

# INE5644 - Data Mining

## Exercício - Árvore de decisão - Parte 2

Bruno Marques do Nascimento\*

30 de Abril de 2018

**Respostas:**

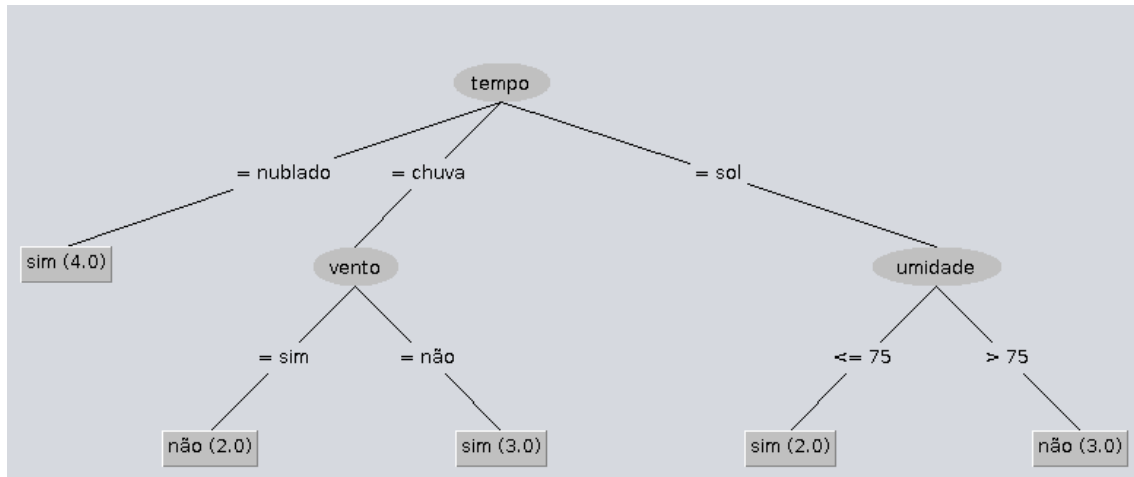
### Exercício slide 1

Tabela 1 – Praticar ou não praticar esportes?

tempo	temperatura	umidade	vento	jogar
nublado	64	65	sim	sim
chuva	65	70	sim	não
chuva	68	80	não	sim
sol	69	70	não	sim
chuva	70	96	não	sim
chuva	71	91	sim	não
sol	72	95	não	não
nublado	72	90	sim	sim
chuva	75	80	não	sim
sol	75	70	sim	sim
sol	80	90	sim	não
nublado	81	75	não	sim
nublado	83	86	não	sim
sol	85	85	não	não

---

\*brunomn95@gmail.com - Universidade Federal de Santa Catarina



Comparando a árvore acima (gerada no weka), com a presente nos slides constata-se que ambas são iguais.

## Exercício slide 2

Tabela 2 – Exercício slide

Peso	Idade	Sexo	Classe
35	criança	Fem	1
50	criança	Masc	1
60	adulto	Fem	1
70	jovem	Masc	2
75	jovem	Masc	2
80	adulto	Masc	2
85	adulto	Masc	2

Através da ordenação é possível ver a presença de 1 limiar de valor **65**.

$$\begin{aligned}
 Entropia(S) &= -\left(\frac{3}{7} \times \log_2 \frac{3}{7}\right) - \left(\frac{4}{7} \times \log_2 \frac{4}{7}\right) \\
 &= -(0.43 \times \log_2 0.43) - (0.57 \times \log_2 0.57) \\
 &= 0.986
 \end{aligned}$$

$$\begin{aligned}
 Entropia(Peso = "< 65") &= -\left(\frac{3}{3} \times \log_2 \frac{3}{3}\right) - \left(\frac{0}{3} \times \log_2 \frac{0}{3}\right) \\
 &= -(1 \times \log_2 1) - \cancel{(0 \times \log_2 \frac{0}{2})} \\
 &= -(1 \times 0) \\
 &= 0
 \end{aligned}$$

**a) Qual o objetivo de se executar poda (*prunning*) em árvores de decisão?**

Existem dois objetivos principais, maximizar o desempenho da árvore de decisão em questão sem perder seu poder de decisão e reduzir seu tamanho físico, ou seja, tornar a árvore de decisão mais leve, irá ocupar menos espaço em disco.

**b) Quais tipos de poda (*prunning*) estão previstos na literatura?**

Na literatura encontram-se presentes podas com abordagem *top-down* e *bottom-up*.

**c) Para o dataset de exemplo, qual o impacto da poda (*prunning*) nas árvores resultado? Quais são as diferenças perceptíveis entre as árvores geradas?**

## Referências

SAVARIS, A. *Aula 20 - Classificação - Conceitos básicos e Árvores de Decisão - Parte 2*. [S.l.], 2018. Acesso em: 24 abril 2018. Nenhuma citação no texto.

SAVARIS, A. *Aula 20 - Classificação - Conceitos básicos e Árvores de Decisão - Parte 2 - Prática*. [S.l.], 2018. Acesso em: 24 abril 2018. Nenhuma citação no texto.

WITTEN, I. H. et al. *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques*. 4th. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2016. ISBN 0128042915, 9780128042915. Nenhuma citação no texto.