Escura	Fina	X
1	1	Escura	Grossa	X
2	2	Clara	Fina	Y
2	2	Escura	Fina	Y
2	2	Escura	Grossa	Y
2	1	Escura	Fina	Z
2	1	Escura	Grossa	Z
1	2	Clara	Fina	Z
1	2	Clara	Grossa	Z

Usando a cor:

Cor	Clara	Escura
X	3	2
Y	1	2
Z	2	2

$$Poder\ discriminativo = \frac{3+2}{12} = 0.41$$

2. Árvores de Classificação

# núcleos	# caudas	Cor	Membrana	Classe
1	1	Clara	Fina	X
2	1	Clara	Fina	X
1	1	Clara	Grossa	X
1	1	Escura	Fina	X
1	1	Escura	Grossa	X
2	2	Clara	Fina	Y
2	2	Escura	Fina	Y
2	2	Escura	Grossa	Y
2	1	Escura	Fina	Z
2	1	Escura	Grossa	Z
1	2	Clara	Fina	Z
1	2	Clara	Grossa	Z

Usando a membrana:

Membrana	Fina	Grossa
X	3	2
Y	2	1
Z	2	2

$$Poder\ discriminativo = \frac{3+2}{12} = 0.41$$

2. Árvores de Classificação

# núcleos	1	2
X	4	1
Y	0	3
Z	2	2

0.58

# caudas	1	2
X	5	0
Y	0	3
Z	2	2

0.67

Cor	Clara	Escura
X	3	2
Y	1	2
Z	2	2

0.41

Membrana	Fina	Grossa
X	3	2
Y	2	1
Z	2	2

0.41

Escolha: # caudas

2. Árvores de Classificação

Caudas

1

# núcleos	Cor	Membrana	Classe
1	Clara	Fina	X
2	Clara	Fina	X
1	Clara	Grossa	X
1	Escura	Fina	X
1	Escura	Grossa	X
2	Escura	Fina	Z
2	Escura	Grossa	Z

2

# núcleos	Cor	Membrana	Classe
2	Clara	Fina	Y
2	Escura	Fina	Y
2	Escura	Grossa	Y
1	Clara	Fina	Z
1	Clara	Grossa	Z

2. Árvores de Classificação

# núcleos	Cor	Membrana	Classe
1	Clara	Fina	X
2	Clara	Fina	X
1	Clara	Grossa	X
1	Escura	Fina	X
1	Escura	Grossa	X
2	Escura	Fina	Z
2	Escura	Grossa	Z

# núcleos	1	2
X	4	1
Y	0	0
Z	0	2

$$\boldsymbol{P.D = 0.86}$$

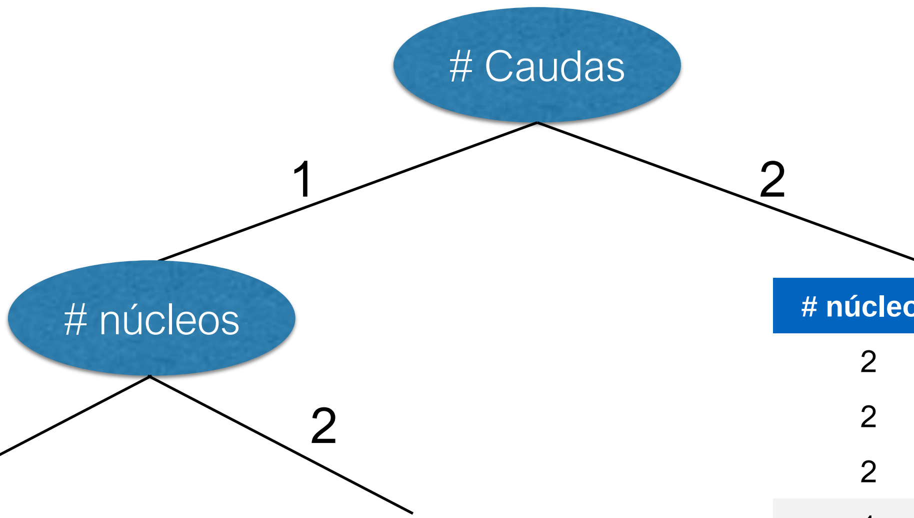
Cor	Clara	Escura
X	3	2
Y	0	0
Z	0	2

$$\boldsymbol{P.D = 0.71}$$

Membrana	Fina	Grossa
X	3	2
Y	0	0
Z	1	1

$$\boldsymbol{P.D = 0.71}$$

2. Árvores de Classificação



Caudas = 1, núcleos = 1:

Cor	Membrana	Classe
Clara	Fina	X
Clara	Grossa	X
Escura	Fina	X
Escura	Grossa	X

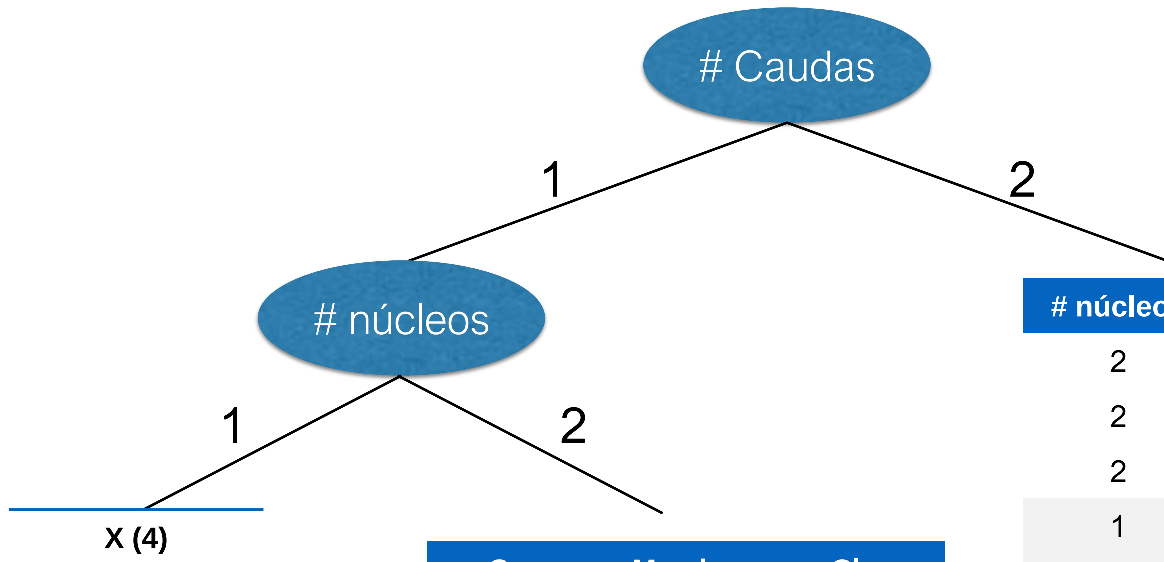
Caudas = 1, núcleos = 2:

