



Universidade do Minho
Departamento de Informática

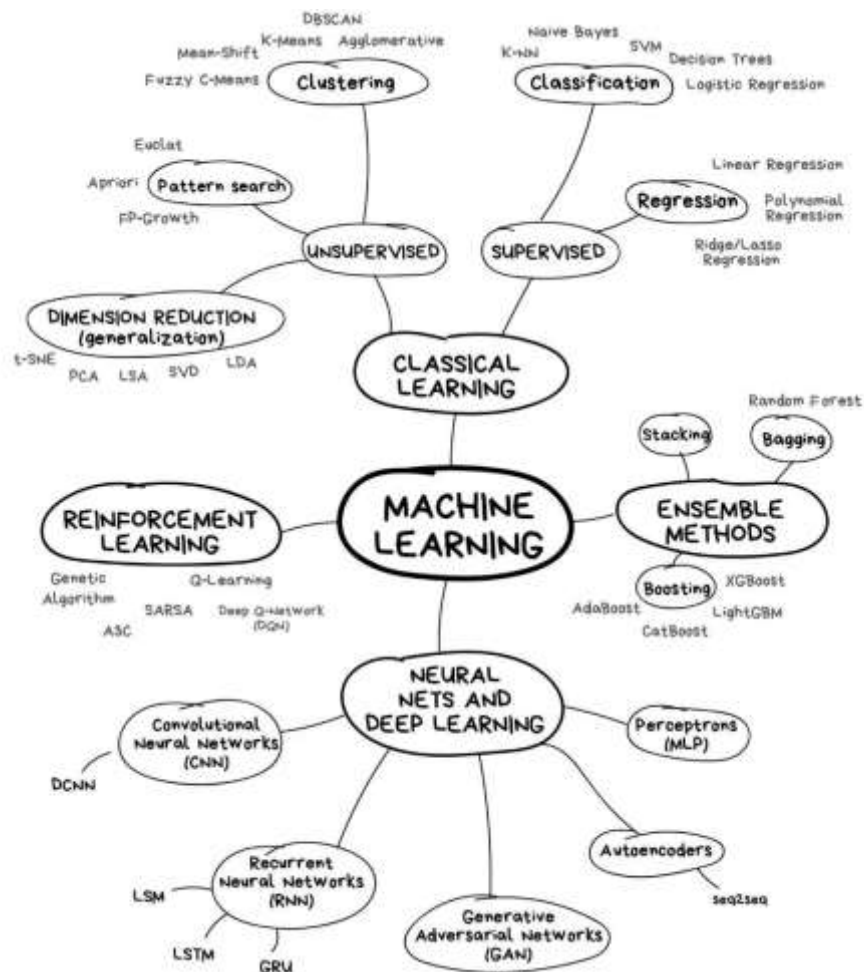
APRENDIZAGEM E DECISÃO INTELIGENTES

LEI/MiEI @ 2022/2023, 2º sem
[ADI³]

- Árvores de Decisão
- Algoritmos
- Poda (Prunning)
- Exemplos



Aprendizagem Automática (Machine Learning)



Classificação

- “*Classification in machine learning is a supervised learning approach in which we learn from the data given to it and make new observations or classifications.*”
- Para uma coleção de dados (registos/conjunto de treino)
- Cada registo é caracterizado por uma tuplo (x, y) , onde x é o conjunto de atributos e y é a classe ou categoria atribuída:
 - x : atributo, preditor, variável independente, entrada
 - y : classe, resposta, variável dependente, saída
- Tarefa:
 - Aprender um modelo que mapeia cada conjunto de atributos x em um das classes predefinidas de y

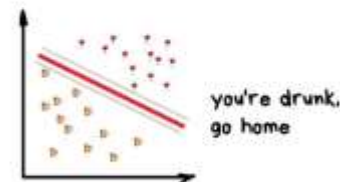
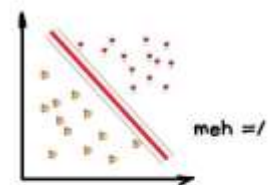


Classificação

Aprendizagem Supervisionada



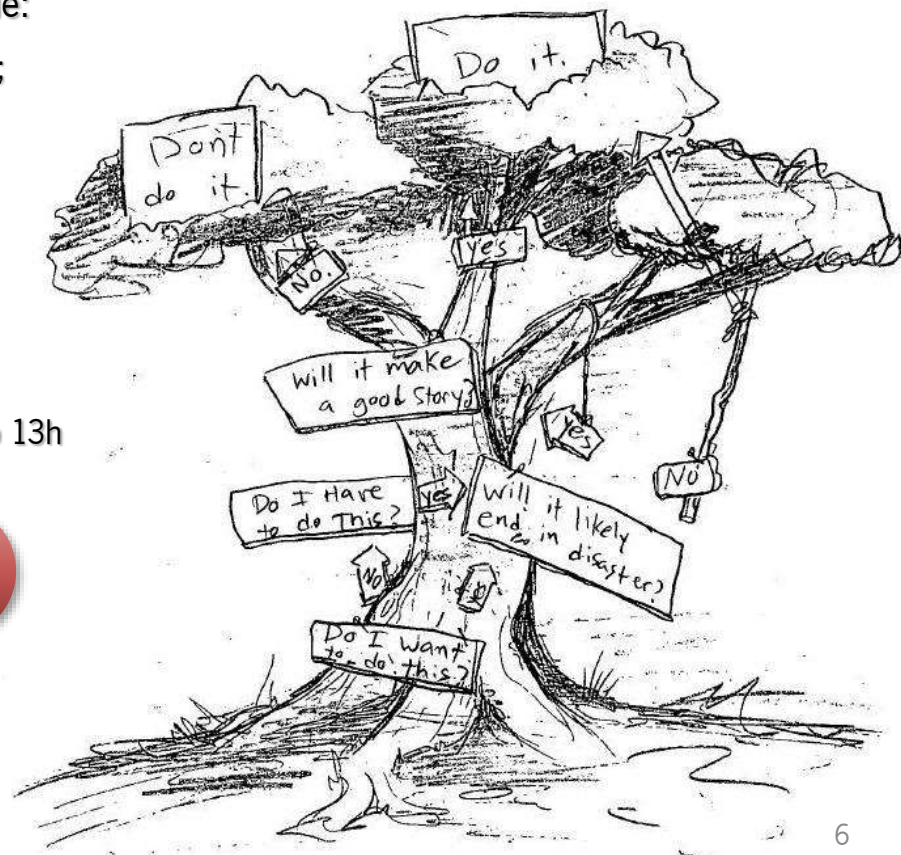
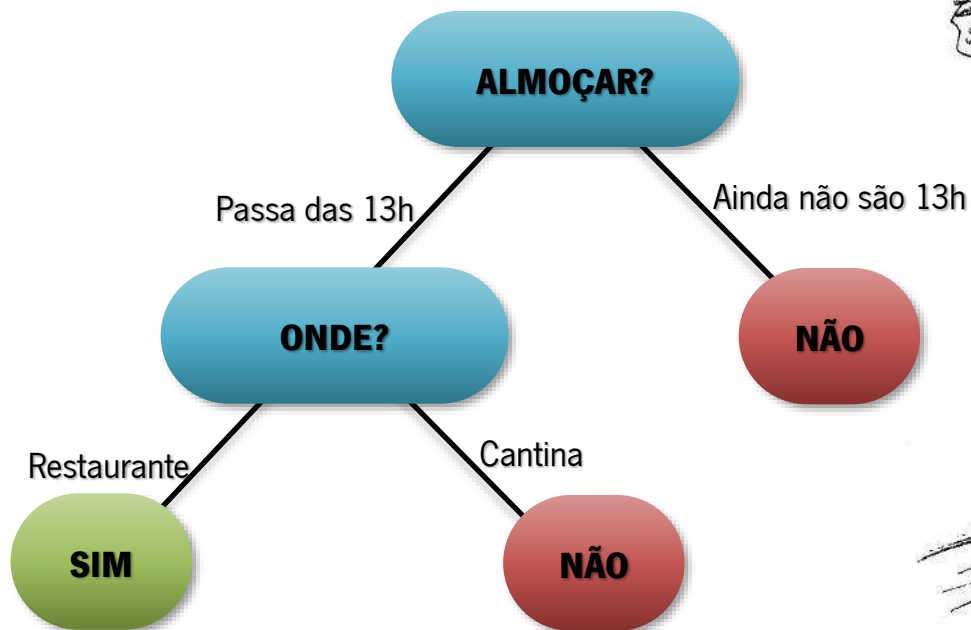
TE TYPES OF ALCOHOL