Cor	Membrana	Classe
Clara	Fina	X
Escura	Fina	Z
Escura	Grossa	Z

Caudas = 2:

# núcleos	Cor	Membrana	Classe
2	Clara	Fina	Y
2	Escura	Fina	Y
2	Escura	Grossa	Y
1	Clara	Fina	Z
1	Clara	Grossa	Z

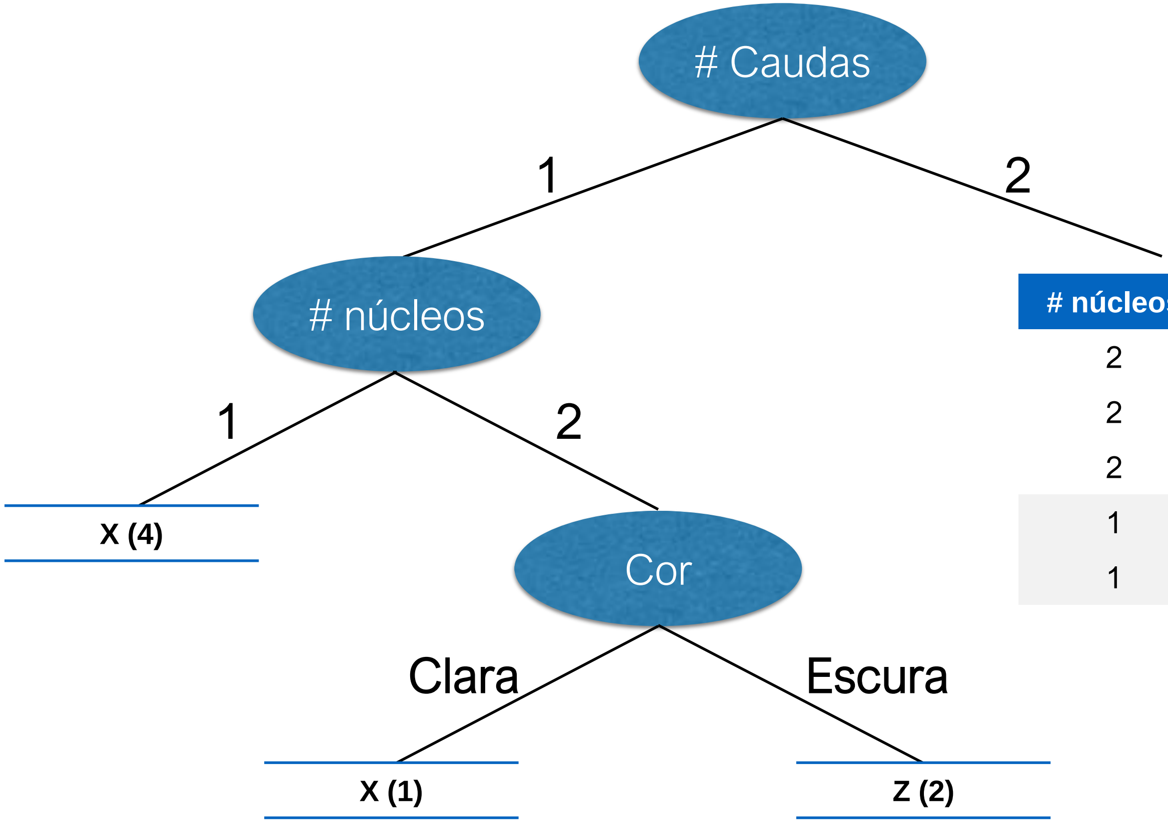
2. Árvores de Classificação



# núcleos	Cor	Membrana	Classe
2	Clara	Fina	Y
2	Escura	Fina	Y
2	Escura	Grossa	Y
1	Clara	Fina	Z
1	Clara	Grossa	Z

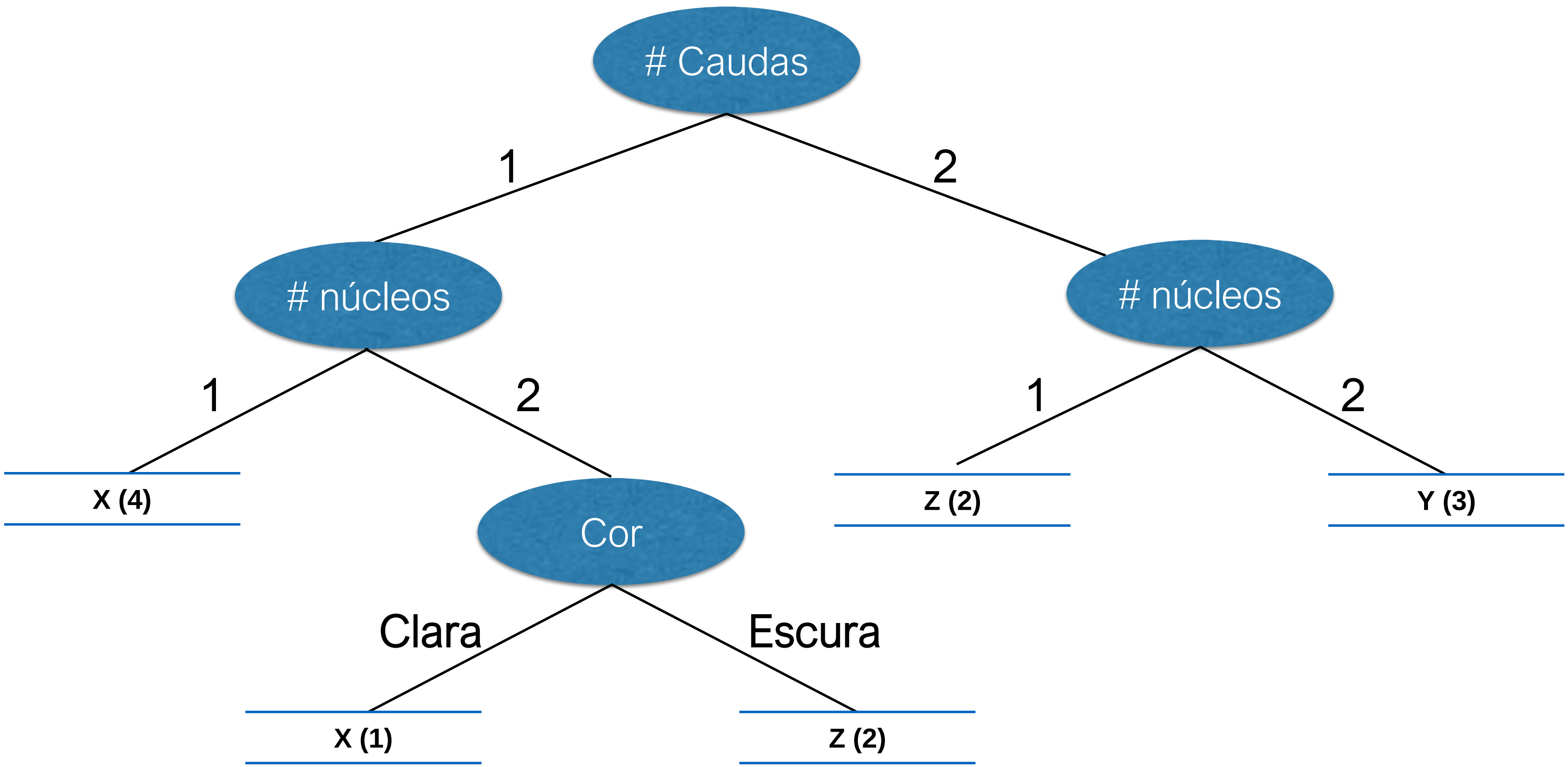
Cor	Membrana	Classe
Clara	Fina	X
Escura	Fina	Z
Escura	Grossa	Z