SUPPORT VECTOR MACHINE

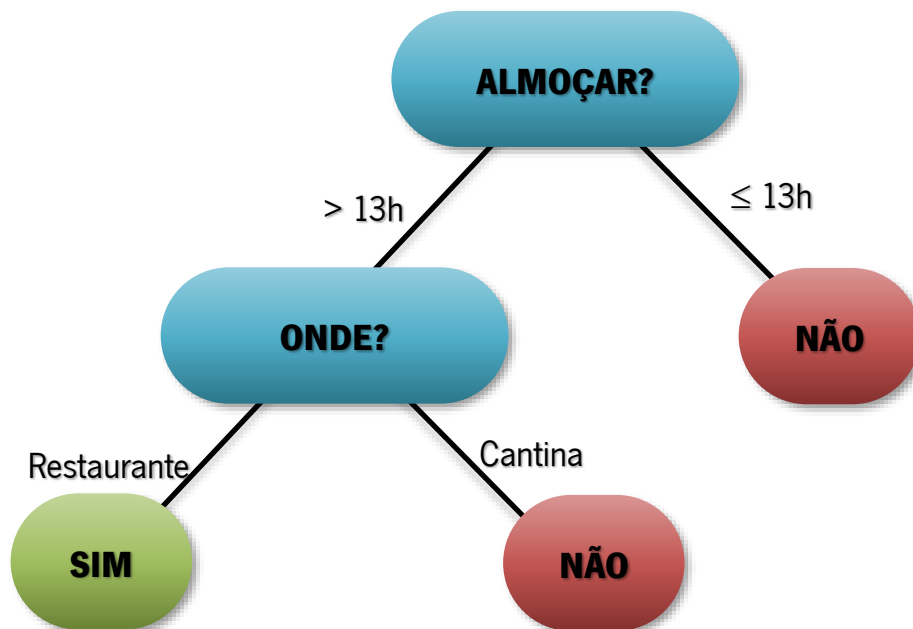
Árvores de Decisão *Decision Trees*

- Uma Árvore de Decisão é um grafo hierarquizado (árvore!) em que:
 - Cada ramo representa a seleção entre um conjunto de alternativas;
 - Cada folha representa uma decisão;



Árvores de Decisão *Decision Trees*

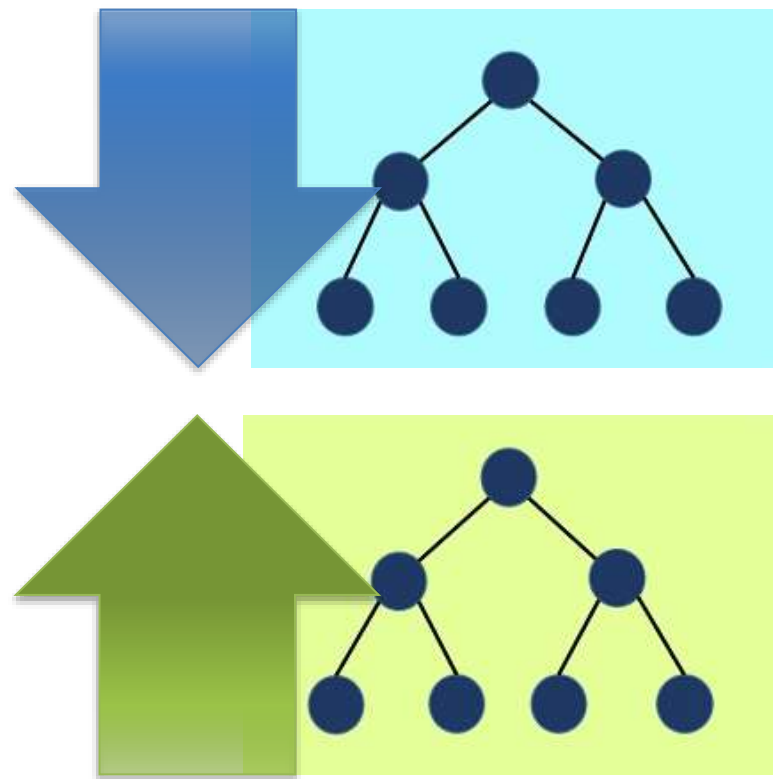
- Uma Árvore de Decisão é um grafo hierarquizado (árvore!) em que:
 - Cada nodo interno testa um atributo do *dataset*;
 - Cada ramo identifica um valor (ou conjunto de valores) do nodo testado;
 - Cada folha representa uma decisão;



ALMOÇAR	ONDE	DECISÃO
12h30	Cantina	NÃO
13h15	Cantina	NÃO
13h10	Restaurante	SIM
11h00	Restaurante	NÃO
13:30	Cantina	NÃO

Modelos de Decisão

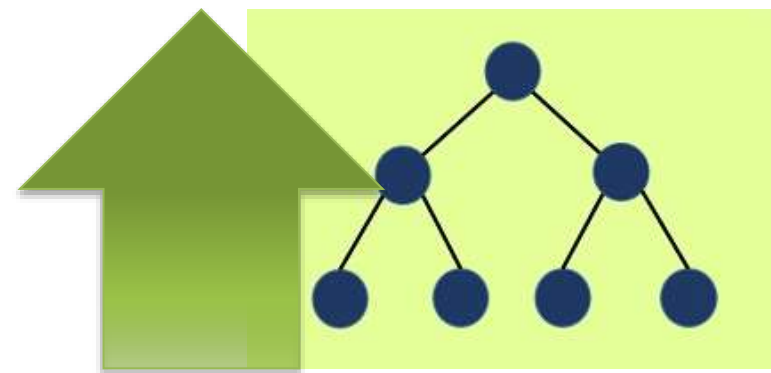
- Paradigmas de criação de modelos de decisão:
 - Top-down:
 - O modelo é construído a partir do conhecimento de especialistas;
 - O “todo” é dividido em “partes”;
 - Bottom-up:
 - O modelo é construído pela identificação de relações entre os atributos do *dataset*;
 - O modelo é induzido por “generalização” dos dados;



Modelos de Decisão

Árvores de Decisão

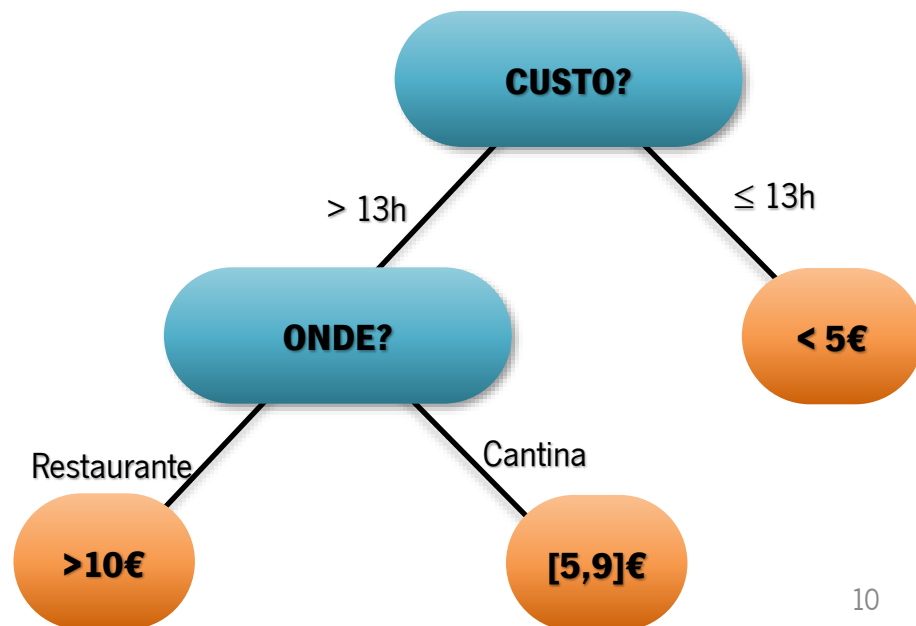
- Árvores de Decisão seguem o Paradigma Bottom-up:
 - Toda a informação sobre cada item de dados (ou objeto) deve estar definido numa coleção fixa e finita de atributos;
 - Deste modo, objetos distintos não podem requerer coleções distintas de atributos;
 - Quando o conjunto dos níveis de decisão é conhecido *a priori*, a construção do modelo segue um paradigma de aprendizagem supervisionado;
 - Quando o conjunto dos níveis de decisão é calculado pelo modelo, a sua construção segue um paradigma de aprendizagem não supervisionado;
 - Os níveis de decisão podem ser de 2 tipos:
 - Contínuos: problemas de regressão;
 - Discretos: problemas de classificação;
 - Quantidade de objetos \gg níveis de decisão;