2. Árvores de Classificação



# núcleos	Cor	Membrana	Classe
2	Clara	Fina	Y
2	Escura	Fina	Y
2	Escura	Grossa	Y
1	Clara	Fina	Z
1	Clara	Grossa	Z

2. Árvores de Classificação

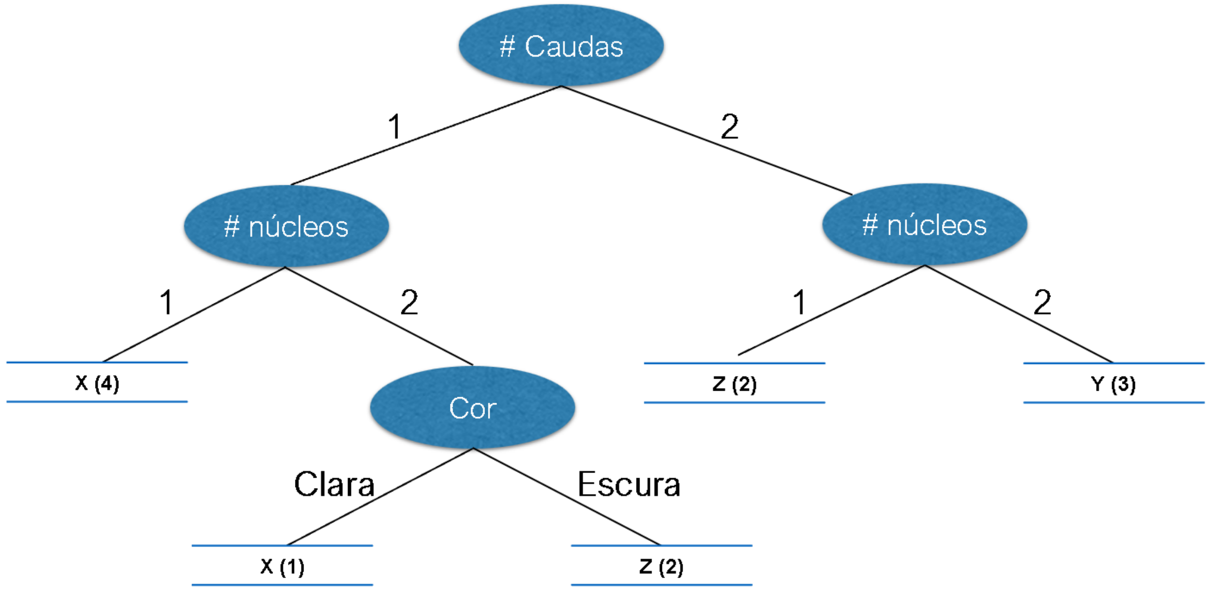


2. Árvores de Classificação

X $\leftrightarrow$ $((\#caudas = 1) \wedge (\#nucleos = 1)) \vee ((\#caudas = 1) \wedge (\#nucleos = 2) \wedge (cor = Clara))$

Y $\leftrightarrow$ $((\#caudas = 2) \wedge (\#nucleos = 2))$

Z $\leftrightarrow$ $((\#caudas = 2) \wedge (\#nucleos = 1)) \vee ((\#caudas = 1) \wedge (\#nucleos = 2) \wedge (cor = Escura))$



2.1 Árvores de Classificação – CART

2.1 Árvores de Classificação – CART

- O algoritmo CART (Classification And Regression Trees) é o algoritmo usado pelo scikit-learn.
- Utiliza o Gini index para medir a "desordem" em cada variável independente:

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

Onde:
- D são os dados de treino para uma determinada classe de uma determinada variável ;
- p_i é a probabilidade de uma observação em D pertencer à classe C_i da variável alvo, estimada por $C_{i,D}/D$;
- $C_{i,D}$ é o número de observações pertencentes à classe C_i e D.