Tipos de Árvores de Decisão

- Contínuo

- O atributo de decisão representa uma sequência, conjunto ou intervalos de possíveis valores;
- As folhas da árvore de decisão identificam intervalos ou conjuntos de valores;



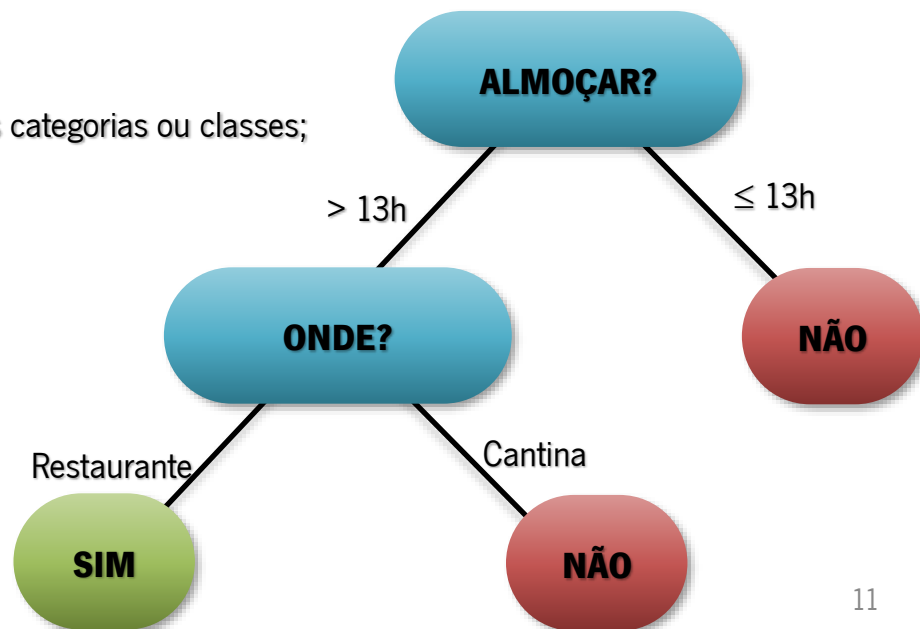
Tipos de Árvores de Decisão

■ Contínuo

- O atributo de decisão representa uma sequência, conjunto ou intervalos de possíveis valores;
- As folhas da árvore de decisão identificam intervalos ou conjuntos de valores;


■ Discreto

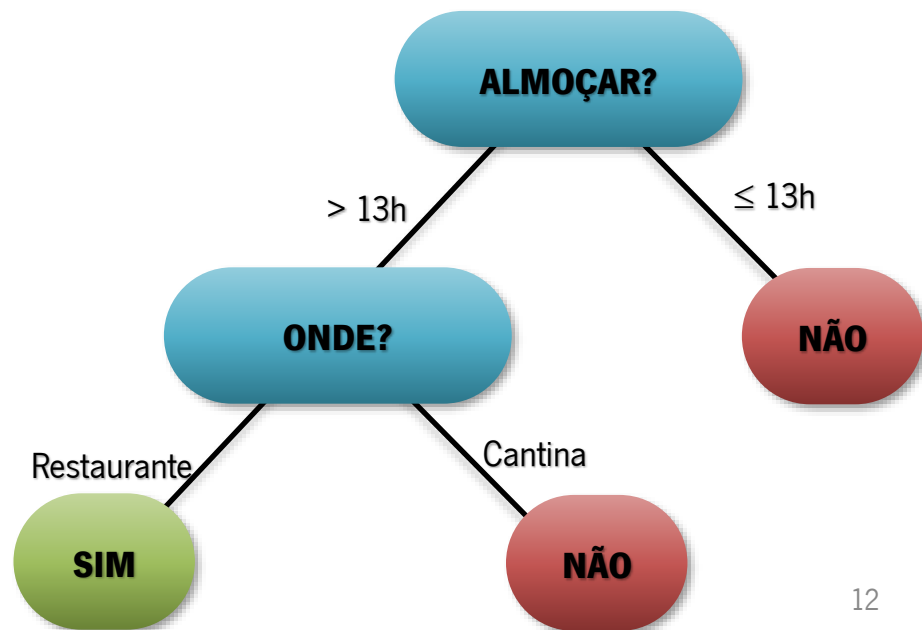
- O atributo de decisão representa uma categoria ou uma classe;
- Os valores representados nas folhas da árvore de decisão são as categorias ou classes;



Ciclo de Execução Resolução de Problemas

- Dada uma árvore de decisão (treinada), o processo de decisão desenvolve-se do seguinte modo:

- 
1. Começar no nodo correspondente ao atributo “raiz”;
 2. Identificar o valor do atributo;
 3. Seguir pelo ramo correspondente ao valor identificado,
 4. Alcançar o nodo relativo ao ramo percorrido;
 5. Voltar a 2. até que o nodo seja uma “folha”;
 6. O nodo alcançado indica a decisão para o problema.



Construção de um Modelo de Árvores de Decisão

- A construção de um modelo baseado em Árvores de Decisão desenvolve-se através de:
 - Observação de exemplos (objetos);
 - Generalização por indução;
 - Criação do modelo;
 - Apresentação de um problema;
 - Obtenção da resolução do problema (previsão);



- Algoritmo ID3: Iterative Dichotomiser 3
link.springer.com/content/pdf/10.1007

- Desenvolvido por Ross Quinlan;
- Constrói uma árvore de decisão, a partir da raiz até às folhas;
- Principal problema:
 - “Qual o melhor atributo para ser a raiz da árvore de decisão?”



- Algoritmo ID3: Iterative Dichotomiser 3

link.springer.com/content/pdf/10.1007

- Desenvolvido por Ross Quinlan;
- Constrói uma árvore de decisão, a partir da raiz até às folhas;
- Principal problema:
 - “Qual o melhor atributo para ser a raiz da árvore de decisão?”

- Noção de Entropia:

- Entropia é uma medida da incerteza associada a um conjunto de objetos;
- A entropia identifica o grau de desorganização dos dados;



- Algoritmo ID3: Iterative Dichotomiser 3
link.springer.com/content/pdf/10.1007

- Entropia:

- Dada uma coleção S de dados contendo exemplos positivos e negativos de um determinado conceito, a entropia de S é definida por:

- $Entropia(S) \equiv -p_+ \log_2 p_+ - p_- \log_2 p_-$

- onde p_+ e p_- representam, respetivamente, a proporção de exemplos positivos e de exemplos negativos em S ;



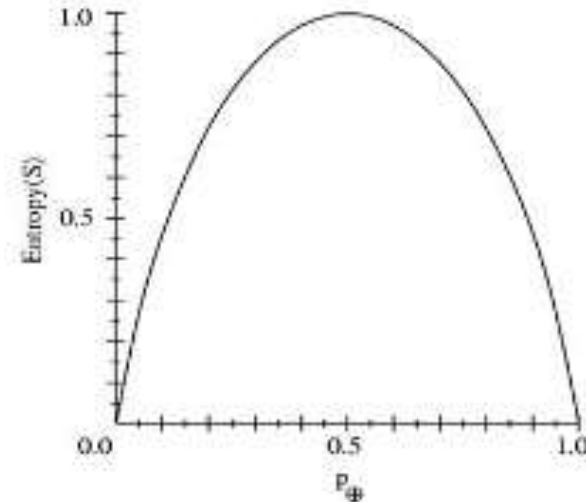
- Algoritmo ID3: Iterative Dichotomiser 3
link.springer.com/content/pdf/10.1007

- Entropia:

- Dada uma coleção S de dados contendo exemplos positivos e negativos de um determinado conceito, a entropia de S é definida por:

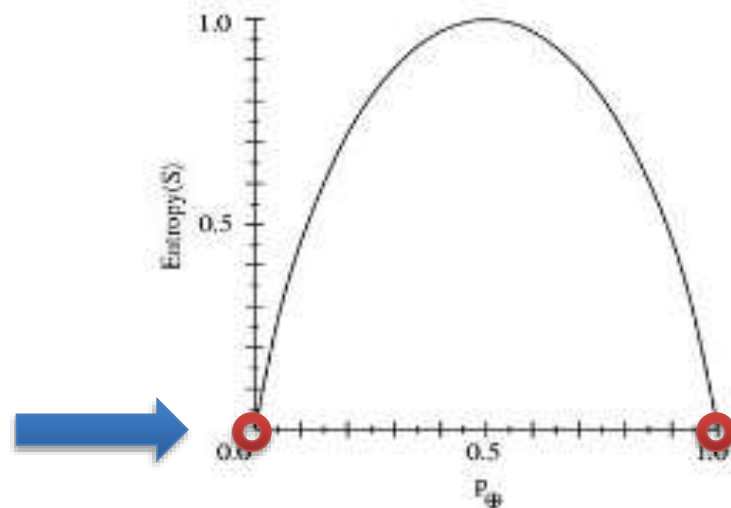
- $Entropia(S) \equiv -p_+ \log_2 p_+ - p_- \log_2 p_-$

- onde p_+ e p_- representam, respetivamente, a proporção de exemplos positivos e de exemplos negativos em S ;
 - Considere a função de entropia definida no gráfico relativa a um atributo binário, sendo que:
 - $p_+ \in [0,1]$
 - $p_- \equiv (1 - p_+) \in [0,1]$



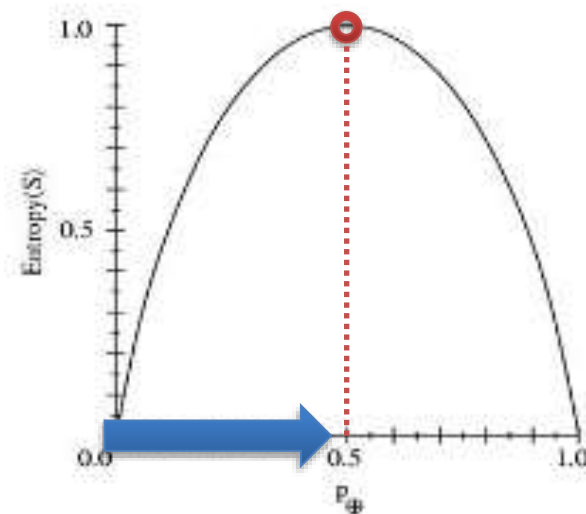
Entropia:

- $Entropia(S) \equiv -p_+ \log_2 p_+ - p_- \log_2 p_-$
- Considere a função de entropia definida no gráfico relativa a um atributo binário, sendo que:
 - $p_+ \in [0,1]$
 - $p_- \equiv (1 - p_+) \in [0,1]$
- A entropia é 0 (zero) quando todos os objetos de S são do mesmo valor;
- Se todos os objetos forem positivos:
 - $p_+ = 1$
 - $p_- = (1 - 1) = 0$
- então:
 - $Entropia(S) = -1 \log_2 1 - 0 \log_2 0$
 $= -1 \times 0 - 0 \times \log_2 0$
 $= 0$



■ Entropia:

- $Entropia(S) \equiv -p_+ \log_2 p_+ - p_- \log_2 p_-$
- Considere a função de entropia definida no gráfico relativa a um atributo binário, sendo que:
 - $p_+ \in [0,1]$
 - $p_- \equiv (1 - p_+) \in [0,1]$
- Quando a entropia é 1 (um), S tem igual proporção de objetos positivos e negativos;
- Sendo:
 - $p_+ = 0,5$
 - $p_- = (1 - 0,5) = 0,5$
- então:
 - $Entropia(S) = -0,5 \log_2 0,5 - 0,5 \log_2 0,5$
 $= -0,5 \times (-1) - 0,5 \times (-1)$
 $= 0,5 + 0,5 = 1$



- Algoritmo ID3: Iterative Dichotomiser 3
link.springer.com/content/pdf/10.1007
- Ganho de Informação:
 - Esta métrica calcula a redução esperada na entropia;
 - Decisão sobre qual o atributo que será selecionado para ser nodo;
 - O atributo com a maior redução de entropia é a melhor escolha para ser nodo;
(para reduzir a profundidade da árvore)



■ Ganho de Informação:

- O ganho de informação $Ganho(S, A)$ de um atributo A relativamente a uma coleção S define-se como:

$$\begin{aligned} \bullet \quad & Ganho(S, A) = Entropia_{original}(S) - Entropia_{relativa}(S) \\ & = Entropia(S) - \sum_{v \in valores(A)} \frac{|S_v|}{|S|} \times Entropia(S_v) \end{aligned}$$

- Sendo:

- S cada valor v de todos os valores possíveis do atributo A ;
- S_v subconjunto de S para o qual o atributo A tem o valor v ;
- $|S_v|$ quantidade de objetos em S_v ;
- $|S|$ quantidade de objetos em S ;



Exemplo

- Aspeto = { Sol, Nuvens, Chuva }
- Temperatura = { Calor, Ameno, Frio }
- Humidade = { Alta, Normal }
- Vento = { Forte, Fraco }

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

- Aspeto = { Sol, Nuvens, Chuva }
- Temperatura = { Calor, Ameno, Frio }
- Humidade = { Alta, Normal }
- Vento = { Forte, Fraco }

- Como calcular o atributo para ser a raiz da árvore de decisão?

Exemplo

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

- Como calcular o atributo para ser a raiz da árvore de decisão?

Exemplo

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

■ Como calcular o atributo para ser a raiz da árvore de decisão?

○ $Entropia(S) = -\left(\frac{9}{14}\right) \log_2 \left(\frac{9}{14}\right) \dots$

Exemplo

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

Exemplo

- Como calcular o atributo para ser a raiz da árvore de decisão?

- $$Entropia(S) = -\left(\frac{9}{14}\right) \log_2 \left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2 \left(\frac{5}{14}\right)$$

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

Exemplo

- Como calcular o atributo para ser a raiz da árvore de decisão?

$$\begin{aligned}
 \circ \text{Entropia}(S) &= -\left(\frac{9}{14}\right) \log_2 \left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2 \left(\frac{5}{14}\right) \\
 &= -(0,643) \times (-0,637) - 0,357 \times (-1,485) \\
 &= 0,410 + 0,530 = 0,940
 \end{aligned}$$

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

Exemplo

- Como calcular o atributo para ser a raiz da árvore de decisão?

$$\begin{aligned}
 \circ \text{Entropia}(S) &= -\left(\frac{9}{14}\right) \log_2 \left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2 \left(\frac{5}{14}\right) \\
 &= -(0,643) \times (-0,637) - 0,357 \times (-1,485) \\
 &= 0,410 + 0,530 = 0,940 \\
 \circ \text{Ganho}(S, \text{Vento}) &= \text{Entropia}(S) \\
 &\quad - \left(\frac{8}{14}\right) \text{Entropia}(S_{\text{Fraco}}) \\
 &\quad - \left(\frac{6}{14}\right) \text{Entropia}(S_{\text{Forte}})
 \end{aligned}$$

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

Exemplo

- Como calcular o atributo para ser a raiz da árvore de decisão?

$$\begin{aligned} \circ \text{Entropia}(S) &= -\left(\frac{9}{14}\right) \log_2 \left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2 \left(\frac{5}{14}\right) \\ &= -(0,643) \times (-0,637) - 0,357 \times (-1,485) \end{aligned}$$

$$= 0,410 + 0,530 = 0,940$$

$$\begin{aligned} \circ \text{Ganho}(S, \text{Vento}) &= \text{Entropia}(S) \\ &\quad - \left(\frac{8}{14}\right) \text{Entropia}(S_{\text{Fraco}}) \\ &\quad - \left(\frac{6}{14}\right) \text{Entropia}(S_{\text{Forte}}) \end{aligned}$$

$$\circ \text{Entropia}(S_{\text{Fraco}}) = -\left(\frac{6}{8}\right) \log_2 \left(\frac{6}{8}\right) - \left(\frac{2}{8}\right) \log_2 \left(\frac{2}{8}\right) = 0,811$$

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

Exemplo

- Como calcular o atributo para ser a raiz da árvore de decisão?

$$\begin{aligned} \circ \text{Entropia}(S) &= -\left(\frac{9}{14}\right) \log_2 \left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2 \left(\frac{5}{14}\right) \\ &= -(0,643) \times (-0,637) - 0,357 \times (-1,485) \end{aligned}$$

$$= 0,410 + 0,530 = 0,940$$

$$\begin{aligned} \circ \text{Ganho}(S, \text{Vento}) &= \text{Entropia}(S) \\ &\quad - \left(\frac{8}{14}\right) \text{Entropia}(S_{\text{Fraco}}) \\ &\quad - \left(\frac{6}{14}\right) \text{Entropia}(S_{\text{Forte}}) \end{aligned}$$

$$\circ \text{Entropia}(S_{\text{Fraco}}) = -\left(\frac{6}{8}\right) \log_2 \left(\frac{6}{8}\right) - \left(\frac{2}{8}\right) \log_2 \left(\frac{2}{8}\right) = 0,811$$

$$\circ \text{Entropia}(S_{\text{Forte}}) = -\left(\frac{3}{6}\right) \log_2 \left(\frac{3}{6}\right) - \left(\frac{3}{6}\right) \log_2 \left(\frac{3}{6}\right) = 1,0$$