2.1 Árvores de Classificação – CART

- Começamos por calcular o grau de desordem na variável alvo:

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

- Se escolhermos a variável A para dividir o "nó" da árvore em v partições, então o grau de desordem é calculado através de:

$$Gini_A(D) = \sum_{j=1}^{v} \frac{D_j}{D} \times Gini(D_j)$$

- A partir do Gini index, podemos calcular o ganho de informação para uma determinada variável independente A:

$$Gain(A) = Gini(D) - Gini_A(D)$$

2.1 Árvores de Classificação – CART

Exemplo: compra de um computador

idade	rendimento	estudante	rating_crédito	compra_pc
<= 30	elevado	não	médio	não
<= 30	elevado	não	excelente	não
[31, …,40]	elevado	não	médio	sim
> 40	médio	não	médio	sim
> 40	baixo	sim	médio	sim
> 40	baixo	sim	excelente	não
[31, …,40]	baixo	sim	excelente	sim
<= 30	médio	não	médio	não
<= 30	baixo	sim	médio	sim
> 40	médio	sim	médio	sim
<= 30	médio	sim	excelente	sim
[31, …,40]	médio	não	excelente	sim
[31, …,40]	elevado	sim	médio	sim
> 40	médio	não	excelente	não

Variável dependente

2.1 Árvores de Classificação – CART

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

compra_pc	
sim	não
9	5

$$Gini_{pc}(D) = Gini(9,5) = 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2 = 0.459$$

2.1 Árvores de Classificação – CART

Calculando o Gini associado à variável idade:

idade	sim	não	Total	Gini
<= 30	2	3	5	0.48
[31, …, 40]	4	0	4	0
> 40	3	2	5	0.48

$$Gini_{\leq 30}(D) = Gini(2,3) = 1 - (2/5)^2 - (3/5)^2 = 0.48$$

$$Gini_{[31,\dots,40]}(D) = Gini(4,0) = 1 - (4/4)^2 - (0/4)^2 = 0$$

$$Gini_{>40}(D) = Gini(3,2) = 1 - (3/5)^2 - (2/5)^2 = 0.48$$

$$Gini_A(D) = \sum_{j=1}^{v} \frac{D_j}{D} \times Gini(D_j)$$

$$Gini_{idade}(D) = \frac{5}{14} Gini_{\leq 30}(3,2) + \frac{4}{14} Gini_{[31,\dots,40]}(4,0) + \frac{5}{14} Gini_{>40}(3,2) = 0.343$$

2.1 Árvores de Classificação – CART

$$Gain(A) = Gini(D) - Gini_A(D)$$

Assim, o ganho de informação na variável idade é:

$$Gain(idade) = Gini_{pc}(D) - Gini_{idade}(D) = 0.459 \; - \; 0.343 = 0.116$$

Se repetirmos o mesmo para as restantes variáveis:

$$Gain(rendimento) = 0.459 \; - \; 0.440 = 0.019$$

$$Gain(estudante) = 0.459 \; - \; 0.368 = 0.091$$

$$Gain(crédito) = 0.459 \; - \; 0.429 = 0.03$$

Idade é a variável com o maior ganho de informação, pelo que começamos com essa variável na segmentação da árvore.

2.1 Árvores de Classificação – CART

- Em alternativa ao Gini index, podemos usar a Entropy:

$$Entropy(D) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

- Se escolhermos a variável A para dividir o "nó" da árvore em v partições, então o grau de desordem é calculado através de:

$$Entropy_A(D) = \sum_{j=1}^{v} \frac{D_j}{D} \times Entropy(D_j)$$

- A partir do Gini index, podemos calcular o ganho de informação para uma determinada variável independente A:

$$Gain(A) = Entropy(D) - Entropy_A(D)$$

2.1 Árvores de Classificação – CART

$$Entropy(D) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

compra_pc	
sim	não
9	5

$$Entropy_{pc}(D) = E(9,5) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right)^2 - \frac{5}{14}\log_2\left(\frac{5}{14}\right)^2$$
$$= 0.940$$

2.1 Árvores de Classificação – CART

Calculando a Entropy associado à variável idade:

idade	sim	não	Total	Gini
<= 30	2	3	5	0.48
[31, …, 40]	4	0	4	0
> 40	3	2	5	0.48

$$Entropy_{\leq 30}(D) = E(2,3) = -(2/5)log_2(2/5) - (3/5)log_2(3/5) = 0.971$$

$$Gini_{[31,\ldots,40]}(D) = E(4,0) = -(4/4)log_2(4/4) - (0/4)log_2(0/4) = 0$$

$$Gini_{>40}(D) = E(3,2) = -(3/5)log_2(3/5) - (2/5)log_2(2/5) = 0.971$$

$$Entropy_A(D) = \sum_{j=1}^{v} \frac{D_j}{D} \times Entropy(D_j)$$

$$Entropy_{idade}(D) = \frac{5}{14} E_{\leq 30}(3,2) + \frac{4}{14} E_{[31,\ldots,40]}(4,0) + \frac{5}{14} E_{>40}(3,2) = 0.694$$

2.1 Árvores de Classificação – CART

$$Gain(A) = Entropy(D) - Entropy_A(D)$$

Assim, o ganho de informação na variável idade é:

$$Entropy(idade) = E_{pc}(D) - E_{idade}(D) = 0.940\ -\ 0.694 = 0.246$$

Se repetirmos o mesmo para as restantes variáveis:

$$Entropy(rendimento) = 0.029$$

$$Entropy(estudante) = 0.151$$

$$Entropy(crédito) = 0.048$$

Idade continua a ser a variável com o maior ganho de informação, pelo que começamos com essa variável na segmentação da árvore.

2.1 Árvores de Classificação – CART

Idade

idade	rendimento	estudante	crédito	compra_pc
<= 30	elevado	não	médio	não
<= 30	elevado	não	excelente	não
<= 30	médio	não	médio	não
<= 30	baixo	sim	médio	sim
<= 30	médio	sim	excelente	sim

idade	rendimento	estudante	crédito	compra_pc
[31, …,40]	elevado	não	médio	sim
[31, …,40]	baixo	sim	excelente	sim
[31, …,40]	médio	não	excelente	sim
[31, …,40]	elevado	sim	médio	sim

idade	rendimento	estudante	crédito	compra_pc
> 40	médio	não	médio	sim
> 40	baixo	sim	médio	sim
> 40	baixo	sim	excelente	não
> 40	médio	sim	médio	sim
> 40	médio	não	excelente	não

2.1 Árvores de Classificação – CART

- Na verdade, a árvore de decisão CART apenas cria ramos binários para uma determinada variável.
- Ou seja, no caso da variável Idade, não seriam criados 3 ramos, mas sim apenas 2… esta é uma das limitações deste algoritmo.
- Assim, antes de iniciarmos o treino do modelo, devemos transformar as variáveis categóricas em variáveis numéricas, por exemplo, através de one-hot encoding, obrigando a árvore a criar ramos binários.
- Algoritmos como o ID3, C4.5, C5 permitem a criação de mais de 2 ramos!

2.1 Árvores de Classificação – CART

Mas e **se a variável Idade fosse numérica**?