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

Exemplo

- Como calcular o atributo para ser a raiz da árvore de decisão?

$$\begin{aligned} \circ \text{Entropia}(S) &= -\left(\frac{9}{14}\right) \log_2 \left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2 \left(\frac{5}{14}\right) \\ &= -(0,643) \times (-0,637) - 0,357 \times (-1,485) \end{aligned}$$

$$= 0,410 + 0,530 = 0,940$$

$$\begin{aligned} \circ \text{Ganho}(S, \text{Vento}) &= \text{Entropia}(S) \\ &\quad - \left(\frac{8}{14}\right) \text{Entropia}(S_{\text{Fraco}}) \\ &\quad - \left(\frac{6}{14}\right) \text{Entropia}(S_{\text{Forte}}) \\ &= 0,940 - \left(\frac{8}{14}\right) \times 0,811 - \left(\frac{6}{14}\right) \times 1,0 \\ &= 0,048 \end{aligned}$$

$$\circ \text{Entropia}(S_{\text{Fraco}}) = -\left(\frac{6}{8}\right) \log_2 \left(\frac{6}{8}\right) - \left(\frac{2}{8}\right) \log_2 \left(\frac{2}{8}\right) = 0,811$$

$$\circ \text{Entropia}(S_{\text{Forte}}) = -\left(\frac{3}{6}\right) \log_2 \left(\frac{3}{6}\right) - \left(\frac{3}{6}\right) \log_2 \left(\frac{3}{6}\right) = 1,0$$

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

Exemplo

- Como calcular o atributo para ser a raiz da árvore de decisão?

- $Entropia(S) = 0,940$

- $Ganho(S, Vento) = 0,048$

- $Ganho(S, Aspeto) = 0,246$

- $Ganho(S, Temperatura) = 0,029$

- $Ganho(S, Humidade) = 0,151$

- Qual o atributo que apresenta maior ganho de informação?

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

Exemplo

- Como calcular o atributo para ser a raiz da árvore de decisão?
 - $Entropia(S) = 0,940$
 - $Ganho(S, Vento) = 0,048$
 - $Ganho(S, Aspeto) = 0,246$
 - $Ganho(S, Temperatura) = 0,029$
 - $Ganho(S, Humidade) = 0,151$
- O atributo Aspeto é o que apresenta maior ganho de informação;
- Será este o atributo selecionado para a raiz da árvore de decisão;

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

- Como calcular o atributo para ser a raiz da árvore de decisão?
- O atributo Aspeto é o que apresenta maior ganho de informação;
- Será este o atributo selecionado para a raiz da árvore de decisão;

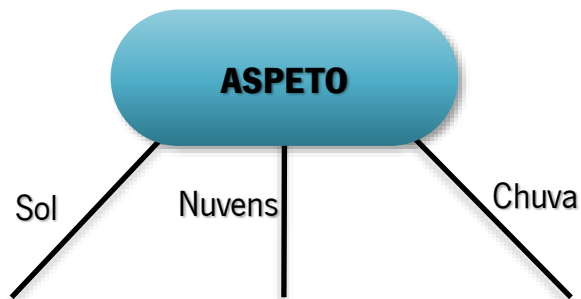
ASPETO

Exemplo

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

Exemplo

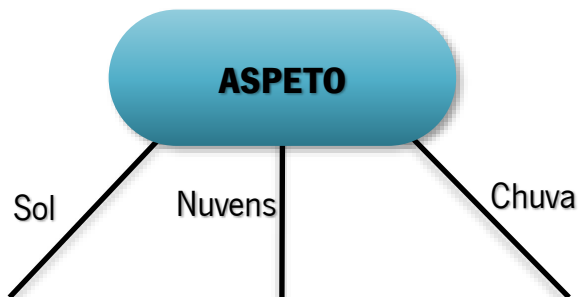
- Como calcular o atributo para ser a raiz da árvore de decisão?
- O atributo Aspeto é o que apresenta maior ganho de informação;
- Será este o atributo selecionado para a raiz da árvore de decisão;
- Este atributo tem 3 valores possíveis, pelo que surgirão 3 ramos desde a raiz, um para cada valor do atributo;



Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

Exemplo

- Como calcular o atributo para ser a raiz da árvore de decisão?
- O atributo Aspeto é o que apresenta maior ganho de informação;
- Será este o atributo selecionado para a raiz da árvore de decisão;
- Este atributo tem 3 valores possíveis, pelo que surgirão 3 ramos desde a raiz, um para cada valor do atributo;

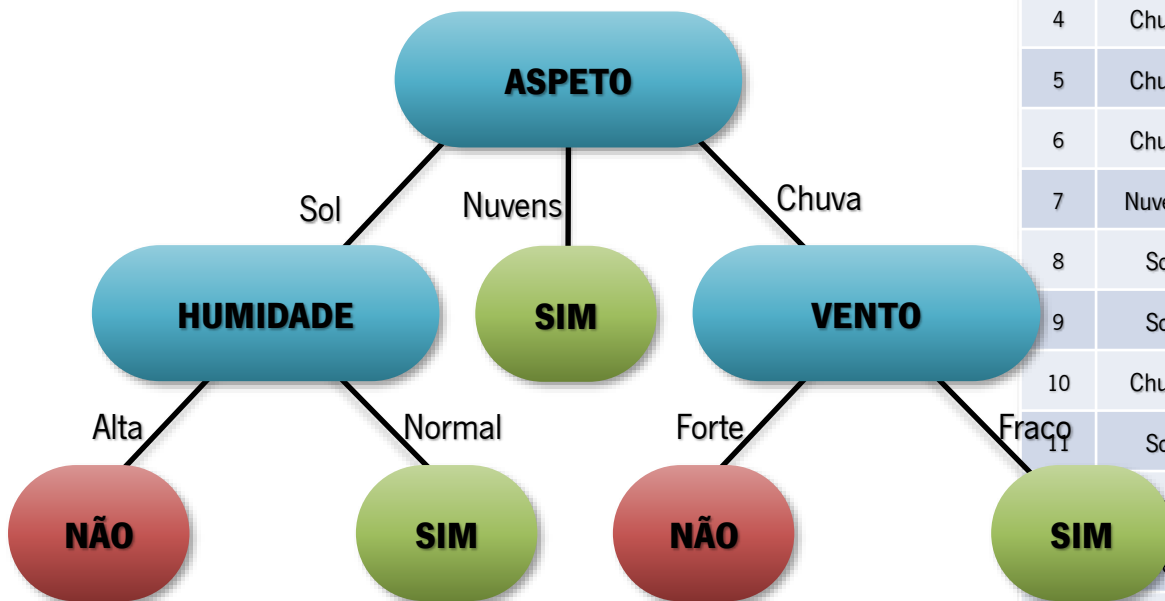


- E agora, que atributo deve ser testado no nodo correspondente ao ramo “Sol”?

Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

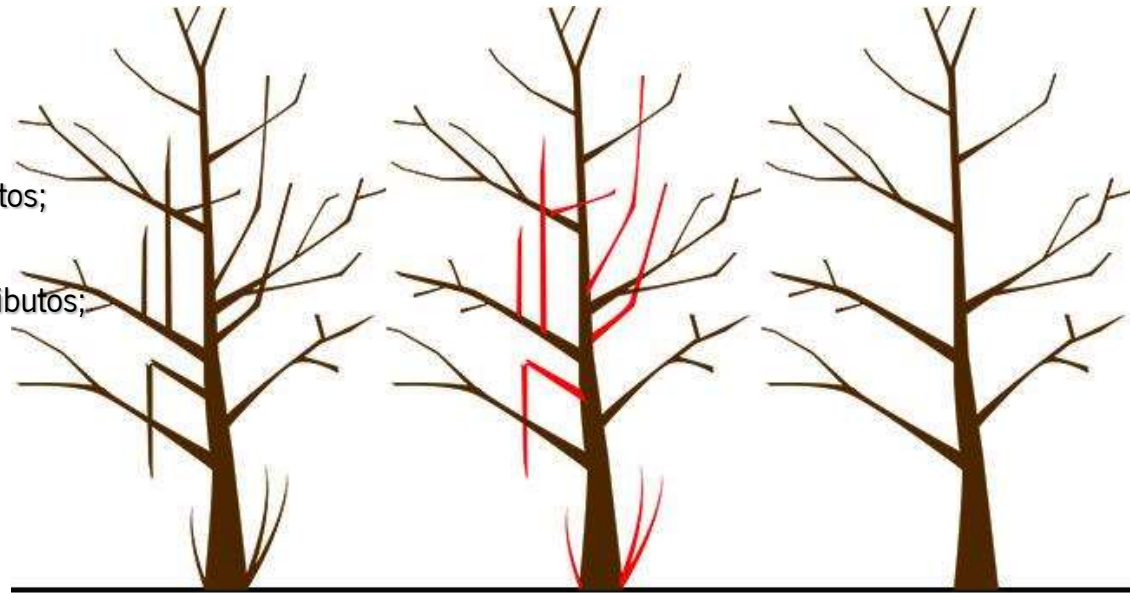
Exemplo

- Prosseguindo o cálculo da entropia para os dados restantes, alcança-se uma árvore de decisão como:



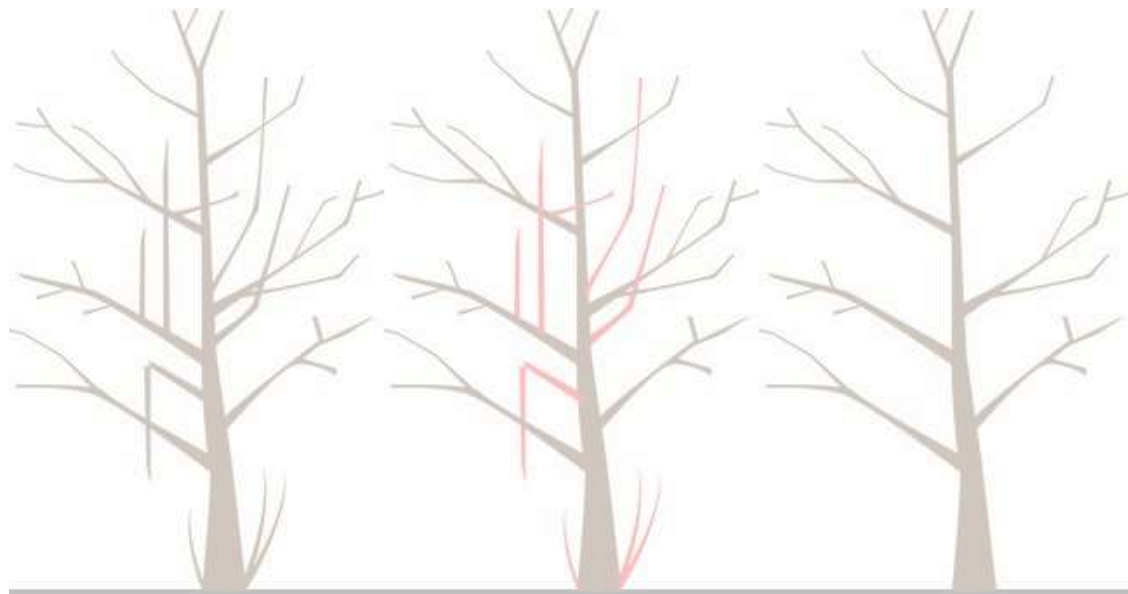
Dia	Aspeto	Temperatura	Humidade	Vento	Decisão
1	Sol	Calor	Alta	Fraco	Não
2	Sol	Calor	Alta	Forte	Não
3	Nuvens	Calor	Alta	Fraco	Sim
4	Chuva	Ameno	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nuvens	Frio	Normal	Forte	Sim
8	Sol	Ameno	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Ameno	Normal	Fraco	Sim
11	Sol	Ameno	Normal	Forte	Sim
12	Nuvens	Ameno	Alta	Forte	Sim
13	Nuvens	Calor	Normal	Fraco	Sim
14	Chuva	Ameno	Alta	Forte	Não

- Algoritmo ID3: Iterative Dichotomiser 3
link.springer.com/content/pdf/10.1007
- Algoritmo C4.5
 - Extensão do algoritmo ID3;
 - Desenvolvido, igualmente, por Ross Quinlan;
 - Melhorias face ao ID3:
 - Manipula atributos contínuos e discretos;
 - Lida com *missing values*;
 - Permite a atribuição de pesos aos atributos;
 - **Permite fazer a poda da árvore;**



■ Algoritmo C4.5

- Manipula atributos contínuos e discretos:
 - Para lidar com atributos contínuos, é definido um limite (*threshold*) usado para dividir os valores acima e abaixo do limite;
- Lida com *missing values*;
- Permite a atribuição de pesos aos atributos;
- Permite fazer a poda da árvore;



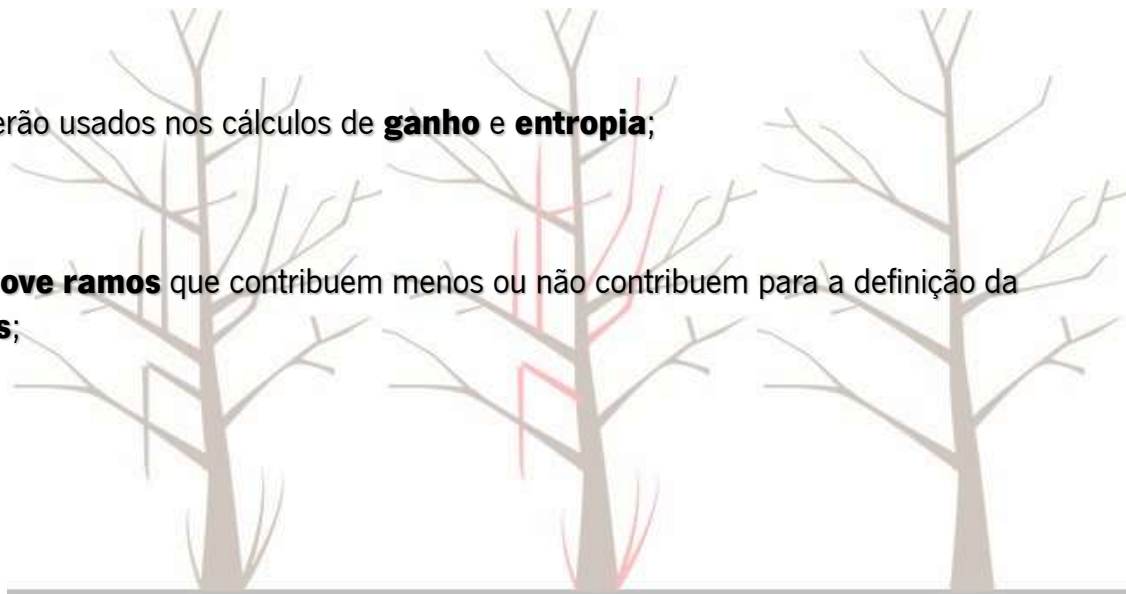
■ Algoritmo C4.5

- Manipula atributos contínuos e discretos:
 - Para lidar com atributos contínuos, é definido um limite (*threshold*) usado para dividir os valores **acima** e **abaixo** do limite;
- Lida com *missing values*:
 - Assinala os *missing values* que não serão usados nos cálculos de **ganho** e **entropia**;
- Permite a atribuição de pesos aos atributos;
- Permite fazer a poda da árvore;