Então seria necessário determinar o melhor ponto de corte:

- Ordenar os valores da variável;
- Cada ponto médio entre cada par de valores adjacentes é um possível ponto de corte;
- Avaliar a medida de grau de desorderm (Gini, Entropy, etc.) para as partições separadas por cada ponto de corte.
- Escolher o ponto de corte que produza o menor grau de desordem.
- Este grau de desordem (e ganho de informação) é comparado com o das restantes variáveis, como visto anteriormente.

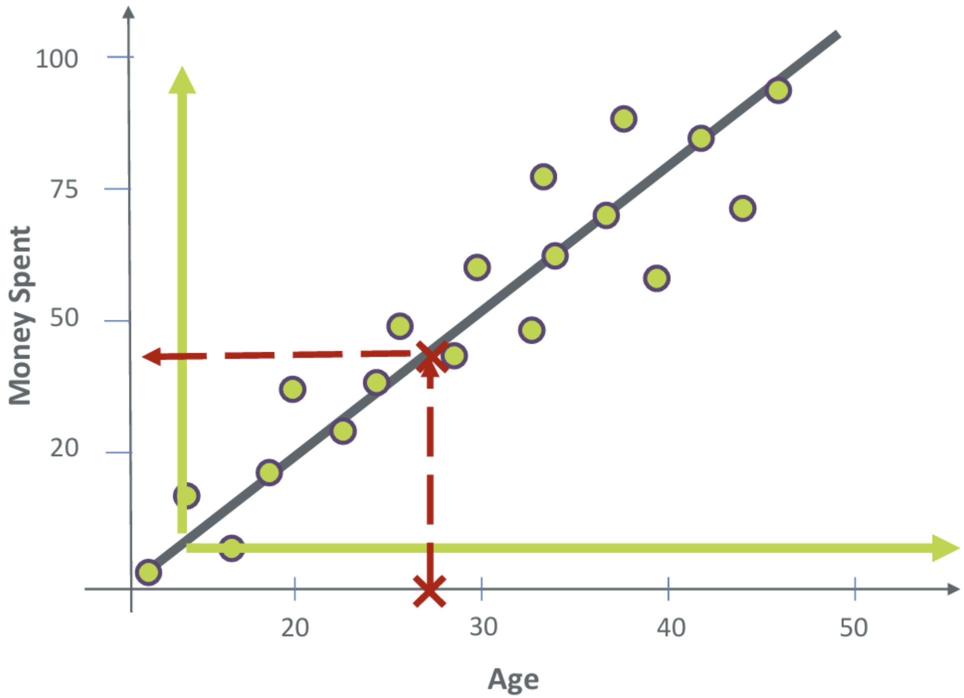
Idade
18
20
22
26
27
29
31
35
41
42
45
48
49

21

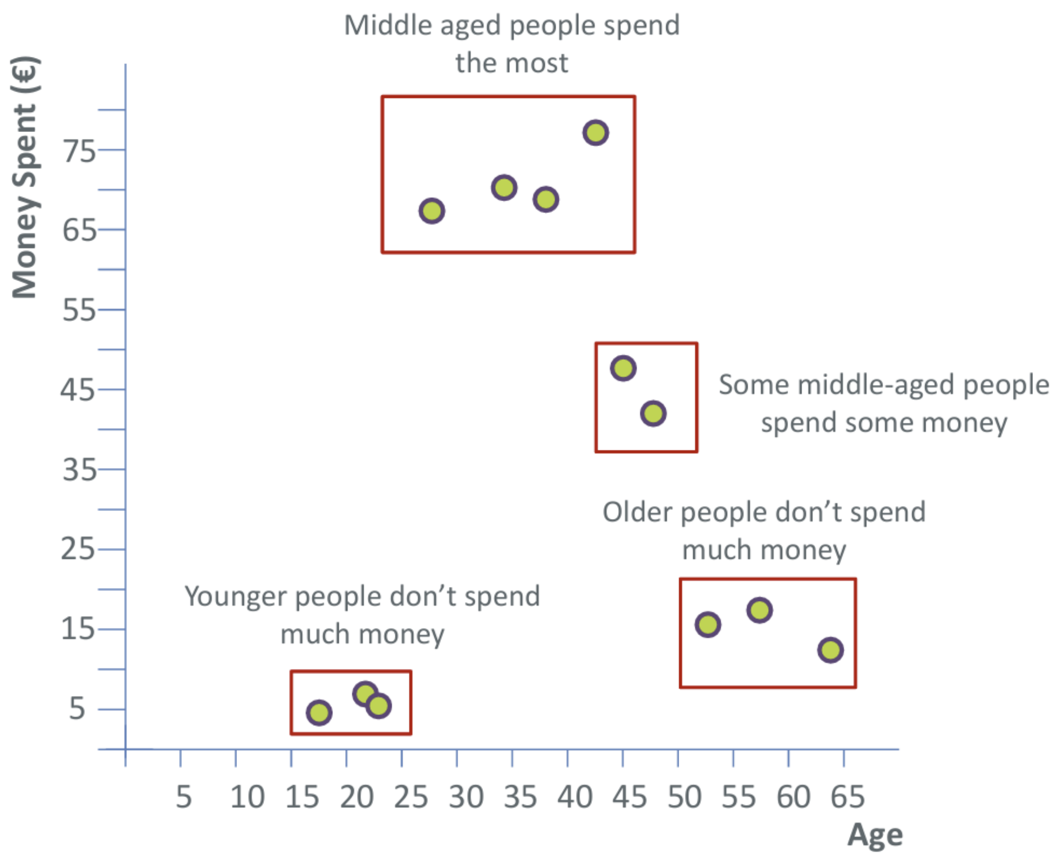
2.2 Árvores de Regressão

2.2 Árvores de Regressão

- Num problema de regressão, há alguns problemas que são fáceis de ajustar utilizando uma regressão linear:



2.2 Árvores de Regressão



2.2 Árvores de Regressão

- Para certos conjuntos de dados, devemos utilizar outros métodos que não a utilização de linhas rectas para fazer previsões.
- Uma opção é a árvore de regressão!
- Uma árvore de regressão é um tipo de árvore de decisão em que cada "leaf node" representa um valor numérico e não uma categoria discreta como nas árvores de decisão de classificação.