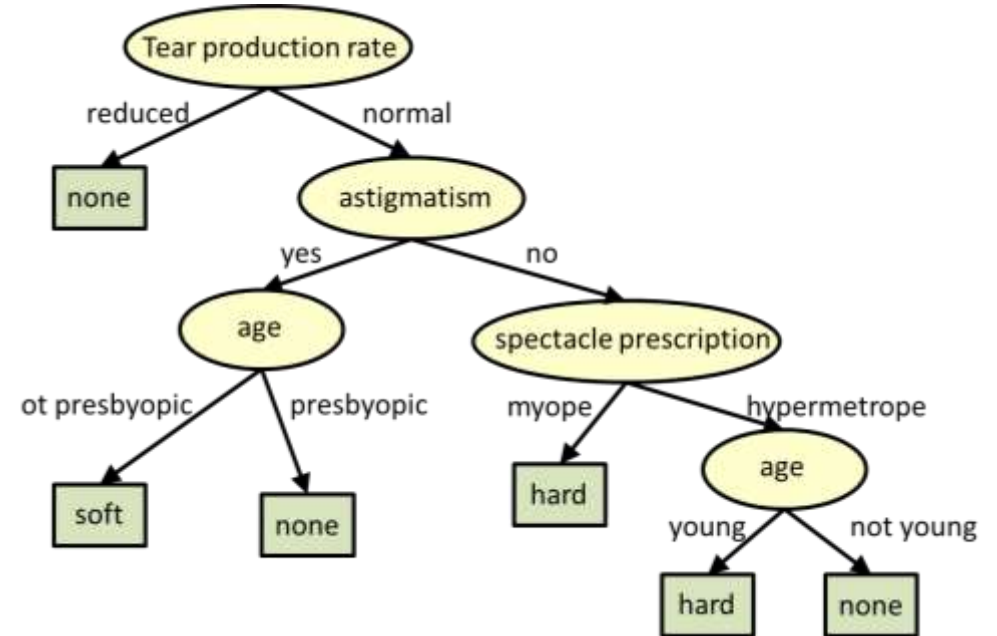


■ Algoritmo C4.5

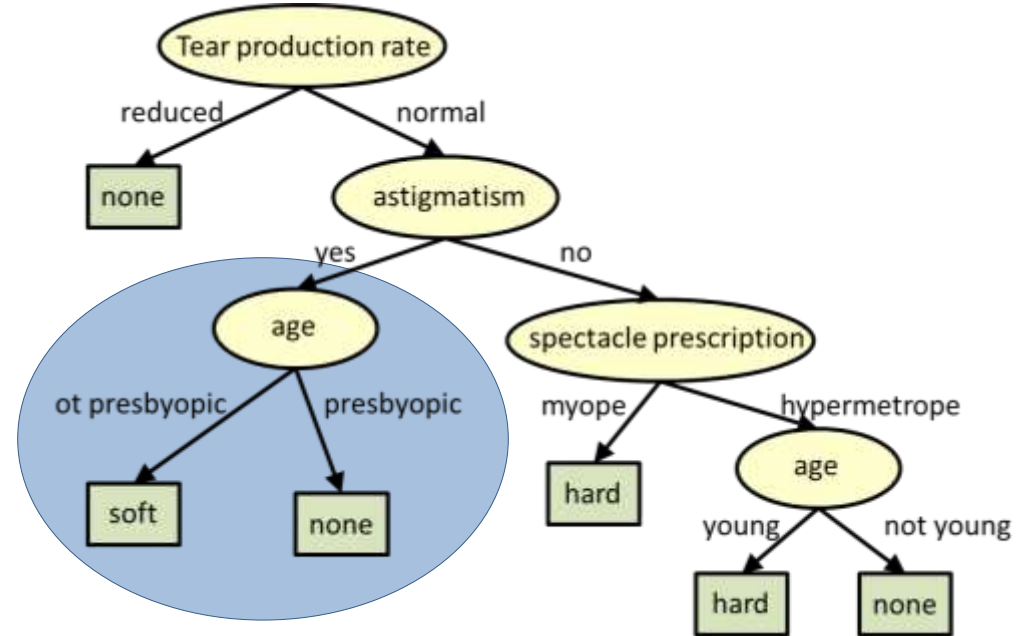
- Manipula atributos contínuos e discretos:
 - Para lidar com atributos contínuos, é definido um limite (*threshold*) usado para dividir os valores **acima** e **abaixo** do limite;
- Lida com *missing values*:
 - Assinala os *missing values* que não serão usados nos cálculos de **ganho** e **entropia**;
- Permite a atribuição de pesos aos atributos;
- Permite fazer a poda da árvore:
 - Retrocede 1 iteração na árvore e **remove ramos** que contribuem menos ou não contribuem para a definição da solução, **substituindo-os por folhas**;



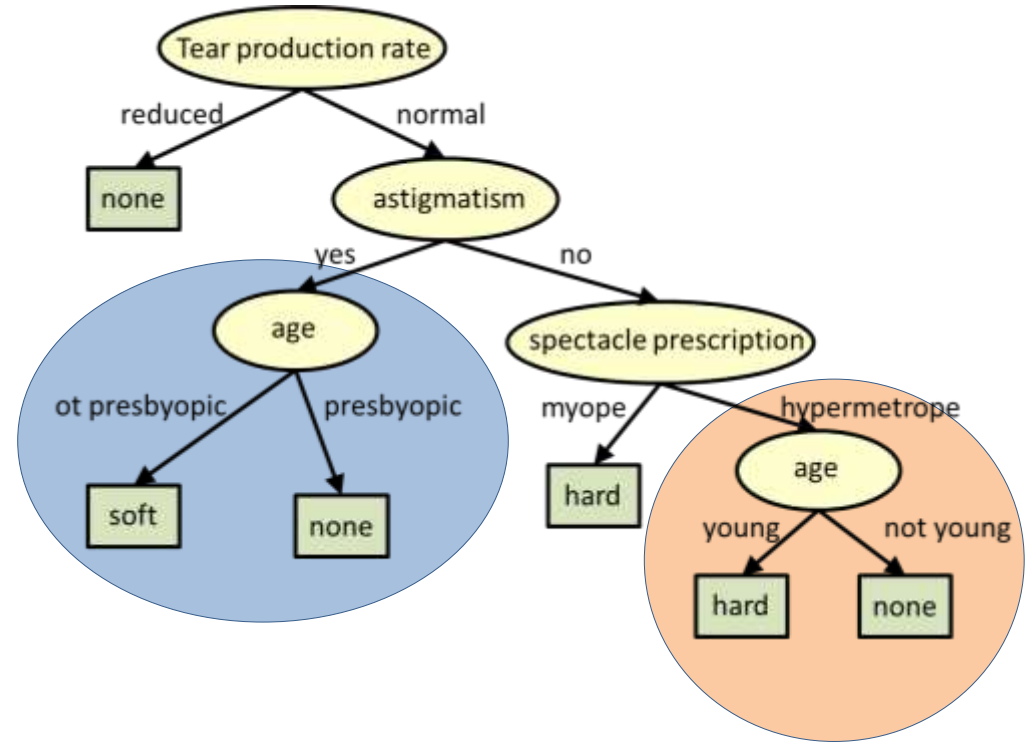
- Algoritmo C4.5
 - Tree Pruning



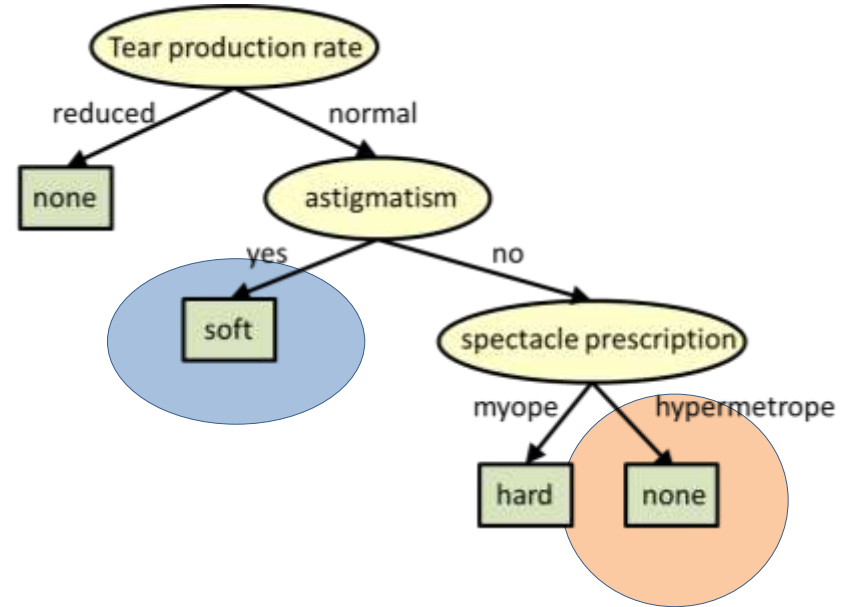
- Algoritmo C4.5
 - Tree Pruning



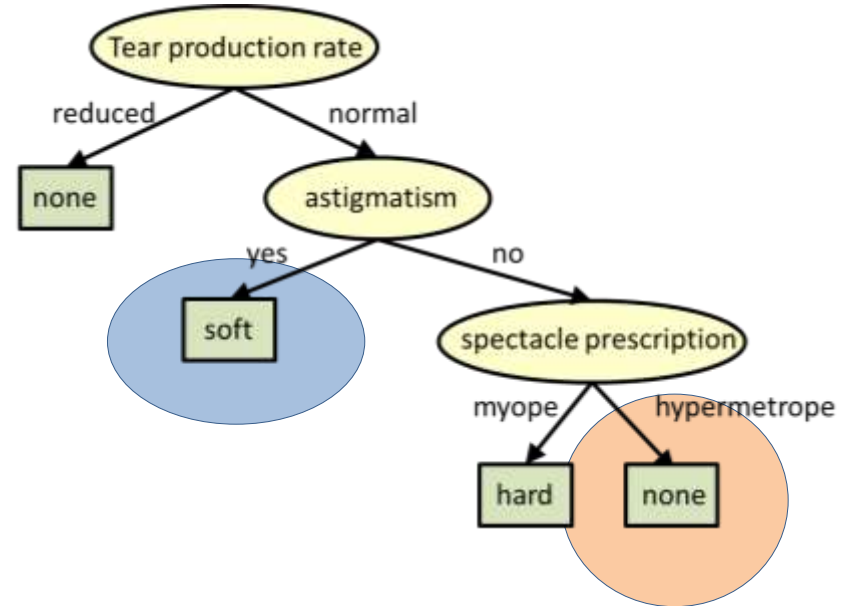
- Algoritmo C4.5
 - Tree Pruning



- Algoritmo C4.5
 - Tree Pruning



- Algoritmo C4.5
 - Tree Pruning
- Porquê?

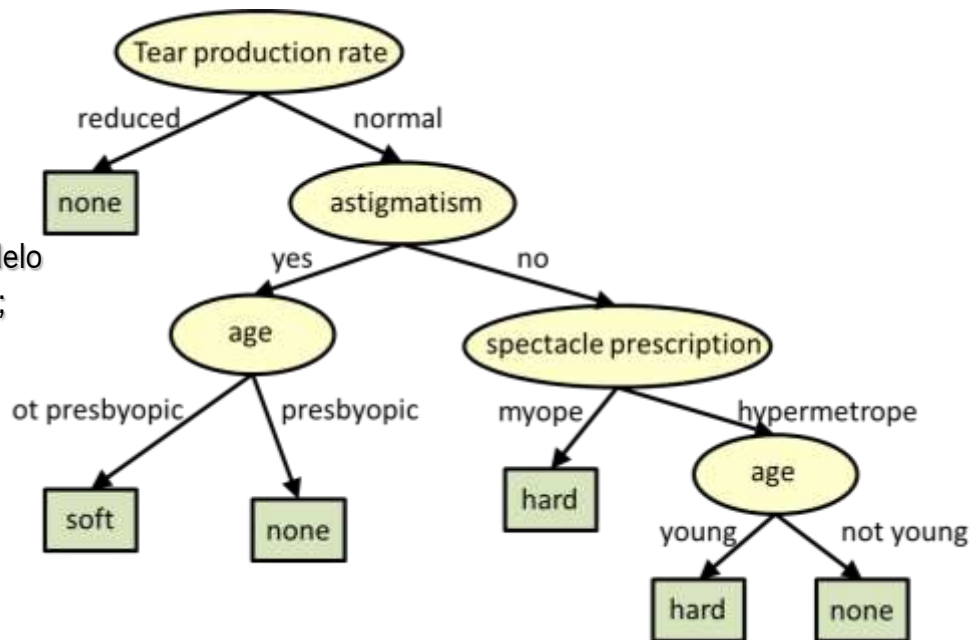


■ Algoritmo C4.5

- Tree Pruning

■ Porquê?

- Porque uma Árvore de Decisão pode resultar num modelo de decisão “demasiado” adaptado aos dados de treino;
- Cada folha pode representar um caso ou conjunto de casos muito específicos;



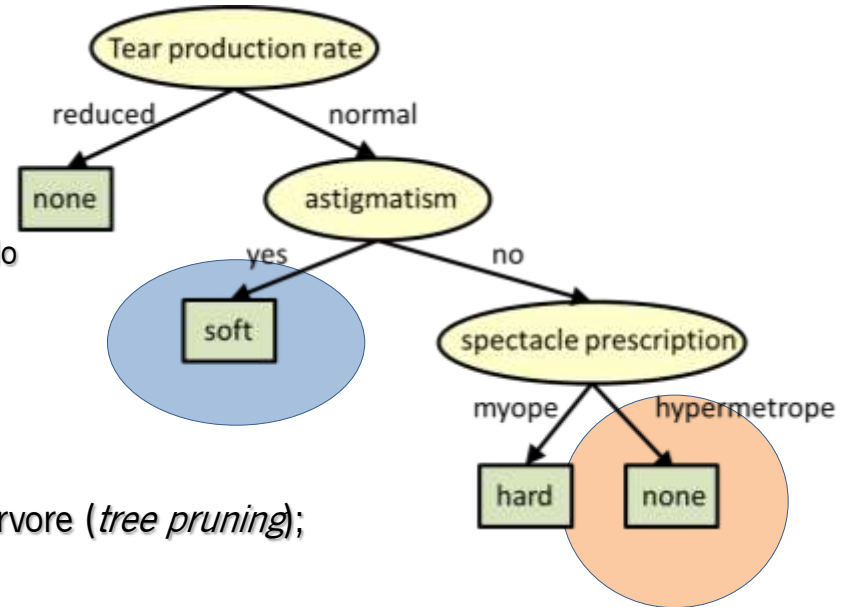
- Algoritmo C4.5

- Tree Pruning

- Porquê?

- Porque uma Árvore de Decisão pode resultar num modelo de decisão “demasiado” adaptado aos dados de treino;
 - Cada folha pode representar um caso ou conjunto de casos muito específicos;

- Para prevenir que isto aconteça, aplica-se a “poda” da árvore (*tree pruning*);



- Algoritmo ID3: Iterative Dichotomiser 3
link.springer.com/content/pdf/10.1007
- Algoritmo C4.5
- Algoritmo J48
 - Implementação do algoritmo C4.5 *open source* em JAVA na plataforma WEKA;
 - WEKA: Waikato Environment for Knowledge Analysis;
(<https://www.cs.waikato.ac.nz/ml/weka>)



- Algoritmo ID3: Iterative Dichotomiser 3
link.springer.com/content/pdf/10.1007
- Algoritmo C4.5
- Algoritmo J48
- Algoritmo CART: Classification and Regression Tree
 - Introduzido por Breiman, praticamente em paralelo com o ID3 de Ross Quinlan;
 - Um mesmo algoritmo que partilha as semelhanças de modelos de classificação e de regressão;

- Algoritmo ID3: Iterative Dichotomiser 3
link.springer.com/content/pdf/10.1007
- Algoritmo C4.5
- Algoritmo J48
- Algoritmo CART: Classification and Regression Tree
- Algoritmo CHAID: Chi-square Automatic Interaction Detection
 - Opera a separação dos dados em modo multi-nível, enquanto que o CART usa modos binários para essa divisão;
 - Adequado para grandes *datasets*;
 - Frequentemente utilizado em estudos de *marketing* para segmentação de mercados;

■ Pontos fortes:

- Facilmente compreensíveis;
- Podem ser convertidas para regras;
- Manipulam *missing values*;
- Configuração simples (não tem demasiados parâmetros de configuração);

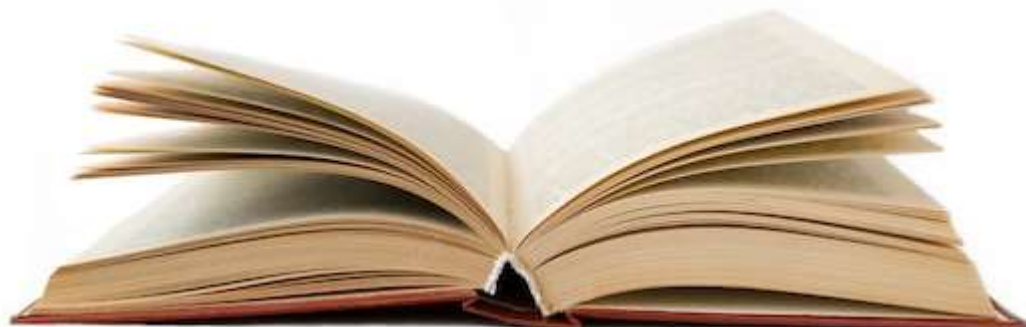
■ Pontos fracos:

- Inadequadas para problemas caracterizados por muitas interações entre os atributos;
- Falta de poder expressivo;
- Não isenta a réplicas de subárvores;



Referências bibliográficas

- Ross Quinlan (1986), “Induction of decision trees”, Machine Learning, 1(1):81-106 (<http://hunch.net/~coms-4771/quinlan.pdf>)
- Ross Quinlan (1993), “C4.5 Programs for Machine Learning”, Morgan Kaufmann
- Breiman, Leo; Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984), “Classification and regression trees”, Monterey, CA
- Pang Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar (2018) “Introduction to Data Mining”, ISBN 9780133128901





Universidade do Minho
Departamento de Informática

APRENDIZAGEM E DECISÃO INTELIGENTES

**LEI/MiEI @ 2022/2023, 2º sem
[ADI³]**