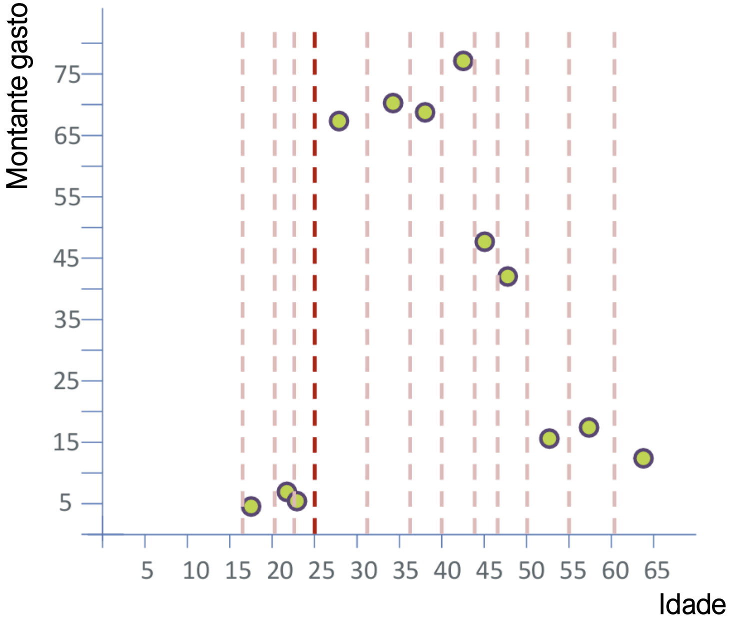
2.2 Árvores de Regressão

- Quando temos mais de 2 preditores, tais como a idade, o número de filhos e o número de meses como cliente, prever o dinheiro montante através de um gráfico é muito difícil, se não impossível…

Idade	#Filhos	#Meses como Cliente	…	Montante gasto
18	0	1	…	5
22	1	1	…	7
23	0	3	…	6
27	2	2	…	68
34	3	2	…	72
38	2	3	…	69
43	1	2	…	77
45	0	1	…	48
48	3	2	…	42
53	2	1	…	16
57	3	3	…	18
64	1	2	…	14

2.2 Árvores de Regressão

- A primeira coisa a fazer, tal como nas árvores de classificação, é decidir qual a variável com menos "grau de desordem" para começar a utilizar na árvore.
- Mas como decidimos qual o ponte de corte a usar nas variáveis, de modo a calcular esse "grau de desordem"?



Para cada variável, ordenamos os valores e testamos diferentes pontes de corte, calculando o erro total da previsão (média da variável alvo) para cada grupo:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2$$

Em alternativa do MSE, o MAE pode ser usado.

2.2 Árvores de Regressão

- Começando na variável Idade:

Idade	Montante gasto (y)
18	5
22	7
23	6
27	68
34	72
…	..
57	18
64	14

Se o ponto de corte for Idade ≤ 20:

- 1º grupo de observações:

$$\bar{y} = 5$$

$$MSE(Idade \leq 20) = \frac{1}{1}(5-5)^2 = 0$$

- 2º grupo de observações:

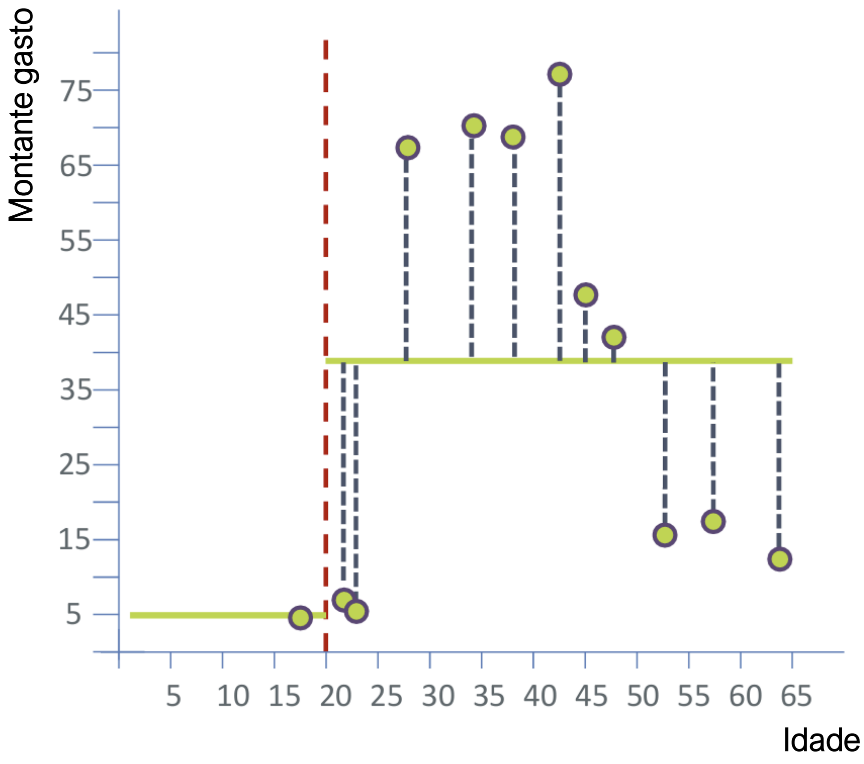
$$\bar{y} = \frac{7+6+\cdots+14}{11} = 39.7$$

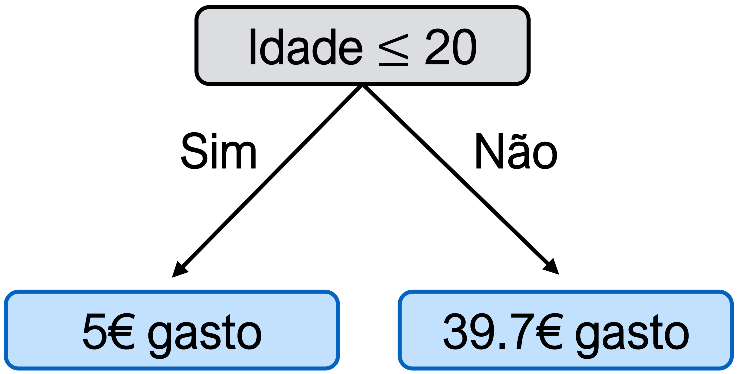
$$MSE(Idade > 20) = \frac{1}{11}((7-39.7)^2 + (6-39.7)^2 + \cdots + (14-39.7)^2) = 733.3$$

$$TOTAL\ MSE = 0 + 733.3 = 733.3$$

2.2 Árvores de Regressão

Se o ponto de corte for Idade ≤ 20:





2.2 Árvores de Regressão

- Começando na variável Idade:

Idade	Montante gasto (y)
18	5
22	7
23	6
27	68
34	72
…	…
57	18
64	14

<u>Se o ponto de corte for Idade ≤ 22.5:</u>

- 1º grupo de observações:

$$\bar{y} = 6$$

$$MSE(Idade \leq 22.5) = \frac{1}{2}((5-6)^2+(7-6)^2) = 1$$

- 2º grupo de observações:

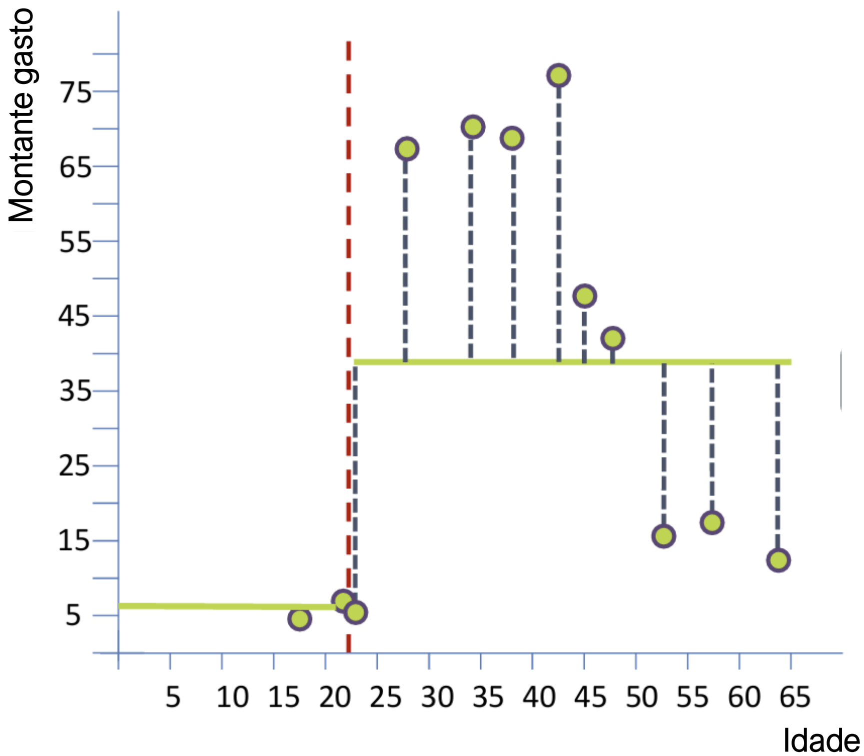
$$\bar{y} = \frac{6+68+\cdots+14}{10} = 43.2$$

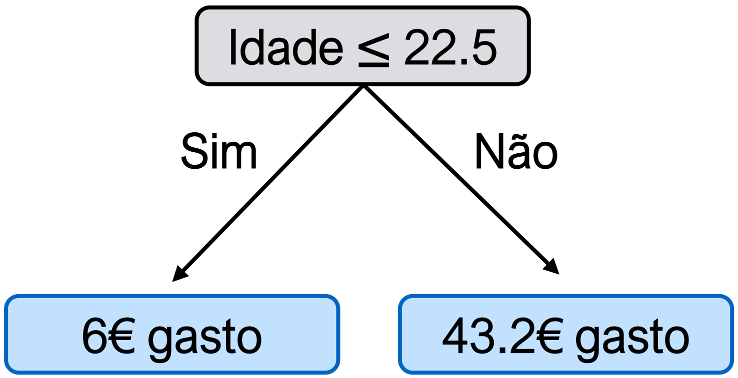
$$MSE(Idade > 22.5) = \frac{1}{10}((6-43.2)^2+(68-43.2)^2+\cdots+(14-43.2)^2) = 688.8$$

$$TOTAL\ MSE = 1 + 688.8 = 689.8$$

2.2 Árvores de Regressão

Se o ponto de corte for Idade ≤ 22.5:





2.2 Árvores de Regressão

- Começando na variável Idade:

Idade	Montante gasto (y)
18	5
22	7
23	6
27	68
34	72
…	…
57	18
64	14

Se o ponto de corte for Idade ≤ 25:

- 1º grupo de observações:

$\bar{y} = 6$

$$MSE(Idade \leq 25) = \frac{1}{3}((5-6)^2+(7-6)^2+(6-6)^2) = 0.7$$

- 2º grupo de observações:

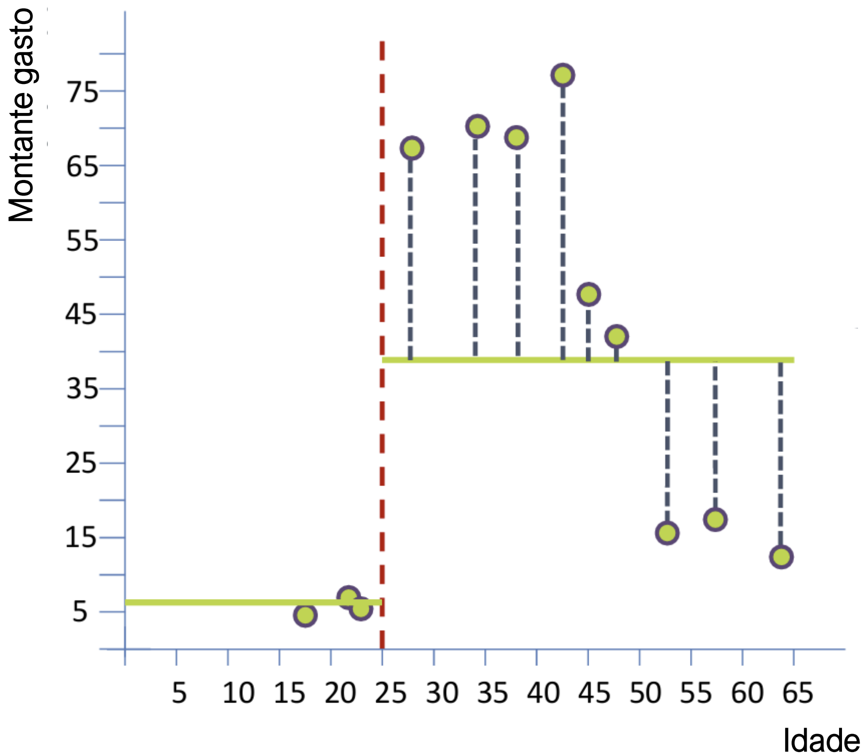
$$\bar{y} = \frac{68 + 72 + \cdots + 14}{10} = 47.1$$

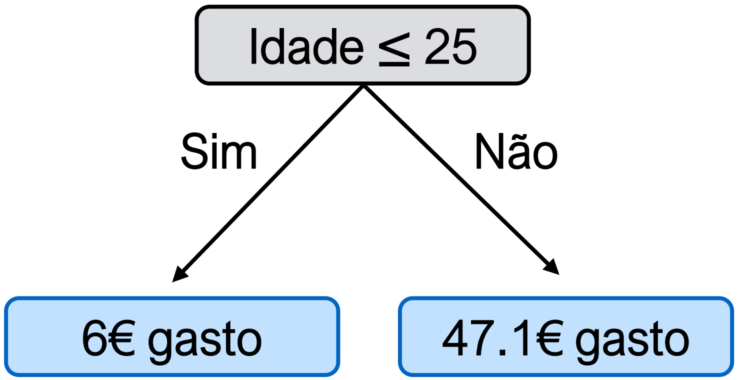
$$MSE(Idade > 25) = \frac{1}{9}((68-47.1)^2 + (72-47.1)^2 + \cdots + (14-47.1)^2) = 596.3$$

$$TOTAL\ MSE = 0.7 + 596.3 = 597.0$$

2.2 Árvores de Regressão

Se o ponto de corte for Idade ≤ 25:

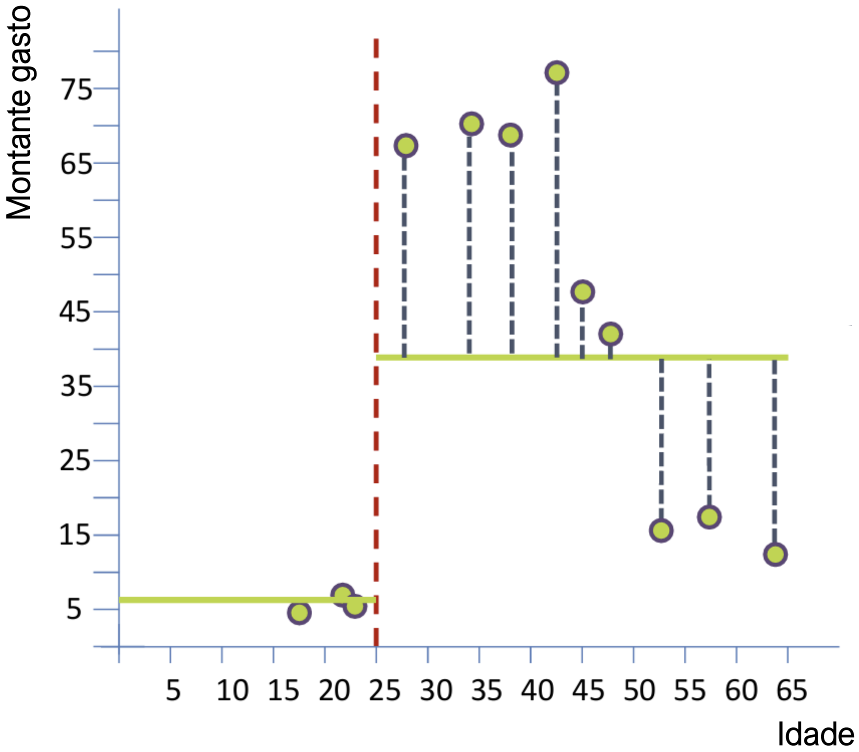




2.2 Árvores de Regressão

- Começando na variável Idade:

Ponto de corte para Idade	*TOTAL MSE*
≤ 20.0	733.3
≤ 22.5	689.8
≤ 25.0	597.0
≤ 30.5	1333.0
≤ 36.0	1558.1
≤ 40.5	1526.6
≤ 44.0	1265.0
≤ 46.5	1056.8
≤ 50.5	828.0
≤ 55.0	816.2
≤ 60.5	782.1



2.2 Árvores de Regressão

- E para a variável #Filhos:

#Filhos	Montante gasto
0	5
1	7
0	6
2	68
3	72
2	69
1	77
0	48
3	42
2	16
3	18
1	14

Ponto de corte para #Filhos	$TOTAL\ MSE$
≤ 0.5	1155.8
≤ 1.5	1301.1
≤ 2.5	1321.6

2.2 Árvores de Regressão

- E para a variável #Meses Cliente:

#Meses	Montante gasto
1	5
1	7
3	6
2	68
2	72
3	69
2	77
1	48
2	42
1	16
3	18
2	14

Ponto de corte para #Meses	$TOTAL\ MSE$
≤ 1.5	1056.7
≤ 2.5	1501.3

2.2 Árvores de Regressão

- Comparando o erro total usando o melhor ponte de corte para cada variável:

Ponto de corte para Idade	$TOTAL\ MSE$
≤ 20.0	733.3
≤ 22.5	689.8
≤ 25.0	597.0
≤ 30.5	1333.0
≤ 36.0	1558.1
≤ 40.5	1526.6
≤ 44.0	1265.0
≤ 46.5	1056.8
≤ 50.5	828.0
≤ 55.0	816.2
≤ 60.5	782.1

Ponto de corte para #Filhos	$TOTAL\ MSE$
≤ 0.5	1155.8
≤ 1.5	1301.1
≤ 2.5	1321.6

Ponto de corte para #Meses	$TOTAL\ MSE$
≤ 1.5	1056.7
≤ 2.5	1501.3

2.2 Árvores de Regressão

- A primeira variável a ser usada é Idade com o ponto de corte ≤ 25:

Idade ≤ 25

Sim | Não

Idade	#Filhos	#Meses como Cliente	…	Montante gasto
18	0	1	…	5
22	1	1	…	7
23	0	3	…	6

Idade	#Filhos	#Meses como Cliente	…	Montante gasto
27	2	2	…	68
34	3	2	…	72
38	2	3	…	69
43	1	2	…	77
45	0	1	…	48
48	3	2	…	42
53	2	1	…	16
57	3	3	…	18
64	1	2	…	14

2.2 Árvores de Regressão - Overfitting

- Para evitar overfitting, podemos definir um número mínimo de samples, abaixo do qual uma determianda ramificação deixa de ser segmentada.

- Por exemplo, se definissemos *min_samples = 7*, o ramo da esquerda terminaria e a 'lead node' Idade ≤ 25 passa a ter um valor previsto de montante gasto de 6!

- Outra forma de evitar overfitting seria defining um nível máximo de profundidade, por exemplo, se *max_depth = 1*, a árvore terminaria.

Obrigado!